Comput Optim Appl
DOI 10.1007/s10589-008-9167-8

Cone-constrained eigenvalue problems: theory and algorithms

A. Pinto da Costa · A. Seeger

Received: 5 July 2007 / Revised: 11 January 2008

Abstract Equilibria in mechanics or in transportation models are not always expressed through a system of equations, but sometimes they are characterized by means of complementarity conditions involving a convex cone. This work deals with the analysis of cone-constrained eigenvalue problems. We discuss some theoretical issues like, for instance, the estimation of the maximal number of eigenvalues in a cone-constrained problem. Special attention is paid to the Paretian case. As a short addition to the theoretical part, we introduce and study two algorithms for solving numerically such type of eigenvalue problems.

Keywords Complementarity conditions · Generalized eigenvalue problems · Convex cones

1 Introduction

This paper is concerned with the analysis of a cone-constrained eigenvalue problem arising in mechanics [13, 14] and in various areas of applied mathematics [8, 10, 15–18]. The notation that we employ is for the most part standard. The Euclidean space $\mathbb{R}^n$ is equipped with the inner product $\langle y, x\rangle = y^T x$ and the associated norm $\|\cdot\|$. Orthogonality with respect to $\langle\cdot,\cdot\rangle$ is indicated by means of the symbol $\perp$. We also use the notation

$$\mathbb{M}_n \equiv \text{real matrices of size } n \times n,$$
$$\Xi(\mathbb{R}^n) \equiv \text{closed convex cones in } \mathbb{R}^n.$$

A. Pinto da Costa
Departamento de Engenharia Civil e Arquitectura and ICIST, Instituto Superior Técnico,
Avenida Rovisco Pais, 1049-001 Lisboa, Portugal
e-mail: apcosta@civil.ist.utl.pt

A. Seeger (✉)
Department of Mathematics, University of Avignon, 33 rue Louis Pasteur, 84000 Avignon, France
e-mail: alberto.seeger@univ-avignon.fr

Given a pair $(A, B) \in \mathbb{M}_n \times \mathbb{M}_n$ and a cone $K \in \Xi(\mathbb{R}^n)$, we are interested in solving an abstract eigenvalue problem of the form:

Problem 1

$$\begin{aligned} &\text{Find } \lambda \in \mathbb{R} \text{ and a nonzero vector } x \in \mathbb{R}^n \text{ such that} \\ &K \ni x \perp (Ax - \lambda Bx) \in K^+. \end{aligned} \tag{1}$$

The specific meanings of the matrices A and B depend on the context. The cone K is interpreted as a constraint set and $K^+ = \{y \in \mathbb{R}^n : \langle y, x\rangle \geq 0,\ \forall x \in K\}$ refers to the dual cone of K. Throughout this work one assumes that

$$K \neq -K \quad \text{and} \quad \langle x, Bx\rangle \neq 0 \quad \forall x \in K \backslash \{0\}. \tag{2}$$

The first hypothesis in (2) says that K is not a linear subspace. Linearly constrained eigenvalue problems fall within the realm of classical linear algebra and therefore we leave them out of the present discussion. The second hypothesis is not restrictive in practice and helps avoiding degenerate situations like unboundedness in the set

$$\sigma_K(A, B) = \{\lambda \in \mathbb{R} :\ (x, \lambda) \text{ solves (1) for some } x \neq 0\}.$$

One refers to $\sigma_K(A, B)$ as the K-spectrum (or set of K-eigenvalues) of the pair (A, B). The first component of a solution (x, λ) is called a K-eigenvector of (A, B).

The orthogonality condition appearing in (1) implies that λ and x are related by the Rayleigh-Ritz ratio $\lambda = \langle x, Ax\rangle / \langle x, Bx\rangle$ as happens in the classical unconstrained setting. However, the "residual" vector

$$y = Ax - \lambda Bx \tag{3}$$

does not need to be zero. That (3) belongs to K^+ is less demanding than the usual condition $Ax = \lambda Bx$, but, on the other hand, x is required to be in K. So, no monotonicity property with respect to K is to be expected from the set $\sigma_K(A, B)$.

An example of Problem 1 arising in mechanics is presented below. It has to do with the search for instabilities of mechanical systems in the presence of obstacles with friction [14].

Example 1 Consider the finite element discretization of an equilibrium configuration of a solid in the presence of an obstacle (cf. Fig. 1). Labels f or c denote the nodes that, at the equilibrium state under consideration, are out of contact (free from any geometrical constraint) or in contact, respectively. A specific type of instability study leads to a complementarity eigenvalue problem involving two different types of variables that, together, represent the rate of change of the equilibrium state: vector x containing the kinematic variables (i.e., the velocities) and vector y containing the static variables (i.e., the reaction rates). The previous vectors may be decomposed in subvectors: x_f and y_f, both with n_f components, grouping the "free" variables (that do not have to satisfy any inequality or complementarity condition), and x_c and y_c,

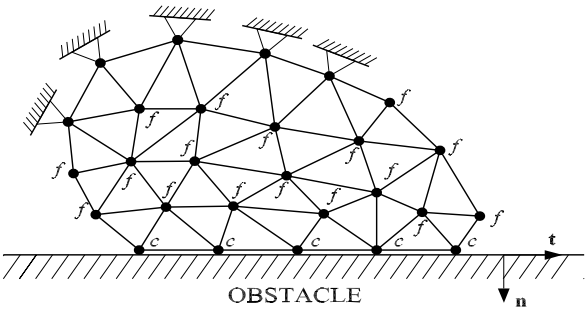

Fig. 1 A finite element discretization of an elastic body in the presence of a rigid obstacle

both with n_c components, grouping the "contact" variables (that must satisfy inequalities and complementarity conditions). One has

$$x = \begin{bmatrix} x_f \\ x_c \end{bmatrix} \in \mathbb{R}^{n_f+n_c}, \quad y = \begin{bmatrix} 0_f \\ y_c \end{bmatrix} \in \mathbb{R}^{n_f+n_c}, \quad 0_c \le x_c \perp y_c \ge 0_c.$$

The above inequalities are to be understood in a componentwise sense. Notice that y_f is a zero-vector since the free nodes cannot have reactions from the obstacle. A necessary and sufficient condition for the occurrence of a so-called *directional instability* is the existence of a nonnegative real number λ and a pair $(x, y) \in \mathbb{R}^{n_f+n_c} \times \mathbb{R}^{n_f+n_c}$, with $x \neq 0$, such that

$$(\lambda M^{\text{mass}} + M^{\text{stiff}})x = y,$$
$$0_c \le x_c \perp y_c \ge 0_c.$$

The effective mass matrix M^{mass} and the effective stiffness matrix M^{stiff} are nonsymmetric in general. We are led to solve a particular case of Problem 1 with $A = M^{\text{stiff}}$, $B = -M^{\text{mass}}$, $K = \mathbb{R}^{n_f} \times \mathbb{R}^{n_c}_+$, and the extra condition that λ must be nonnegative. It is worthwhile mentioning that, in this example, the convex cone K is not pointed.[1]

For the reader's convenience we recall below some general facts concerning the cardinality of K-spectra. The symbol I_n indicates the identity matrix of size $n \times n$ and card$[S]$ stands for the cardinality of a set S in $\mathbb{R}$.

Proposition 1 *Under* (2) *the set* $\sigma_K(A, B)$ *is nonempty and compact. Furthermore,*

(a) $\sigma_K(A, B)$ *has finitely many elements in case* K *is a polyhedral convex cone.*
(b) *for each* $n \ge 3$, *one can find a nonsymmetric matrix* $A \in \mathbb{M}_n$ *and a sequence* $\{K_\nu\}_{\nu\in\mathbb{N}}$ *of polyhedral convex cones in* $\mathbb{R}^n$ *such that* card$[\sigma_{K_\nu}(A, I_n)] \to \infty$ *as* $\nu \to \infty$.
(c) *for each* $n \ge 3$, *one can find a nonsymmetric matrix* $A \in \mathbb{M}_n$ *and a nonpolyhedral convex cone* $K \in \Xi(\mathbb{R}^n)$ *such that* $\sigma_K(A, I_n)$ *contains an interval of positive length* (*which implies, in particular, that* $\sigma_K(A, I_n)$ *is uncountable*).

[1] A closed convex cone K is called pointed if $K \cap -K = \{0\}$ (see, e.g., [4]).

Proof Combine [10, Theorem 3.3] and [18, Theorem 2.5] for the existential part. The compactness result is obvious. See [17, Proposition 6.1] for part (a) and the references [16, 18] for part (c). Part (b) is explained in [18, Proposition 4.6]. □

Already in a polyhedral context one should worry about the size of K-spectra. A polyhedral cone-constrained eigenvalue problem in some Euclidean space of small dimension, say $n = 3$, may lead to a huge K-spectrum, for instance with more than 1 million elements! This strange phenomenon has to do with the facial structure of polyhedral convex cones but we will not elaborate here on this issue. In the non-polyhedral case the situation can be even worse: Iusem and Seeger [7] succeeded in constructing a symmetric matrix A and a nonpolyhedral convex cone K such that $\sigma_K(A, I_n)$ behaves like the Cantor ternary set, i.e., it is uncountable and totally disconnected.

2 The Pareto eigenvalue problem

The Pareto eigenvalue problem is the prototype of a cone-constrained eigenvalue problem. Its precise formulation is as follows (cf. [15, 17]):

Problem 2

$$\begin{aligned} &\text{Find } \lambda \in \mathbb{R} \text{ and a nonzero vector } x \in \mathbb{R}^n \text{ such that} \\ &\mathbb{R}^n_+ \ni x \perp (Ax - \lambda x) \in \mathbb{R}^n_+. \end{aligned} \tag{4}$$

This particular model exhibits already many of the mathematical difficulties arising in the general context of Problem 1. The first challenge that one has to face is a possible exponential growth in the cardinality of

$$\sigma_{\mathbb{R}^n_+}(A) = \{\lambda \in \mathbb{R} : \ (x, \lambda) \text{ solves (4) for some } x \neq 0\}$$

with respect to the dimension n of the underlying Euclidean space. One refers to $\sigma_{\mathbb{R}^n_+}(A)$ as the Pareto spectrum (or set of Pareto eigenvalues) of A.

Proposition 2 *Let $A \in \mathbb{M}_n$. Then,*

(a) $\sigma_{\mathbb{R}^n_+}(P^T A P) = \sigma_{\mathbb{R}^n_+}(A)$ *for any permutation matrix P of size $n \times n$.*
(b) $\sigma_{\mathbb{R}^n_+}(A - \gamma I_n) = \sigma_{\mathbb{R}^n_+}(A) - \gamma$ *for all $\gamma \in \mathbb{R}$.*
(c) $\sigma_{\mathbb{R}^n_+}(\beta A) = \beta \sigma_{\mathbb{R}^n_+}(A)$ *for all $\beta \geq 0$.*

Proof The proof is easy and therefore omitted. □

Remark 1 In contrast to classical eigenvalue analysis, a matrix A and its transpose A^T may have different Pareto spectra. For instance, the Pareto spectra of

$$A = \begin{bmatrix} 8 & -1 \\ 3 & 4 \end{bmatrix} \quad \text{and} \quad A^T = \begin{bmatrix} 8 & 3 \\ -1 & 4 \end{bmatrix}$$

are $\{5, 7, 8\}$ and $\{4\}$, respectively. As one can see, matrix transposition may change even the number of Pareto eigenvalues.

In the sequel $\mathcal{J}(n)$ denotes the collection of all nonempty subsets of $\{1, \ldots, n\}$, the symbol $|J|$ stands for the cardinality of a set J in $\mathcal{J}(n)$, and A^J refers to the principal submatrix of A formed with the rows and columns of A indexed by J.

Lemma 1 (Cf. [17]) *Let $A \in \mathbb{M}_n$. Then, $\lambda \in \mathbb{R}$ is a Pareto eigenvalue of A if and only if there are an index set $J \in \mathcal{J}(n)$ and a vector $\xi \in \mathbb{R}^{|J|}$ such that*

$$A^J \xi = \lambda \xi, \quad \xi \in \operatorname{int}(\mathbb{R}^{|J|}_+), \tag{5}$$

$$\sum_{j \in J} A_{ij} \xi_j \geq 0 \quad \forall i \notin J. \tag{6}$$

In such a case, the vector $x \in \mathbb{R}^n$ defined by

$$x_j = \begin{cases} \xi_j & \text{if } j \in J, \\ 0 & \text{if } j \notin J, \end{cases}$$

is a Pareto eigenvector of A and λ is the corresponding Pareto eigenvalue.

Computing a Pareto spectrum is a much harder problem than computing a usual spectrum. In the first case one has to take into consideration all the possible ways of selecting the index set J.

Example 2 The Pareto eigenvalues of the matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \tag{7}$$

are necessarily in the set $\{a,\ d,\ \lambda_-,\ \lambda_+\}$. Here $\lambda_\pm$ are the usual eigenvalues of A, i.e.,

$$\lambda_\pm = \frac{a+d}{2} \pm \frac{\sqrt{\Delta}}{2} \quad \text{with } \Delta = (a-d)^2 + 4bc.$$

Different cases are possible depending on the sign of the terms $b,\ c,\ a-d,\ \Delta$. We refer to these four terms as the "discriminating factors" of the matrix (7). For instance, when the sign of Δ is negative, the eigenvalues $\lambda_\pm$ are complex numbers and therefore they cannot apply for membership in the Pareto spectrum. Table 1 is constructed by working out all the possible combinations with $bc \geq 0$ and Table 2 summarizes the situation when $bc < 0$. The last column in each table indicates the cardinality of the Pareto spectrum.

A preliminary lesson that can be drawn from Example 2 is this:

$$\begin{cases} \text{if the number of Pareto eigenvalues of the matrix (7) is even, then} \\ \text{at least one of the discriminating factors } \{b,\ c,\ a-d,\ \Delta\} \text{ is equal to zero.} \end{cases} \tag{8}$$

Table 1 The Pareto spectrum of the matrix (7) when $bc \geq 0$

b	c	$a-d$	Pareto spectrum of A	Cardinality
$-$	$-$	Any	λ_-	1
0	0	0	a	1
		$\pm$	a, d	2
$+$	$+$	$\pm$	a, d, λ_+	3
		0	a, λ_+	2
$-$	0	$-$ or 0	a	1
		$+$	a, d	2
0	$-$	$+$ or 0	d	1
		$-$	a, d	2
$+$	0	$+$ or 0	a	1
		$-$	a, d	2
0	$+$	$-$ or 0	d	1
		$+$	a, d	2

Table 2 The Pareto spectrum of the matrix (7) when $bc < 0$

b	c	$a-d$	Δ	Pareto spectrum of A	Cardinality
$-$	$+$	$-$ or 0	Any	a	1
		$+$	$-$	a	1
		$+$	0	a, λ_+	2
		$+$	$+$	a, λ_-, λ_+	3
$+$	$-$	$+$ or 0	Any	d	1
		$-$	$-$	d	1
		$-$	0	d, λ_+	2
		$-$	$+$	d, λ_-, λ_+	3

Roughly speaking, this principle says that a 2×2 matrix is more likely to have an odd number of Pareto eigenvalues. A probabilistic formulation of this statement will be given in Proposition 10. Based on extensive numerical experimentation it is very tempting to conjecture that a principle similar to (8) could be stated for higher dimensional matrices as well. We shall not indulge however on this matter because the task of identifying the appropriate discriminating factors and examining all the sign combinations is simply awful.

2.1 The exponential growth phenomenon

Observe that (5) is a classical eigenvalue problem for the matrix A^J except that now the eigenvectors must satisfy a certain interiority condition. There are

$$r_n = 2^n - 1$$

ways of selecting the index set J and therefore we are led to solve the same number of classical eigenvalue problems. As indicated in [17, Proposition 5.3], the upper bound

$$\text{card}[\sigma_{\mathbb{R}^n_+}(A)] \leq r_n \tag{9}$$

applies to

$$\begin{array}{l} \text{any symmetric matrix } A \text{ of size } n \times n, \\ \text{any matrix } A \in \mathbb{M}_n \text{ whose off-diagonal entries are nonnegative,} \\ \text{any matrix } A \in \mathbb{M}_n \text{ whose off-diagonal entries are nonpositive,} \end{array} \tag{10}$$

and, more generally, to any $A \in \mathbb{M}_n$ such that each principal submatrix A^J has at most one eigenvalue associated to an eigenvector in the interior of $\mathbb{R}^{|J|}_+$. The next proposition shows that the bound (9) is sharp within the first two classes mentioned in (10). This is what we call the exponential growth phenomenon.

Proposition 3 *For each $n \geq 2$, there is an $n \times n$ symmetric matrix A with positive entries such that* $\text{card}[\sigma_{\mathbb{R}^n_+}(A)] = r_n$.

Proof Take $n \geq 2$ and consider the $n \times n$ symmetric matrix A whose general entry is given by $A_{ij} = a^{i+j}$. Here a denotes a real number greater than or equal to $\sqrt{2}$. This is a rare example of a nontrivial matrix whose Pareto spectrum $\sigma_{\mathbb{R}^n_+}(A)$ can be computed explicitly.[2] Given an arbitrary index set $J = \{i_1, \ldots, i_\ell\}$, with $1 \leq i_1 < \cdots < i_\ell \leq n$, the principal submatrix

$$A^J = (M_{j,k})_{1 \leq j,k \leq \ell}$$

has $M_{j,k} = a^{i_j + i_k}$ as entry in the $\{j,k\}$-position. The vector $\xi = (a^{i_1}, \ldots, a^{i_\ell})^T$ belongs to the interior of $\mathbb{R}^{|J|}_+$ and

$$(A^J \xi)_j = \sum_{k=1}^{\ell} M_{j,k} a^{i_k} = \sum_{k=1}^{\ell} a^{i_j + i_k} a^{i_k} = \lambda_J a^{i_j} = \lambda_J \xi_j \quad \forall j \in \{1, \ldots, \ell\}$$

with

$$\lambda_J = \sum_{k=1}^{\ell} a^{2i_k} = \sum_{i \in J} a^{2i}.$$

Since A has positive entries, one does not have to worry about the condition (6). Lemma 1 tells us that λ_J is a Pareto eigenvalue of A. For completing the proof of the proposition we need to check that

$$\lambda_I \neq \lambda_J \quad \text{whenever } I \neq J. \tag{11}$$

[2] We are indebted to Dr. Charki Amara (Avignon) for building this example.

Take $I, J \in \mathcal{J}(n)$ with $I \neq J$. Since $I \Delta J = (I \backslash J) \cup (J \backslash I)$ is nonempty, one can define

$$m = \max\{k \in \{1, \ldots, n\} : k \in I \Delta J\}.$$

Suppose, for instance, that $m \in I$. In such a case, $m \notin J$ and

$$\lambda_I - \lambda_J = \sum_{i \in I} a^{2i} - \sum_{i \in J} a^{2i} = \sum_{i \in I,\, i \leq m} a^{2i} - \sum_{i \in J,\, i \leq m-1} a^{2i}.$$

This implies that $\lambda_I - \lambda_J \geq b^m - \sum_{i=1}^{m-1} b^i$ with $b = a^2$. But

$$b^m - \sum_{i=1}^{m-1} b^i = b^m - \left(\frac{b^m - 1}{b-1} - 1 \right) = \frac{b^{m+1} - 2b^m + b}{b-1}.$$

If $a \geq \sqrt{2}$, then $b \geq 2$ and $b^{m+1} \geq 2b^m$. Hence,

$$\frac{b^{m+1} - 2b^m + b}{b-1} \geq \frac{b}{b-1} \geq \frac{2}{b-1} > 0,$$

showing in this way that $\lambda_I \neq \lambda_J$. Since there are $2^n - 1$ ways of choosing the index set J, there are as many elements in the Pareto spectrum of this special matrix A. □

Remark 2 If $a < \sqrt{2}$, then one cannot guarantee the injectivity condition (11). Consider for instance $a = \sqrt{b}$ with $b = (1 + \sqrt{5})/2 \approx 1.618$. Notice that $I = \{1, 2\}$ and $J = \{3\}$ yield the same Pareto eigenvalue because $\lambda_I = b + b^2 = b^3 = \lambda_J$.

The particular choice $A_{i,j} = (\sqrt{2})^{i+j}$ yields the Pareto spectrum $\sigma_{\mathbb{R}^n_+}(A) = \{2, 4, 6, \ldots, 2^{n+1} - 2\}$ whose cardinality is r_n. Proposition 3 does not mean that one should expect getting systematically such a large number of Pareto eigenvalues. However, "almost all" matrices with positive entries, be them symmetric or not, have a Pareto spectrum with such large cardinality. A measure theoretic justification supporting this statement is given in Sect. 2.2 (cf. Proposition 6).

2.2 Introducing the Perron map

The Perron map is a useful tool for analyzing the Pareto spectrum of a matrix belonging to the class

$$\mathbb{P}_n = \{A \in \mathbb{M}_n : A \text{ is positive}\}. \tag{12}$$

Positivity in (12) is understood in the componentwise sense. The famous Perron theorem asserts that a matrix A in $\mathbb{P}_n$ admits the real number

$$\rho(A) = \text{spectral radius of } A$$

as an algebraically simple eigenvalue. Moreover, $\rho(A)$ is positive and there exists a vector $x \in \text{int}(\mathbb{R}^n_+)$ such that $Ax = \rho(A)x$. Perron's theorem can be applied of course to any principal submatrix A^J.

For computational purposes it is useful to label the index sets $J_1, J_2, \ldots, J_{r_n}$ by using the binary decomposition ordering: one defines

$$J_k = \{j \in \{1, 2, \ldots, n\} : b_j(k) = 1\},$$

where the $\{0, 1\}$-coefficients $b_1(k), b_2(k), \ldots, b_n(k)$ are uniquely determined by the binary expansion

$$k = b_1(k)2^0 + b_2(k)2^1 + \cdots + b_n(k)2^{n-1}$$

of the integer $k \in \{1, 2, \ldots, r_n\}$.

Definition 1 The Perron map on $\mathbb{P}_n$ refers to the vector-valued function $\chi : \mathbb{P}_n \to \mathbb{R}^{r_n}$ whose k-th component $\chi_k : \mathbb{P}_n \to \mathbb{R}$ is given by

$$\chi_k(A) = \text{spectral radius of the principal submatrix } A^{J_k}.$$

The upward Perron map $\chi^{\uparrow} : \mathbb{P}_n \to \mathbb{R}^{r_n}$ is defined by $\chi^{\uparrow}(A) = (\chi_1^{\uparrow}(A), \chi_2^{\uparrow}(A), \ldots, \chi_{r_n}^{\uparrow}(A))$, where the numbers

$$\chi_1^{\uparrow}(A) \le \chi_2^{\uparrow}(A) \le \cdots \le \chi_{r_n}^{\uparrow}(A) \tag{13}$$

are obtained by rearranging in nondecreasing order the components of the vector $\chi(A)$.

Example 3 Consider the matrix $A \in \mathbb{P}_3$ whose general entry is $A_{i,j} = (\sqrt{3})^{i+j}$. As explained in the proof of Proposition 3, for this particular matrix one gets $\chi_k(A) = \sum_{j \in J_k} 3^j$.

As seen in Table 3, the values of $\chi_1(A), \ldots, \chi_7(A)$ are already arranged in increasing order. This is not necessarily the case for an arbitrary $A \in \mathbb{P}_n$. The usefulness of the upward Perron map is obvious. Among other things, one can write

$$\chi_1^{\uparrow}(A) = \text{smallest Pareto eigenvalue of } A, \tag{14}$$

$$\chi_{r_n}^{\uparrow}(A) = \text{largest Pareto eigenvalue of } A, \tag{15}$$

Table 3 The Perron map χ evaluated at the matrix A of Example 3. Here χ has $r_3 = 7$ components

k	Binary expansion of k	$\chi_k(A)$
1	001	3
2	010	9
3	011	12
4	100	27
5	101	30
6	110	36
7	111	39

the full Pareto spectrum of $A \in \mathbb{P}_n$ being given by

$$\sigma_{\mathbb{R}^n_+}(A) = \{\chi_1(A), \chi_2(A), \ldots, \chi_{r_n}(A)\} \tag{16}$$

$$= \{\chi_1^{\uparrow}(A), \chi_2^{\uparrow}(A), \ldots, \chi_{r_n}^{\uparrow}(A)\}. \tag{17}$$

We mention below a useful characterization of the extremal terms (14) and (15).

Proposition 4 *Let $A \in \mathbb{P}_n$. The smallest and the largest Pareto eigenvalues of A are*

$$\chi_1^{\uparrow}(A) = \min\{A_{1,1}, \ldots, A_{n,n}\} \quad \textit{and} \quad \chi_{r_n}^{\uparrow}(A) = \rho(A), \tag{18}$$

respectively. Furthermore,

$$\chi_{k_1}^{\uparrow}(A) < \chi_{k_2}^{\uparrow}(A) < \cdots < \chi_{k_p}^{\uparrow}(A) \tag{19}$$

for any subset $\{k_1, k_2, \ldots, k_p\}$ of $\{1, 2, \ldots, r_n\}$ such that $J_{k_1} \subsetneq J_{k_2} \subsetneq \cdots \subsetneq J_{k_p}$. In particular, $\chi_1^{\uparrow}(A) < \chi_3^{\uparrow}(A) < \chi_7^{\uparrow}(A) < \cdots < \chi_{r_n}^{\uparrow}(A)$ and A has at least n distinct Pareto eigenvalues.

Proof According to [6, Corollary 8.1.20], the spectral radii of the principal submatrices of $A \in \mathbb{P}_n$ obey to the monotonicity principle

$$I, J \in \mathcal{J}(n) \quad \text{and} \quad I \subset J \quad \Longrightarrow \quad \rho(A^I) \leq \rho(A^J). \tag{20}$$

This implication yields immediately the formulas announced in (18). As noticed by Frobenius as early as 1912, the inequality in (20) is strict if I is strictly contained in J. This observation leads directly to the chain of inequalities in (19). In particular, the spectral radii of the leading principal submatrices

$$\begin{bmatrix} A_{1,1} & \cdots & A_{1,q} \\ \vdots & & \vdots \\ A_{q,1} & \cdots & A_{q,q} \end{bmatrix}, \quad q = 1, \ldots, n$$

are arranged in (strictly) increasing order. This takes care of the last part of the proposition. □

The monotonicity principle (20) serves also to estimate the second smallest component and the second largest component of the vector $\chi^{\uparrow}(A)$. Clearly, $\chi_2^{\uparrow}(A)$ and $\chi_{r_n-1}^{\uparrow}(A)$ are to be found in the sets

$$\underbrace{\{A_{1,1}, \ldots, A_{n,n}\}}_{n \text{ terms}} \cup \underbrace{\{\rho(A^J) : |J| = 2\}}_{n(n-1)/2 \text{ terms}} \quad \text{and} \quad \underbrace{\{\rho(A^J) : |J| = n-1\}}_{n \text{ terms}},$$

respectively.

Some of the inequalities in (13) could occur as an equality, and therefore the set (17) could have less that r_n elements. In practice, however, this situation is rare.

A topological justification of this statement is given in Proposition 5. We recall first a couple of continuity results. As one can see from the well known variational formulas

$$\rho(A) = \max_{x \in \mathrm{int}(\mathbb{R}^n_+)} \min_{1 \le i \le n} \frac{1}{x_i} \sum_{j=1}^{n} A_{i,j} x_j \tag{21}$$

$$= \min_{x \in \mathrm{int}(\mathbb{R}^n_+)} \max_{1 \le i \le n} \frac{1}{x_i} \sum_{j=1}^{n} A_{i,j} x_j, \tag{22}$$

the spectral radius function $\rho : \mathbb{P}_n \to \mathbb{R}$ is both upper-semicontinuous and lower-semicontinuous. This yields the continuity of each function

$$A \in \mathbb{P}_n \mapsto \chi_k(A) = \rho(V_k^T A V_k)$$

with V_k denoting a matrix of size $n \times |J_k|$ which does not depend on A. More precisely, in the columns of V_k we stack the canonical vectors $\{e_j\}_{j \in J_k} \subset \mathbb{R}^n$ so that $V_k^T A V_k = A^{J_k}$ for all $A \in \mathbb{P}_n$. The continuity of the Perron map χ yields in turn the continuity of each $\chi_k^{\uparrow}$.

We need to recall also the following two lemmas. By a Perron eigenvector of $A \in \mathbb{P}_n$ one understands a vector $x \in \mathrm{int}(\mathbb{R}^n_+)$ such that $Ax = \rho(A)x$. Such vector x is unique up to normalization.

Lemma 2 *Let $A \in \mathbb{P}_n$. For any $E \in \mathbb{M}_n$, one has*

$$\lim_{t \to 0} \frac{\rho(A + tE) - \rho(A)}{t} = \frac{\langle y, Ex \rangle}{\langle y, x \rangle},$$

where x and y are Perron eigenvectors of A and A^T, respectively. In particular, the partial derivatives of ρ at A are given by

$$\frac{\partial \rho}{\partial A_{i,j}}(A) = \frac{y_i x_j}{\langle y, x \rangle}. \tag{23}$$

Proof This lemma is a particular formulation of [6, Theorem 6.3.12]. □

Formula (23) shows that $\rho(A)$ depends on each entry of the matrix A. More precisely, if an arbitrary entry of A increases, then so does the spectral radius of A.

Lemma 3 *If the index sets $I, J \in \mathcal{J}(n)$ are distinct, then there is no open set in $\mathbb{P}_n$ over which the function*

$$A \in \mathbb{P}_n \mapsto \rho(A^I) - \rho(A^J) \tag{24}$$

is constant.

Proof One of the index sets contains an element which is not in the other. Suppose, for instance, that $\ell \in I \setminus J$. Let us examine the behavior of (24) around a reference matrix, say $A_* \in \mathbb{P}_n$. By Lemma 2, a slight perturbation in the (ℓ, ℓ)-th entry of A_*

modifies the value of $\rho(A_*^I)$. However, the term $\rho(A_*^J)$ remains unchanged because $\ell \notin J$. This argument proves that, in any neighborhood of A_*, the difference (24) is subject to changes. □

Lemma 3 is equivalent to saying that, for any pair of distinct integers $k, \ell \in \{1, 2, \ldots, r_n\}$, there is no open set in $\mathbb{P}_n$ over which the function $\chi_k - \chi_\ell$ is constant. We now are ready to state:

Proposition 5 *For each* $k \in \{1, 2, \ldots, r_n\}$, *let* $\Omega(\mathbb{P}_n, k) = \{A \in \mathbb{P}_n : \text{card}[\sigma_{\mathbb{R}^n_+}(A)] = k\}$. *Then,*

(a) $\Omega(\mathbb{P}_n, r_n)$ *is an open set in* $\mathbb{M}_n$.
(b) $\Omega(\mathbb{P}_n, k)$ *has empty interior in* $\mathbb{M}_n$ *when* $k < r_n$.

Proof That A belongs to $\Omega(\mathbb{P}_n, r_n)$ amounts to saying that

$$A \in \mathbb{P}_n \quad \text{and} \quad \chi_1^{\uparrow}(A) < \chi_2^{\uparrow}(A) < \cdots < \chi_{r_n}^{\uparrow}(A). \tag{25}$$

Since $\mathbb{P}_n$ is an open set in the space $\mathbb{M}_n$, a slight perturbation of A is still in $\mathbb{P}_n$. On the other hand, the inequalities in (25) remain strict after perturbation because the functions $\chi_k^{\uparrow} : \mathbb{P}_n \to \mathbb{R}$ are continuous. This takes care of (a). Now, consider any $k < r_n$. There is no loss of generality in assuming that $k \geq n$ because otherwise $\Omega(\mathbb{P}_n, k)$ is empty by Proposition 4. Let $A \in \mathbb{P}_n$ be a matrix in $\Omega(\mathbb{P}_n, k)$, i.e.,

$$\begin{aligned} \chi_{\psi(j)}(A) &= \chi_{\psi(j+1)}(A) \quad \forall j \in J_{\text{eq}}, \\ \chi_{\psi(j)}(A) &< \chi_{\psi(j+1)}(A) \quad \forall j \in J_{\text{str}} \end{aligned} \tag{26}$$

for a suitable permutation ψ and a suitable partition $J_{\text{eq}} \cup J_{\text{str}} = \{1, 2, \ldots, r_n - 1\}$ with $|J_{\text{str}}| = k - 1$. Consider any $\ell \in J_{\text{eq}}$. In view of Lemma 3, the equality $\chi_{\psi(\ell)}(A) = \chi_{\psi(\ell+1)}(A)$ can be broken by perturbing A. The perturbed matrix, say A', can be taken as near from A as one wishes. Notice that (26) remains in force after perturbation and that $\sigma_{\mathbb{R}^n_+}(A')$ has at least $k + 1$ elements. Summarizing, outside of the set $\Omega(\mathbb{P}_n, k)$, one can find a positive matrix which is arbitrarily close to A. This takes care of (b). □

Remark 3 We mention in passing that the variational formulas (21)–(22) have other uses besides guaranteeing the continuity of ρ. By applying them to any principal submatrix A^J of $A \in \mathbb{P}_n$, one gets

$$\min_{i \in J} \sum_{j \in J} A_{i,j} \leq \rho(A^J) \leq \max_{i \in J} \sum_{j \in J} A_{i,j}.$$

The above sandwich yields easily computable lower and upper bounds for the components $\chi_k(A)$ of the Perron map. A tighter sandwich is obtained by applying Brauer's theorem (cf. [11, Sect. 2.1]).

For each dimension $n \in \{2, 3, \ldots, 6\}$, we performed the numerical experiment which consists in generating randomly a collection of 2000 matrices of size $n \times n$. Each matrix A was generated according to a uniform distribution on the hypercube $[0, 1]^{n\times n}$, i.e., the components $A_{i,j}$, with $i, j \in \{1, \ldots n\}$, were independent random variables following a uniform distribution on the interval $[0, 1]$. We computed the corresponding Pareto-spectra by using the enumerative method suggested by Lemma 1. It turned out that in all cases the maximal cardinality r_n was attained! If one takes into account the following proposition, the outcome of this experiment is not surprising altogether.

Proposition 6 *Let $\mathbb{M}_n$ be equipped with a probability measure P that is absolutely continuous with respect to the $n \times n$-dimensional Lebesgue measure. If P is concentrated on*

$$\mathrm{cl}(\mathbb{P}_n) = \{A \in \mathbb{M}_n : A \textit{ is nonnegative}\}, \tag{27}$$

then

$$P(\Omega(\mathbb{P}_n, k)) = \begin{cases} 1 & \textit{if } k = r_n, \\ 0 & \textit{if } k < r_n. \end{cases}$$

In particular, if $X \in \mathbb{M}_n$ is a random matrix with uniform distribution on the hypercube $[0, 1]^{n\times n}$, then $\mathrm{Prob}[\mathrm{card}[\sigma_{\mathbb{R}^n_+}(X)] = r_n] = 1$.

Proof The boundary of $\mathbb{P}_n$ has zero Lebesgue measure in $\mathbb{M}_n$. So, that P is absolutely continuous and concentrated on (27) amounts to saying that P is expressible in the form

$$P(\Xi) = \int_{\Xi\cap\mathbb{P}_n} f(A) d\mu_{n\times n}(A).$$

Here μ_m denotes the m - dimensional Lebesgue measure and $f : \mathbb{M}_n \to \mathbb{R}$ refers to the density function of P, i.e., the Radon-Nikodym derivative of P with respect to $\mu_{n\times n}$. Let $k < r_n$. We shall prove that $\Omega(\mathbb{P}_n, k)$ has zero Lebesgue measure in $\mathbb{M}_n$. For this we rely not just on the continuity of the spectral radius function ρ, but on a stronger property, namely, its analyticity on $\mathbb{P}_n$. Recall that a function of several real variables is called analytic at a point if in a neighborhood of this point it has a power series expansion. The analyticity of ρ at a given $A_0 \in \mathbb{P}_n$ can be shown, for instance, by using a general analyticity result of functions involving eigenvalues (cf. [5]). Alternatively, one can apply the "real" version of Theorem 2 in [1]. It is important to observe that for all A near A_0, the term $\rho(A)$ is a simple eigenvalue of A. That $\rho : \mathbb{P}_n \to \mathbb{R}$ is analytic implies in turn that each component of the Perron map is analytic. Note that if A belongs to $\Omega(\mathbb{P}_n, k)$, then at least two of the terms in (16) coincide. Thus,

$$\Omega(\mathbb{P}_n, k) \subset \bigcup_{\psi\in\mathrm{Perm}(r_n)} \{A \in \mathbb{P}_n : \chi_{\psi(2)}(A) - \chi_{\psi(1)}(A) = 0\}$$

with $\mathrm{Perm}(m)$ denoting the set of permutations on $\{1, 2, \ldots, m\}$. Observe that each set in the above union has zero Lebesgue measure in $\mathbb{M}_n$ because each function

$\chi_{\psi(2)} - \chi_{\psi(1)}$ is nonconstant and analytic on $\mathbb{P}_n$. This proves that $\Omega(\mathbb{P}_n, k)$ has zero Lebesgue measure in $\mathbb{M}_n$ and yields the equality $P(\Omega(\mathbb{P}_n, k)) = 0$. For the second part of the proposition we take f as the uniform density function on $[0, 1]^{n\times n}$. □

2.3 The Pareto capacity of $\mathbb{M}_n$

The case of a nonsymmetric matrix $A \in \mathbb{M}_n$ with possibly negative entries is more involved. According to Lemma 1, there are still r_n classical eigenvalue problems to be solved, but now each problem may yield several solutions, i.e., several Pareto eigenvalues. In fact, for a general $A \in \mathbb{M}_n$, the cardinality of $\sigma_{\mathbb{R}^n_+}(A)$ may go beyond the bound $r_n = 2^n - 1$ considered in (9). Let us illustrate this point with a low dimensional example.

Example 4 Consider a 3×3 matrix of the form

$$A = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} = \begin{bmatrix} 8 & -1 & \gamma \\ 3 & 4 & \varepsilon \\ \nu & -\delta & 6 \end{bmatrix} \tag{28}$$

where $\gamma, \varepsilon, \nu, \delta$ are positive parameters. The particular submatrix

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} 8 & -1 \\ 3 & 4 \end{bmatrix} \tag{29}$$

has 3 Pareto eigenvalues and this is the most one can get for a 2×2 matrix. The Pareto spectrum of (29) is $\{5, 7, 8\}$. We are allowing A to have exactly 2 negatives entries, both in the same column. This choice is dictated by the wish of complying to the sign-constraint (6) for as many index sets J as possible. A particular case of (28) with 9 Pareto eigenvalues is displayed in Table 4. For ease of visualization, the Pareto eigenvalues are being arranged in increasing order. Note that the index set J_3 produces two Pareto eigenvalues. The same remark applies to J_6 and J_7. No Pareto eigenvalue is produced by J_2.

Table 4 Pareto spectrum of the matrix (28) when $\gamma = 4$, $\varepsilon = 1/2$, $\nu = 2$ and $\delta = 1/2$

λ	x_1	x_2	x_3	y_1	y_2	y_3	Index set
4.133975	0	1	0.267949	0.071797	0	0	J_6
4.602084	0.185257	1	0.092628	0	0	0	J_7
5	0.333333	1	0	0	0	0.166667	J_3
5.866025	0	0.267949	1	3.732051	0	0	J_6
6	0	0	1	4	0.5	0	J_4
7	1	1	0	0	0	1.5	J_3
8	1	0	0	0	3	2	J_1
9.397916	1	0.602084	0.5	0	0	0	J_7
10	1	0	0.5	0	3.25	0	J_5

The next lemma is a slight improvement with respect to [17, Proposition 5.2]. This new upper bound for the cardinality of $\sigma_{\mathbb{R}^n_+}(A)$ will be further improved in Proposition 7.

Lemma 4 *A matrix of size $n \times n$ has at most $n2^{n-1} - (n-1)$ Pareto eigenvalues.*

Proof Consider an arbitrary $A \in \mathbb{M}_n$. We suppose that A has at least one negative off-diagonal entry, otherwise we are done because $2^n - 1 \leq n2^{n-1} - (n-1)$. As a preliminary upper estimate for the cardinality of $\sigma_{\mathbb{R}^n_+}(A)$ one may consider

$$\operatorname{card}[\sigma_{\mathbb{R}^n_+}(A)] \leq n2^{n-1}. \tag{30}$$

The bound (30) was suggested in [17] and follows easily from the fact that each principal submatrix A^J has at most $|J|$ real eigenvalues. Notice that

$$n2^{n-1} = \sum_{J \in \mathcal{J}(n)} |J| = \sum_{d=1}^{n} d\, C_d^n \quad \text{with } C_d^n = \frac{n!}{d!(n-d)!}.$$

For sharpening the bound (30) we must take into account the nonnegativity condition (6). Let $j_1, \ldots, j_p$ indicate the columns of A containing a negative off-diagonal entry, i.e.,

$$\{j_1, \ldots, j_p\} = \{j \in \{1, \ldots, n\} : A_{i,j} < 0 \text{ for some } i \neq j\}.$$

Two cases must be distinguished:
Case $p \in \{n-1, n\}$. We take away the p diagonal terms $\{A_{j,j} : j \in \{j_1, \ldots, j_p\}\}$ from the Pareto spectrum of A. These diagonal terms where counted as Pareto eigenvalues, but in fact they are not because the nonnegativity condition (6) is violated when $J = \{j_k\}$ with $k \in \{1, \ldots, p\}$. Hence,

$$\operatorname{card}[\sigma_{\mathbb{R}^n_+}(A)] \leq n2^{n-1} - p \leq n2^{n-1} - (n-1).$$

Case $p \in \{1, \ldots, n-2\}$. Besides the p diagonal terms considered above, we must drop a few additional candidates for membership in $\sigma_{\mathbb{R}^n_+}(A)$. If one considers the index set $J_+ = \{1, \ldots, n\} \backslash \{j_1, \ldots, j_p\}$, then one gets a corresponding matrix A^{J_+} whose off-diagonal entries are nonnegative. The $(n-p) \times (n-p)$ matrix A^{J_+} produces at most one Pareto eigenvalue and not $|J_+| = n - p$ as counted before. So, we must substract also $n - p - 1$ candidates from our initial estimate $n2^{n-1}$. Again, one ends up with

$$n2^{n-1} - (n-p-1) - p = n2^{n-1} - (n-1)$$

remaining candidates. □

The main merit of the bound derived in Lemma 4 is its simplicity. The bound proposed in the next proposition is sharper but it requires evaluating a more involved expression, namely, the discrete Fenchel conjugate

$$n \mapsto \varphi^*(n) = \sup_{p \in \mathbb{N}} \{np - \varphi(p)\}$$

of a certain function $\varphi : \mathbb{N} \to \mathbb{N}$ of integer variable.

Proposition 7 *A matrix of size $n \times n$ has at most*

$$q_n = n2^{n-1} - \frac{n(n-1)}{2} + \varphi^*(n)$$

Pareto eigenvalues. Here $\varphi^* : \mathbb{N} \to \mathbb{N}$ stands for the discrete Fenchel conjugate of

$$p \mapsto \varphi(p) = p\,2^{p-1} + \frac{p(p+1)}{2}. \tag{31}$$

Given the special form (31) of the function φ, the supremum in the definition of φ^* could equally be taken just over $\{1, \ldots, n-1\}$.

Proof A matrix $A \in \mathbb{M}_n$ has C_2^n principal submatrices of size 2×2, i.e., of the form

$$A^{\{k,l\}} = \begin{bmatrix} A_{k,k} & A_{k,\ell} \\ A_{\ell,k} & A_{\ell,\ell} \end{bmatrix}.$$

Let m denote the number of matrices $A^{\{k,l\}}$ satisfying the sign condition

$$A_{k,\ell} A_{\ell,k} < 0. \tag{32}$$

Each of the remaining $C_2^n - m$ matrices $A^{\{k,l\}}$ produces at most one Pareto eigenvalue and not two as considered in (30). We must therefore substract the number $C_2^n - m$ from the upper bound (30). But this is not all. Each of the m matrices satisfying (32) contains a negative off-diagonal entry and this fact eliminates additional candidates for membership in the Pareto spectrum. Let p denote the number of columns containing at least one of these m negative off-diagonal entries. Let us examine the different possibilities depending on the size of m:

- $1 \le m \le n-1$. In order to loose as few candidates as possible we place the m negative off-diagonal entries in the same column, say, the first one. We have $p = 1$ and only one additional candidate is lost.
- $n \le m \le (n-1) + (n-2)$. The next group of $m - (n-1)$ negative off-diagonal entries are to be placed in another column, say, the second one. They must be placed below the diagonal entry. This time $p = 2$ and we lose 4 additional candidates because the first two columns of A contain a negative element on the same row.
- $1 + \sum_{s=1}^{p-1}(n-s) \le m \le \sum_{s=1}^{p}(n-s)$. This corresponds to the general case and the most conservative configuration looks like in the following matrix

$$A = \begin{bmatrix} * & + & + & + & + & + \\ - & * & + & + & + & + \\ - & - & * & + & + & * \\ - & - & - & * & * & * \\ - & - & - & * & * & * \\ - & - & * & * & * & * \end{bmatrix}.$$

No Pareto-eigenvalue is produced by the upper-left block of size $p \times p$, neither by the principal submatrices contained in this block. This time $p2^{p-1}$ additional candidates are lost.

The analysis of the general case leads to minimize the total loss $p2^{p-1} + C_2^n - m$ with respect to all integers $m \geq 1$ and $p \geq 1$ satisfying

$$1 + \sum_{s=1}^{p-1}(n-s) \leq m \leq \sum_{s=1}^{p}(n-s) \leq C_2^n. \tag{33}$$

The best strategy is taking p as small as possible and m as large as possible. The second inequality in (33) becomes active at the optimum and therefore

$$\begin{aligned} \operatorname{card}[\sigma_{\mathbb{R}^n_+}(A)] &\leq n2^{n-1} - \min_p \left\{ p2^{p-1} + C_2^n - \sum_{s=1}^{p}(n-s) \right\}, \\ &= n2^{n-1} - C_2^n + \max_p \left\{ \sum_{s=1}^{p}(n-s) - p\,2^{p-1} \right\} \end{aligned}$$

where optimization is carried out with respect to $p \geq 1$ such that $\sum_{s=1}^{p}(n-s) \leq C_2^n$. The latter constraint on p amount to saying that $p \leq n-1$. It is not difficult to see that such constraint on p is superfluous. In fact, the optimal integer p is very small while compared to n. A matter of simplification completes the proof of the proposition. □

Computing $\varphi^*(n)$ is not so difficult after all. Given the special structure of φ, everything boils down to finding the largest element in a collection of $n-1$ integers. If the dimension n is big, then the simplest way of computing $\varphi^*(n)$ is by finding the unique root of the nonlinear equation

$$\big[1 + p\ln(2)\big]2^{p-1} + p = n. \tag{34}$$

The term on the left-hand side of (34) corresponds to the derivative of (31) when viewed as a function of a real variable. If p_n denotes such root, then the supremum in the definition of $\varphi^*(n)$ is attained at the lower integer part of p_n or at the upper integer part of p_n. On the other hand, one can check that

$$\frac{n(n-1)}{2} - \varphi^*(n) \geq n-1, \tag{35}$$

so Proposition 7 is stronger than Lemma 4. The difference between both sides on (35) is negligible for small values of n but it becomes more relevant as n increases.

We mention in passing that the sign condition (32) alone does not guarantee that $A^{\{k,\ell\}}$ produces two Pareto eigenvalues. This minor technical point is clarified next.

Proposition 8 *Let $A \in \mathbb{M}_n$. The 2×2 submatrix $A^{\{k,\ell\}}$ produces exactly two Pareto eigenvalues of A if and only if the following conditions are in force*:

(i) $A_{k,\ell}A_{\ell,k} < 0$.
(ii) $(A_{k,k} - A_{\ell,\ell})^2 + 4A_{k,\ell}A_{\ell,k} > 0$.
(iii) $(A_{\ell,\ell} - A_{k,k} \pm \sqrt{(A_{k,k} - A_{\ell,\ell})^2 + 4A_{k,\ell}A_{\ell,k}})/A_{k,\ell} > 0$.

(iv) *For all $i \notin \{k, \ell\}$, one has*

$$A_{i,k} + \frac{A_{i,\ell}}{2A_{k,\ell}}\big(A_{\ell,\ell} - A_{k,k} \pm \sqrt{(A_{k,k} - A_{\ell,\ell})^2 + 4A_{k,\ell}A_{\ell,k}}\big) \geq 0.$$

Proof This is a matter of working out Lemma 1 for the index set $J = \{k, \ell\}$. Condition (ii) expresses the fact that $A^{\{k,\ell\}}$ has two distinct real eigenvalues. The inequalities in (i) and (iii) say that the corresponding eigenvectors can be taken with positive components. Finally, the inequalities in (iv) take care of (6). □

As one sees from Proposition 8, a matrix $A^{\{k,\ell\}}$ must comply to plenty of sign restrictions in order to produce two Pareto eigenvalues of A. The analysis of the higher dimensional principal submatrices of A is too complicated to be treated by hand.

Definition 2 The Pareto capacity of the space $\mathbb{M}_n$ is understood as the maximal number

$$\pi_n = \max_{A \in \mathbb{M}_n} \text{card}[\sigma_{\mathbb{R}^n_+}(A)] \tag{36}$$

of Pareto eigenvalues that a matrix in $\mathbb{M}_n$ can achieve.

In a similar way one can define the Pareto capacity of any subset of $\mathbb{M}_n$. Evaluating π_n and finding a matrix $A \in \mathbb{M}_n$ that achieves the supremum in (36) is not an easy matter.

Proposition 9 *The first two Pareto capacities are $\pi_1 = 1$ and $\pi_2 = 3$. For all integers $k, \ell \geq 1$, one can write $\pi_k + \pi_\ell \leq \pi_{k+\ell}$. In particular, $\{\pi_n\}_{n\geq 1}$ is an increasing sequence.*

Proof That $\pi_1 = 1$ and $\pi_2 = 3$ is clear. Suppose that $C \in \mathbb{M}_k$ achieves the Pareto capacity of the space $\mathbb{M}_k$ and that $D \in \mathbb{M}_\ell$ achieves the Pareto capacity of the space $\mathbb{M}_\ell$. The Pareto spectrum of the partitioned matrix

$$\begin{bmatrix} C & 0 \\ 0 & D \end{bmatrix} \in \mathbb{M}_{k+\ell}$$

contains the π_k Pareto eigenvalues of C as well as the π_ℓ Pareto eigenvalues of D. There is no loss of generality in assuming that $\sigma_{\mathbb{R}^k_+}(C) \cap \sigma_{\mathbb{R}^\ell_+}(D) = \emptyset$, otherwise we pick up a constant γ large enough and change D by the matrix $D + \gamma I_\ell$. We have proven in this way that $\pi_{k+\ell}$ is greater than or equal to $\pi_k + \pi_\ell$. □

The sequence $\{\pi_n\}_{n\geq 1}$ increases with respect to n in an exponential way. Indeed, in view of Propositions 3 and 7, the term π_n is sandwiched as follows:

$$r_n \leq \pi_n \leq q_n. \tag{37}$$

Table 5 Bounds (37) for the Pareto capacity of spaces $\mathbb{M}_n$ for $n \in \{2, \ldots, 6\}$

n	r_n	π_n	q_n	$n2^{n-1} - (n-1)$
2	3	Exactly 3	3	3
3	7	9 or 10	10	10
4	15	At least 17, at most 27	28	29
5	31	At most 71	73	76
6	63	Not examined	182	187

Table 5 gives the numerical values of these bounds when n ranges from 2 to 6. The upper bound q_n is sharp when $n = 2$. For larger values of n there is room for improvement, but obtaining a sharper and still easily computable upper bound is a tough job. In fact, it is difficult to see which is the sign pattern that produces a matrix achieving the Pareto capacity of the space $\mathbb{M}_n$.

The Pareto capacity of $\mathbb{M}_3$ is either 9 or 10. We have not found yet a 3×3 matrix with 10 Pareto eigenvalues. Although we seriously doubt that such a matrix exists, at this point in time we cannot discard such a possibility. An example of 4×4 matrix with 17 Pareto eigenvalues is

$$A = \begin{bmatrix} 34 & -61 & 58 & 58 \\ 30 & -63 & 10 & 9 \\ 98 & -83 & 45 & 74 \\ 99 & -84 & 46 & 44 \end{bmatrix}.$$

A 4×4 matrix with more than 17 Pareto eigenvalues is likely to exist but we do not have yet experimental evidence of this fact. That more than 27 Pareto eigenvalues is impossible follows by examining carefully all the possible sign patterns.

2.4 Expected number of Pareto eigenvalues

Encountering a matrix $A \in \mathbb{M}_n$ that possesses as much as π_n Pareto eigenvalues is a sort of worst scenario situation. Extensive computational testing suggests that the expected (or average) value of $\text{card}[\sigma_{\mathbb{R}^n_+}(A)]$ grows at most linearly with respect to n. The information provided by Table 6 has been obtained as follows. For each dimension $n \in \{2, 3, \ldots, 10\}$, one generates randomly a sample of 10000 matrices of size $n \times n$, each matrix following a uniform probability distribution over the hypercube $[-1, 1]^{n \times n}$. For each randomly generated matrix A one solves Problem 2 by using Lemma 1 and one counts the number of elements in $\sigma_{\mathbb{R}^n_+}(A)$.

Observe that only odd numbers of Pareto eigenvalues are showing up in Table 6. The following proposition gives a probabilistic justification of the obtained results for the dimension $n = 2$.

Proposition 10 *Let $X \in \mathbb{M}_2$ be a random matrix uniformly distributed over the hypercube $[-1, 1]^{2\times 2}$. Then, $Z = \text{card}[\sigma_{\mathbb{R}^2_+}(X)]$ is a discrete random variable with*

Table 6 Cardinality counting for the Pareto spectra of 10000 randomly generated matrices with possibly negative entries. The last row shows the expected cardinality for each $n \in \{2, 3, \ldots, 10\}$

♯ sols.	Dimension n								
	2	3	4	5	6	7	8	9	10
1	6592	5786	5456	5259	5336	5460	5311	5395	5354
3	3408	2981	2686	2492	2322	2149	2220	2098	2074
5	–	875	1003	1049	958	912	881	908	923
7	–	358	526	542	546	532	515	468	511
9	–	–	210	268	302	296	291	302	287
11	–	–	79	176	182	175	184	195	181
13	–	–	33	90	119	115	144	121	150
15	–	–	7	51	57	88	93	90	105
17	–	–	–	29	52	62	68	76	65
19	–	–	–	26	45	50	54	66	59
21	–	–	–	14	20	36	44	44	44
23	–	–	–	4	24	30	38	33	35
25	–	–	–	–	13	19	24	30	20
27	–	–	–	–	7	11	23	23	36
29	–	–	–	–	6	12	15	16	11
31	–	–	–	–	3	16	13	12	14
33 or more	–	–	–	–	8	37	82	123	131
Average	1.682	2.161	2.550	2.943	3.183	3.408	3.751	3.966	4.105

weight coefficients

$$p_k = \text{Prob}[Z = k] = \begin{cases} 95/144 & \text{if } k = 1, \\ 0 & \text{if } k = 2, \\ 49/144 & \text{if } k = 3. \end{cases}$$

The expected number of Pareto eigenvalues of X is $E[Z] = \sum_{k=1}^{3} k p_k = 121/72 \approx 1.68$.

Proof If $X \in \mathbb{M}_2$ is a random matrix distributed according to a density function $f : \mathbb{M}_2 \to \mathbb{R}$, then

$$p_k = \int_{\Omega_k} f(A) d\mu_{2\times 2}(A) \tag{38}$$

with $\Omega_k = \{A \in \mathbb{M}_2 : \text{card}[\sigma_{\mathbb{R}^2_+}(A)] = k\}$. If f is the uniform density on the hypercube $[-1, 1]^{2\times 2}$, then the integral (38) becomes

$$p_k = 2^{-4}\, \mu_{2\times 2}(\Omega_k \cap [-1, 1]^{2\times 2}).$$

We now recall the information provided by Tables 1 and 2. The region $\Omega_2 \cap [-1, 1]^{2\times 2}$ is negligible with respect to the measure $\mu_{2\times 2}$ as a consequence of the

principle stated in (8). This shows that $p_2 = 0$. For evaluating p_3 we must compute the 2×2-dimensional volume of the region $\Omega_3 \cap [-1, 1]^{2\times 2}$. This amounts to evaluating the 4-dimensional volume of the set $\widehat{\Omega}_3$ of all vectors $(a, b, c, d) \in [-1, 1]^4$ satisfying one of the following mutually exclusive conditions:

$$b > 0, \quad c > 0, \tag{39}$$

$$b < 0, \quad c > 0, \quad a - d > 0, \quad (a - d)^2 + 4bc > 0, \tag{40}$$

$$b > 0, \quad c < 0, \quad a - d < 0, \quad (a - d)^2 + 4bc > 0. \tag{41}$$

A matter of iterated integration shows that the portion (39) contributes with 4 units to the volume of $\Omega_3 \cap [-1, 1]^{2\times 2}$. Integration over (40) produces

$$\int_{-1}^{0}\left[\int_{0}^{1}\left[\int_{\substack{a-d>\sqrt{-4bc}\\ -1\le a,d\le 1}} d\mu_2(a, d)\right] d\mu_1(c)\right] d\mu_1(b)$$

$$= \int_{-1}^{0}\left[\int_{0}^{1} 2(1 - \sqrt{-bc})^2 d\mu_1(c)\right] d\mu_1(b) = \frac{13}{18}$$

additional units. The contribution of the portion (41) is also $13/18$ as one can see by exchanging the roles of b and c, as well as the roles of a and d. In short,

$$p_3 = \frac{1}{16}\left[4 + \frac{13}{18} + \frac{13}{18}\right] = \frac{49}{144} \approx 0.34 \quad \text{and} \quad p_1 = 1 - \frac{49}{144} = \frac{95}{144} \approx 0.66.$$

This completes the proof of the proposition. □

It is worth mentioning that the integral (38) serves to compute the weight coefficients p_k even if the random matrix X does not follow a uniform distribution. One can consider, in fact, any absolutely continuous law, i.e., any probability law admitting a density function with respect to the Lebesgue measure. Although the weight coefficients p_1 and p_3 depend on f, the coefficient p_2 does not. In other words, no matter which is the density function that is being employed, a 2×2 random matrix has 2 Pareto eigenvalues with probability zero.

The values announced by Proposition 10 are consistent with those appearing in the second column of Table 6. A probabilistic justification of the experimental data for higher dimensional matrices could be developed along the same lines, but such a task would necessarily be cumbersome and tedious.

2.4.1 Asymptotic behavior

Table 6 is a statistical sample giving a hint on how $\text{card}[\sigma_{\mathbb{R}^n_+}(A)]$ behaves when $A \in \mathbb{M}_n$ is generated according to a uniform probability distribution as explained before. Allowing negatives entries in A has as consequence the violation of the condition (6) on a large number of occasions. The later fact partially explains why the expected value of $\text{card}[\sigma_{\mathbb{R}^n_+}(A)]$ does not grow exponentially. The lack of symmetry in the random matrices $A \in \mathbb{M}_n$, and in their corresponding principal submatrices, is

another reason explaining the slow growth of the expected value of card$[\sigma_{\mathbb{R}^n_+}(A)]$. Indeed, nonsymmetric matrices usually have a big proportion of complex eigenvalues, which, of course, do not count as Pareto eigenvalues.

In an interesting paper of 1994, Edelman et al. [3] prove that if E_n^{normal} denotes the expected number of real eigenvalues of an $n \times n$ matrix whose entries are independent random variables with standard normal distributions, then

$$\lim_{n\to\infty} \frac{E_n^{\text{normal}}}{\sqrt{n}} = \sqrt{\frac{2}{\pi}},$$

i.e., E_n^{normal} behaves like $\sqrt{2n/\pi}$ for large n. Numerical experiments with matrices whose entries are independent random variables uniformly distributed on $[-1, 1]$ are also reported in [3]. The asymptotic behavior of E_n^{uniform} does not differ significantly with respect to the normally distributed case. Additional information on this topic can be found in [2].

Let $\mathcal{E}_n^{\text{uniform}}$ denotes the expected number of Pareto eigenvalues of an $n \times n$ matrix whose entries are independent random variables with uniform distribution on $[-1, 1]$. The last row in Table 6 suggests the existence of a constant c such that

$$\mathcal{E}_n^{\text{uniform}} \approx c\sqrt{n} \quad \text{for large } n.$$

Additional numerical testing reported in Table 7 confirms this asymptotic behavior, but we do not have yet a formal proof. Table 7 contains also information on the behavior of $\mathcal{E}_n^{\text{normal}}$, the expected number of Pareto eigenvalues of an $n \times n$ matrix whose entries are independent random variables with standard normal distributions.[3]

2.5 Spectral histograms

Among a large set of randomly generated matrices of size 3×3, one case with 9 Pareto eigenvalues was detected:

$$A = \begin{bmatrix} 0.347338 & -0.612421 & 0.583729 \\ 0.308260 & -0.629640 & 0.073888 \\ 0.985363 & -0.832666 & 0.453387 \end{bmatrix}. \tag{42}$$

Up to a scalar multiplication by 100, the matrix

$$A = \begin{bmatrix} 34 & -61 & 58 \\ 30 & -63 & 10 \\ 98 & -83 & 45 \end{bmatrix} \tag{43}$$

is not far from (42) and admits 9 Pareto eigenvalues as well. The details are displayed in Table 8. There are some similarities with the case mentioned in Example 4, but there are also some differences. For instance, the index set J_7 produces now 3 Pareto eigenvalues. This is a very productive index set indeed.

[3] IST cluster took 8 hours to produce the entry of Table 7 relative to $n = 20$ (normal distribution) by using 40 processors in parallel.

Table 7 Expected cardinality of the Pareto spectrum of a random matrix with possibly negative entries, computed from a sample of 10000 randomly generated matrices. Uniform distributions and standard normal distributions are considered

Expected cardinality	Dimension n									
	2	3	4	5	6	7	8	9	10	
$\mathcal{E}_n^{\text{unif}}$	1.682	2.161	2.550	2.943	3.183	3.408	3.751	3.966	4.105	
$\frac{\mathcal{E}_n^{\text{unif}}}{\sqrt{n}}$	1.189	1.248	1.275	1.316	1.299	1.288	1.326	1.322	1.298	
$\mathcal{E}_n^{\text{normal}}$	1.707	2.217	2.641	3.031	3.349	3.465	3.736	4.090	4.283	
$\frac{\mathcal{E}_n^{\text{normal}}}{\sqrt{n}}$	1.207	1.280	1.320	1.356	1.367	1.310	1.321	1.363	1.354	
Expected cardinality	Dimension n									
	11	12	13	14	15	16	17	18	19	20
$\mathcal{E}_n^{\text{unif}}$	4.389	4.716	4.756	4.607	4.788	4.978	5.142	5.652	5.808	5.745
$\frac{\mathcal{E}_n^{\text{unif}}}{\sqrt{n}}$	1.323	1.362	1.319	1.231	1.236	1.245	1.247	1.332	1.332	1.285
$\mathcal{E}_n^{\text{normal}}$	4.484	4.481	4.597	4.624	4.901	5.178	5.469	5.306	5.553	5.712
$\frac{\mathcal{E}_n^{\text{normal}}}{\sqrt{n}}$	1.352	1.294	1.275	1.236	1.266	1.295	1.327	1.251	1.274	1.277

Table 8 Pareto spectrum of the matrix (43)

λ	x_1	x_2	x_3	y_1	y_2	y_3	Index set
−44.590788	0.479759	1	0.401644	0	0	0	J_7
−38.166419	0.815397	1	0.037168	0	0	0	J_7
−37.352790	0.854907	1	0	0	0	0.780887	J_3
8.352790	1	0.420446	0	0	0	63.102976	J_3
34	1	0	0	0	30	98	J_1
36.672749	0	0.100328	1	51.879972	0	0	J_6
45	0	0	1	58	10	0	J_4
98.757208	0.712877	0.194034	1	0	0	0	J_7
115.092658	0.715231	0	1	0	31.456936	0	J_5

Two matrices may possess the same number of Pareto eigenvalues, but not for the same reasons. For instance, in one case the Pareto eigenvalues could be produced by the low dimensional principal submatrices, and in the other case the production of Pareto eigenvalues could be mainly due to the high dimensional principal submatrices.

Definition 3 The spectral histogram of $A \in \mathbb{M}_n$ is the vector $h(A) = (h_1(A), h_2(A), \ldots, h_{r_n}(A))$ whose k-th component indicates the number of Pareto eigenvalues of A produced by J_k. The aggregated spectral histogram of A is the vector $H(A) =$

$(H_1(A), H_2(A), \ldots, H_n(A))$ defined by

$$H_d(A) = \sum_{k \text{ s.t. } |J_k|=d} h_k(A) \quad \forall d \in \{1, \ldots, n\},$$

that is, $H_d(A)$ indicates the number of Pareto eigenvalues produced by all the $d \times d$ principal submatrices of A.

For instance, the spectral histograms of the matrices associated with Tables 4 and 8 are

$$(1, 0, 2, 1, 1, 2, 2),$$
$$(1, 0, 2, 1, 1, 1, 3),$$

respectively. Their aggregated versions are $(2, 5, 2)$ and $(2, 4, 3)$, respectively. In general, one has

$$\text{card}[\sigma_{\mathbb{R}^n_+}(A)] = \sum_{k=1}^{r_n} h_k(A) = \sum_{d=1}^{n} H_d(A) \quad \forall A \in \mathbb{M}_n.$$

The function $H : \mathbb{M}_n \to \mathbb{N}^n$ obeys to a number of interesting rules among which one can mention:

Proposition 11 *Let $A \in \mathbb{M}_n$. Then,*

(a) $H(P^T A P) = H(A)$ *for any permutation matrix P of size $n \times n$.*
(b) $H(A - \gamma I_n) = H(A) - \gamma$ *for all $\gamma \in \mathbb{R}$.*
(c) $H(\beta A) = \beta H(A)$ *for all $\beta \geq 0$.*
(d) $H_d(A) \leq d\, C_d^n$ *for all $d \in \{1, \ldots, n\}$. Simultaneous attainment of these bounds is impossible.*
(e) $H_1(A) = n$ *if and only if $A_{i,j} \geq 0$ for all $i, j \in \{1, \ldots, n\}$, $i \neq j$.*
(f) $H_n(A) = n$ *if and only if A is spectrally saturated in the sense that it admits n distinct real eigenvalues with corresponding eigenvectors in the interior of $\mathbb{R}^n_+$. This is yet equivalent to saying that*

$$A = [\lambda_1 u^1, \ldots, \lambda_n u^n]\, U^{-1} \tag{44}$$

for some vector $\lambda \in \mathbb{R}^n$ with distinct entries and some positive matrix $U = [u^1, \ldots, u^n]$ with $\det(U) = 1$.

Proof For proving (a) we identify the index set J_k with the k-th face of the Pareto cone $\mathbb{R}^n_+$, i.e., with the subcone

$$F_k = \{x \in \mathbb{R}^n_+ : x_j = 0,\ \forall j \notin J_k\}.$$

The image of a face under a permutation matrix is another face of the same dimension. This key observation leads easily to the invariance property stated in (a). The other

statements of the proposition are more or less straightforward, except possibly for the last part of (f). Saying that A is spectrally saturated means that one can find nonzero vectors $u^1, \ldots, u^n$ in the interior of $\mathbb{R}^n_+$ and distinct real numbers $\lambda_1, \ldots, \lambda_n$ such that $Au^i = \lambda_i u^i$ for all $i \in \{1, \ldots, n\}$. Written in full extent, these equations are

$$\sum_{j=1}^{n} A_{i,j} u^i_j = \lambda_i u^i_j \quad \forall i, j \in \{1, \ldots, n\}. \tag{45}$$

We look at (45) as a linear system of n^2 equations in which the unknown variables are the entries of A. Notice that $U = [u^1, \ldots, u^n]$ is a nonsingular matrix with positive entries. By renumbering the vectors u^i if necessary, one may suppose that the determinant of U is positive. In such case, one can take $\det(U) = 1$ by invoking a simple homogeneity argument. The system (45) admits then as unique solution the matrix A given by (44). □

3 Numerical methods

3.1 Scaling-and-projection algorithm

Recall that a normalizing function for a closed convex cone K is any continuous function $\phi : K \to \mathbb{R}$ such that

- $\phi(x) > 0$ for all nonzero vector $x \in K$,
- $\phi(tx) = t\phi(x)$ for all $t > 0$ and $x \in K$, (46)
- $K_\phi = \{x \in K : \phi(x) = 1\}$ is compact.

One says that $x \in K$ is a ϕ-normalized vector if $\phi(x) = 1$. A closed convex cone in a finite dimensional space admits always a normalizing function, think for instance of

$$\phi(x) = \|x\|, \tag{47}$$

$$\phi(x) = \sqrt{\pm\langle x, Bx\rangle}, \tag{48}$$

where the sign in (48) depends on whether the quadratic form $x \mapsto \langle x, Bx\rangle$ is positive or negative on K. Another interesting normalizing function is

$$\phi(x) = \langle e, x\rangle \quad \text{with } e \in \operatorname{int}(K^+), \tag{49}$$

but this choice makes sense only if K is pointed.

Below we introduce and discuss the Scaling-and-Projection Algorithm (SPA) whose aim is finding a solution to

Problem 3

Find $\lambda \in \mathbb{R}$ and vectors $x, y \in \mathbb{R}^n$ such that

$$(A - \lambda B)x = y,$$

$$K \ni x \perp y \in K^+, \tag{50}$$

$$\phi(x) = 1. \tag{51}$$

This is just another way of writing Problem 1 except that now one has added the normalization condition (51). The complementarity system (50) simply means that $x \in K$, $y \in K^+$, and $\langle x, y\rangle = 0$. This system remains obviously unchanged if y is multiplied by a positive scalar:

$$x \in K, \quad sy \in K^+, \quad \langle x, sy\rangle = 0. \tag{52}$$

The parameter $s > 0$ is interpreted as a scaling factor that puts the primal vector x and the dual (or residual) vector y in a right balance. This is the "scaling" part of the algorithm. As we shall see latter, the choice of the scaling parameter does have an importance when it comes to convergence issues. In view of Moreau's orthogonal decomposition theorem [12], one can write (52) in the equivalent form

$$x = \Pi_K[x - sy]$$

with $\Pi_K(z)$ denoting the element from K at smallest distance from z. This is the "projection" part of the algorithm. The following lemma will be used in the sequel.

Lemma 5 *Let $K \in \Xi(\mathbb{R}^n)$. If y is orthogonal to $x \in K\backslash\{0\}$, then the function $s \in \mathbb{R} \mapsto \Pi_K[x - sy]$ never vanishes.*

Proof Suppose, on the contrary, that $\Pi_K[x - \hat{s}y] = 0$ for some $\hat{s} \in \mathbb{R}$. In such a case the vector $\hat{s}y - x$ is in the dual cone K^+ and therefore

$$0 \leq \langle x, \hat{s}y - x\rangle = -\|x\|^2,$$

contradicting that x is a nonzero vector. □

3.1.1 Description of the SPA

The SPA generates a sequence $\{x^t\}_{t\geq 0}$ lying in the compact set K_ϕ and a bounded sequence $\{\lambda_t\}_{t\geq 0}$ in $\mathbb{R}$. It generates also a bounded sequence $\{y^t\}_{t\geq 0}$ of residual vectors.

- *Initialization*: Take any nonzero vector u in K and define

$$x^0 = \frac{u}{\phi(u)}.$$

- *Iteration*: One has a current point x^t in K_ϕ. Compute

$$\lambda_t = \frac{\langle x^t, Ax^t\rangle}{\langle x^t, Bx^t\rangle} \quad \text{and} \quad y^t = Ax^t - \lambda_t Bx^t.$$

By construction, y^t is orthogonal to x^t. Select any step-size or scaling factor $s_t > 0$ and compute the projection

$$v^t = \Pi_K[x^t - s_t y^t]. \tag{53}$$

By Lemma 5 one knows that $v^t \neq 0$. The projection v^t belongs to K but it is not necessarily a ϕ-normalized vector. Proceed then to a ϕ-normalization

$$x^{t+1} = \frac{1}{\phi(v^t)} v^t. \tag{54}$$

Remark 4 If at the t-th iteration one has $y^t = 0$, then one should stop. There is no need to continue because $x^{t+1} = x^t$ and a solution has been found. Indeed, (x^t, λ_t) is a classical solution (i.e. the residual term y^t vanishes) for the pair (A, B) with x^t being a nonzero vector in K.

Computing v^t is a matter of projecting onto the cone K and this could be hard if K has a complicated structure. There are however plenty of interesting cones for which the projection map admits an explicit and easily computable formula. This is true, for instance, for the Pareto cone or positive orthant, for the Loewner cone of positive semidefinite symmetric matrices, for the Lorentz or ice-cream cone and, more generally, for any revolution cone.

Selecting the scaling factor is perhaps the most delicate part concerning the use of the SPA. Is there an "optimal" way of selecting the scaling factor s_t? This question is too complex to be settled properly at this point in time. Anyway, it seems natural asking the sequence $\{s_t\}_{t\geq 0}$ to be bounded. Not only that, it would be convenient to have convergence toward some positive number. Some possible options for consideration are:

I. *The Constant Rule*. One chooses $s_t = s$ for all $t \geq 0$, with s denoting a positive constant.
II. *The Blind Rule*. One considers a sequence $\{s_t\}_{t\geq 0}$ converging to some positive scalar s.
III. *The Feedback Rule*. One uses a scaling factor $s_t = \Psi(x^t, \lambda_t, y^t)$ that depends on the current information at stage t.

3.1.2 A convergence result

As mentioned before, the sequence $\{x^t\}_{t\geq 0}$ generated by the SPA remains in the compact set K_ϕ.

Theorem 1 *Suppose that the SPA is implemented with the Blind Rule* (*or with any other rule that guarantees convergence of scaling factors toward a positive scalar*). *Assume convergence of* $\{x^t\}_{t\geq 0}$ *toward some limit that one denotes by* $\bar{x}$. *Then*,

$$\{\lambda_t\}_{t\geq 0} \to \bar{\lambda} = \frac{\langle \bar{x}, A\bar{x}\rangle}{\langle \bar{x}, B\bar{x}\rangle}, \tag{55}$$

$$\{y^t\}_{t\geq 0} \to \bar{y} = A\bar{x} - \bar{\lambda}B\bar{x}, \tag{56}$$

and $(\bar{x}, \bar{\lambda}, \bar{y})$ *is a solution to Problem* 3.

Proof The limit $\bar{x}$ is necessarily in K_ϕ. The conclusions (55) and (56) are immediate. Observe, incidentally, that $\bar{y}$ is orthogonal to $\bar{x}$. Plugging (53) into (54) one gets

$$x^{t+1} = \frac{1}{\phi(\Pi_K[x^t - s_t y^t])}\,\Pi_K[x^t - s_t y^t]. \tag{57}$$

The projection mapping Π_K and the normalizing function ϕ being continuous, one can pass to the limit in (57). One obtains in this way

$$\bar{r}\bar{x} = \Pi_K[\bar{x} - s\bar{y}] \tag{58}$$

with s denoting the limit of $\{s_t\}_{t\geq 0}$ and $\bar{r} = \phi(\Pi_K[\bar{x} - s\bar{y}])$ being a positive scalar in view of Lemma 5 and the property (46) of the normalizing function ϕ. In order to complete the proof one must check that $\bar{r} = 1$. To do this one rewrites (58) as

$$\bar{r}\bar{x} = \Pi_K[\bar{r}\bar{x} - (s\bar{y} - \bar{x} + \bar{r}\bar{x})].$$

Moreau's orthogonal decomposition theorem yields

$$s\bar{y} - \bar{x} + \bar{r}\bar{x} \in K^+,$$
$$\langle \bar{r}\bar{x}, s\bar{y} - \bar{x} + \bar{r}\bar{x}\rangle = 0.$$

Since $\bar{x} \perp \bar{y}$, the last equality reduces to $\bar{r}\,(\bar{r} - 1)\langle \bar{x}, \bar{x}\rangle = 0$. So, $\bar{r} = 1$ as needed. □

The convergence assumption on $\{x^t\}_{t\geq 0}$ is essential to make sure that one can pass to the limit in (57). Suppose, on the contrary, that $\bar{x}$ is just a cluster point and not a limit point of $\{x^t\}_{t\geq 0}$. This means that $\bar{x} = \lim_{t\to\infty} x^{\varphi(t)}$ for some strictly increasing function $\varphi : \mathbb{N} \to \mathbb{N}$. It is very tempting trying to pass to the limit on both sides of

$$x^{\varphi(t)+1} = \frac{1}{\phi(\Pi_K[x^{\varphi(t)} - s_{\varphi(t)} y^{\varphi(t)}])}\,\Pi_K[x^{\varphi(t)} - s_{\varphi(t)} y^{\varphi(t)}] \tag{59}$$

but the term on the left-hand side may not converge. Taking a new subsequence if necessary, one may assume that $x^{\varphi(t)+1}$ does converge. The problem now is that $\hat{x} = \lim_{t\to\infty} x^{\varphi(t)+1}$ may be different from $\bar{x}$. After passing to the limit in (59) one arrives at $\bar{r}\hat{x} = \Pi_K[\bar{x} - s\bar{y}]$. This time one gets

$$s\bar{y} - \bar{x} + \bar{r}\hat{x} \in K^+,$$
$$\langle \bar{r}\hat{x}, s\bar{y} - \bar{x} + \bar{r}\hat{x}\rangle = 0,$$

but there is no way of guaranteeing that $\bar{r}\hat{x} - \bar{x} = 0$. In short, what we are saying is that a cluster point $(\bar{x}, \bar{\lambda}, \bar{y})$ produced by the SPA is not necessarily a solution to Problem 3. In view of this observation, it is crucial to choose $\{s_t\}_{t\geq 0}$ so as to guarantee the convergence of $\{x^t\}_{t\geq 0}$.

3.1.3 The SPA as fixed point algorithm

The SPA implemented with a constant scaling factor $s > 0$ corresponds in fact to a fixed point algorithm

$$x^{t+1} = F_s(x^t) \tag{60}$$

for the nonlinear operator $F_s : K_\phi \to K_\phi$ given by

$$F_s(x) = \frac{1}{\phi(\Pi_K[x - sy])} \Pi_K[x - sy],$$
$$y = Ax - \frac{\langle x, Ax\rangle}{\langle x, Bx\rangle} Bx.$$

Proposition 12 *Suppose that K is pointed. If ϕ is any of the normalizing functions (47), (48), (49), then the nonlinear operator $F_s : K_\phi \to K_\phi$ satisfies the relation*

$$\|F_s(x') - F_s(x)\| \leq L_s \|x' - x\| \quad \forall x', x \in K_\phi$$

for a suitable Lipschitz constant L_s.

Proof The compactness of K_ϕ is essential at several stages. First of all, it ensures the Lipschitzness of the map

$$x \in K_\phi \mapsto Ax - \frac{\langle x, Ax\rangle}{\langle x, Bx\rangle} Bx.$$

The projection map $\Pi_K : \mathbb{R}^n \to \mathbb{R}^n$ being nonexpansive, it follows that

$$x \in K_\phi \mapsto G_s(x) = \Pi_K \left[x - s\left(Ax - \frac{\langle x, Ax\rangle}{\langle x, Bx\rangle} Bx \right) \right]$$

is Lipschitz as well. The Lipschitz constant of G_s depends of course on s. The map G_s transforms the compact set K_ϕ into another compact set, namely $G_s(K_\phi) = \{G_s(x) : x \in K_\phi\}$. By Lemma 5 one knows that $G_s(K_\phi)$ does not contain the origin of $\mathbb{R}^n$. For completing the proof one just needs to observe that

$$v \in G_s(K_\phi) \mapsto \frac{v}{\phi(v)}$$

is a Lipschitz map because $\phi(v)$ remains away from 0 when v ranges over $G_s(K_\phi)$. □

The pointedness assumption in Proposition 12 has been introduced just to make sure that one can use the linear form $x \mapsto \phi(x) = \langle e, x\rangle$ as normalizing function. The advantage of this choice is that K_ϕ is not just compact, but also convex. Since $F_s : K_\phi \to K_\phi$ is continuous, the Schauder fixed point theorem (cf. [19, Theorem 1.26]) tells us that F_s has at least one fixed point on K_ϕ. By the way, this proves that Problem 1 is solvable when K is pointed.

3.2 The convexified scaling-and-projection algorithm (CSPA)

Applying the SPA is a matter of performing fixed point iterations of the Picard type (60) on the nonlinear operator F_s. In view of Proposition 12, it is natural to look for a scaling factor s that renders the Lipschitz constant

$$L_s = \sup_{\substack{x',x\in K_\phi \\ x'\neq x}} \frac{\|F_s(x') - F_s(x)\|}{\|x' - x\|}$$

as small as possible. Obtaining $L_s < 1$ is out of the question because the Banach Contraction Principle would imply a unique fixed point. Recall that we are dealing with eigenvalue problems that have several solutions in general.

If the operator F_s were nonexpansive, i.e., $L_s = 1$, then the Krasnoselskii iteration scheme

$$x^{t+1} = (1-\alpha)x^t + \alpha F_s(x^t) \tag{61}$$

would converge to a fixed point of F_s provided the parameter α is taken in the open interval $]0, 1[$. Historically speaking, Krasnoselskii [9] suggested the midpoint $\alpha = 1/2$ for his iteration scheme, but this is not always the best choice. Note that x^{t+1} is obtained as convex combination of x^t and $F_s(x^t)$. If ϕ is the linear normalizing function (49), then K_ϕ is convex and x^{t+1} remains in K_ϕ.

By writing (61) in a more explicit form, one gets the iteration model

$$\begin{aligned}
\lambda_t &= \frac{\langle x^t, Ax^t\rangle}{\langle x^t, Bx^t\rangle}, \\
y^t &= Ax^t - \lambda_t Bx^t, \\
v^t &= \Pi_K[x^t - sy^t], \\
x^{t+1} &= (1-\alpha)x^t + \alpha\frac{1}{\phi(v^t)}v^t.
\end{aligned}$$

This is what we call the CSPA. Besides the choice of the initial point u, the CSPA leaves open the possibility of playing with two parameters: the scaling factor s and the convexity coefficient α.

Proposition 13 *Let K be pointed and ϕ be the linear normalizing function (49). In case of convergence, the CSPA produces a solution to Problem 3.*

Proof The CSPA is similar to the SPA, except that (60) has been changed by (61). Observe that the new operator $(1-\alpha)I_n + \alpha F_s$ has the same fixed points as F_s. □

Table 9 Solution set of the Pareto eigenproblem associated with matrix (62)

Sol. name	λ	x_1	x_2	x_3	y_1	y_2	y_3
S_1	4.133975	0	1	0.267949	0.071797	0	0
S_2	4.602084	0.185257	1	0.092628	0	0	0
S_3	5	0.333333	1	0	0	0	0.166667
S_4	5.866025	0	0.267949	1	3.732051	0	0
S_5	6	0	0	1	4	0.5	0
S_6	7	1	1	0	0	0	1.5
S_7	8	1	0	0	0	3	2
S_8	9.397916	1	0.602084	0.5	0	0	0
S_9	10	1	0	0.5	0	3.25	0

3.2.1 Numerical experience with the CSPA

By way of illustration, we apply the CSPA to the Pareto eigenproblem associated with the matrix

$$A = \begin{bmatrix} 8 & -1 & 4 \\ 3 & 4 & 1/2 \\ 2 & -1/2 & 6 \end{bmatrix} \tag{62}$$

whose Pareto spectrum is shown in Table 9. This matrix has 9 Pareto eigenvalues in all.

Table 10 displays the outcome of the CSPA when implemented with the linear normalizing function $\phi(x) = x_1 + x_2 + x_3$ and initialized[4] at $u = (0, 1, 0)$. The notation S_j/m indicates that the solution S_j was found within m iterations and the symbol ∞ indicates that convergence did not occur within 2000 iterations.

Table 10 suggests somehow that the CSPA has better chances of convergence if either the scaling factor or the Krasnoselskii parameter is small. In fact, it is very tempting to conjecture that the convergence of the CSPA depends on the product $s\alpha$. In Table 10 one sees that convergence occurs if the product $s\alpha$ ranges between 0.04 and 0.3.

4 By way of conclusion

In this work we have studied some theoretical aspects concerning the so-called cone-constrained eigenvalue problem. Special attention was paid to the Paretian case. It has been observed that the number of solutions to Problem 2 can grow exponentially with respect to the dimension n of the underlying Euclidean space. The Pareto spectrum of an $n \times n$ matrix can be computed explicitly by solving $2^n - 1$ classical eigenvalue

[4] We initialized the CSPA also at $u = (0, 1, 1)$ and got an array more or less similar to that of Table 10. The detected Pareto eigenvalues depend of course on the initial point.

Table 10 Influence of the scaling parameter s and the Krasnoselskii parameter α on the detected solution and on the number of iterations required by the CSPA to achieve convergence. The matrix under consideration is (62) and the initial vector is $u = (0, 1, 0)$

s	Krasnoselskii parameter α								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	∞	$S_2/158$	$S_2/106$	$S_2/79$	$S_2/62$	$S_2/51$	$S_2/43$	$S_2/37$	$S_2/32$
0.2	$S_2/160$	$S_2/80$	$S_2/52$	$S_2/37$	$S_2/28$	$S_2/22$	$S_2/17$	$S_2/14$	$S_2/10$
0.3	$S_2/107$	$S_2/52$	$S_2/33$	$S_2/22$	$S_2/16$	$S_2/11$	$S_2/5$	$S_2/11$	$S_2/17$
0.4	$S_2/81$	$S_2/38$	$S_2/23$	$S_2/14$	$S_2/8$	$S_2/11$	$S_2/18$	$S_2/31$	$S_2/59$
0.5	$S_2/64$	$S_2/29$	$S_2/16$	$S_2/8$	$S_2/11$	$S_2/23$	$S_2/49$	$S_3/184$	$S_3/169$
0.6	$S_2/53$	$S_2/23$	$S_2/11$	$S_2/11$	$S_2/22$	$S_2/58$	$S_1/192$	∞	∞
0.7	$S_2/45$	$S_2/19$	$S_2/6$	$S_2/17$	$S_2/47$	$S_3/193$	∞	∞	∞
0.8	$S_2/39$	$S_2/15$	$S_2/11$	$S_2/24$	$S_1/198$	∞	∞	∞	∞
0.9	$S_2/34$	$S_2/12$	$S_2/16$	$S_2/53$	∞	∞	∞	∞	∞
1.0	$S_2/30$	$S_2/9$	$S_2/23$	$S_1/209$	∞	∞	∞	∞	∞
2.0	$S_2/11$	$S_1/246$	∞	∞	∞	∞	∞	∞	∞
3.0	$S_2/23$	∞	∞	∞	∞	∞	∞	∞	$S_5/24$
4.0	$S_2/315$	∞	∞	∞	∞	∞	∞	∞	$S_7/23$
5.0	$S_3/1205$	∞	∞	∞	∞	∞	∞	$S_7/91$	$S_7/23$
8.1	$S_1/445$	∞	∞	∞	∞	∞	∞	$S_5/25$	$S_7/21$
8.2	$S_3/466$	∞	∞	∞	∞	∞	∞	$S_7/24$	$S_9/10$
8.3	$S_3/429$	∞	∞	∞	∞	∞	∞	$S_7/24$	$S_5/21$

problems of different sizes (cf. [17]). For $n \geq 10$, such an explicit method is too time consuming and alternative search techniques are to be sought.

In Sect. 3 we introduced and studied the SPA. In case of convergence, this algorithm generates a sequence $\{(x^t, \lambda_t, y^t)\}_t \geq 0$ leading to a solution (x, λ, y) to Problem 3. A variant of the SPA called CSPA was also studied. The latter algorithm behaves similarly to the former one except that it offers the possibility of choosing an additional adjustment parameter.

The SPA and the CSPA can be applied to cone-constrained eigenvalue problems involving an arbitrary closed convex cone K, and not just the Pareto cone as in the explicit method of [17].

The SPA and the CSPA are useful methods only if we are not aiming at finding all the solutions to Problem 3. This is not just because the number of solutions could be very large, but also because some solutions could be extremely hard to be detected. It happens in practice that some solutions are hard to be found even if these algorithms are initialized at nearby initial points. The existence of such "rare" solutions is somewhat intrinsic to the nature of the cone-constrained eigenvalue problem. An interesting open question is understanding why some solutions are so hard to detect and others are not. An appropriate theoretical analysis remains to be done.

Acknowledgements A. Pinto da Costa would like to thank Profs. Carlos Varelas da Rocha, Miguel Casquilho and Moitinho de Almeida (respectively from Departamento de Matemática, Engenharia Química e Biológica and Engenharia Civil e Arquitectura of IST) for their kind assistance in guiding

him to the algorithm for the selection of the principal submatrices of a square matrix, in the verification of Table 6 by another random number generator and in the parallel implementation of some programs developed during this study.

References

1. Chu, K.W.E.: On multiple eigenvalues of matrices depending on several parameters. SIAM J. Numer. Anal. **27**, 1368–1385 (1990)
2. Edelman, A., Kostlan, E.: How many zeros of a random polynomial are real? Bull. Am. Math. Soc. **32**, 1–37 (1995)
3. Edelman, A., Kostlan, E., Shub, M.: How many eigenvalues of a random matrix are real? J. Am. Math. Soc. **7**, 247–267 (1994)
4. Facchinei, F., Pang, J.S.: Finite-Dimensional Variational Inequalities and Complementarity Problems, vol. I. Springer, New York (2003)
5. Fan, M.K.H., Tsing, N.K., Verriest, E.I.: On analyticity of functions involving eigenvalues. Linear Algebra Appl. **207**, 159–180 (1994)
6. Horn, R.A., Johnson, C.R.: Matrix Analysis. Cambridge Univ. Press, Cambridge (1985)
7. Iusem, A., Seeger, A.: On convex cones with infinitely many critical angles. Optimization **56**, 115–128 (2007)
8. Júdice, J.J., Sherali, H.D., Ribeiro, I.M.: The eigenvalue complementarity problem. Comput. Optim. Appl. **37**, 139–156 (2007)
9. Krasnoselskii, M.A.: Two remarks on the method of successive approximations. Usp. Mat. Nauk **10**, 123–127 (1955) (in Russian)
10. Lavilledieu, P., Seeger, A.: Existence de valeurs propres pour les systèmes multivoques: résultats anciens et nouveaux. Ann. Sci. Math. Que. **25**, 47–70 (2000)
11. Minc, H.: Nonnegative Matrices. Wiley-Interscience, New York (1988)
12. Moreau, J.J.: Décomposition orthogonale d'un espace hilbertien selon deux cones mutuellement polaires. C. R. Acad. Sci. Paris **255**, 238–240 (1962)
13. Pinto da Costa, A., Figueiredo, I.N., Júdice, J.A., Martins, J.A.C.: A complementarity eigenproblem in the stability of finite dimensional elastic systems with frictional contact. In: Ferris, M., Mangasarian, O., Pang, J.S. (eds.) Complementarity: Applications, Algorithms and Extensions. Applied Optimization Series, vol. 50, pp. 67–83. Kluwer Academic, Dordrecht (1999)
14. Pinto da Costa, A., Martins, J.A.C., Figueiredo, I.N., Júdice, J.J.: The directional instability problem in systems with frictional contacts. Comput. Methods Appl. Mech. Eng. **193**, 357–384 (2004)
15. Queiroz, M., Júdice, J., Humes, C.: The symmetric eigenvalue complementarity problem. Math. Comput. **73**, 1849–1863 (2004)
16. Quittner, P.: Spectral analysis of variational inequalities. Comment. Math. Univ. Carol. **27**, 605–629 (1986)
17. Seeger, A.: Eigenvalue analysis of equilibrium processes defined by linear complementarity conditions. Linear Algebra Appl. **292**, 1–14 (1999)
18. Seeger, A., Torki, M.: On eigenvalues induced by a cone constraint. Linear Algebra Appl. **372**, 181–206 (2003)
19. Singh, S., Watson, B., Srivastava, P.: Fixed Point Theory and Best Approximation: The KKM-Map Principle. Kluwer Academic, Dordrecht (1997)