J Glob Optim (2009) 44:1–28
DOI 10.1007/s10898-007-9225-2

Local minima of quadratic forms on convex cones

Alberto Seeger · Mounir Torki

Received: 15 June 2007 / Accepted: 8 August 2007 / Published online: 7 September 2007

Abstract We study the local minima and the critical values of a quadratic form on the trace of a convex cone. This variational problem leads to the development of a spectral theory that combines matrix algebra and facial analysis of convex cones.

Keywords Convex cones · Conic quadratic programming · Linear complementarity problems · Cone-constrained eigenvectors

Mathematics Subject Classifications 15A18 · 58C40

1 Introduction

1.1 Formulation of the problem and aim of this work

The Euclidean space $\mathbb{R}^n$ is equipped with the inner product $\langle u, v\rangle = u^T v$ and the associated norm $\|\cdot\|$. The dimension n is assumed to be greater than or equal to 2. The symbol $\mathbb{S}_n$ refers to the unit sphere in $\mathbb{R}^n$. We also use the notation

$$\begin{aligned} \mathrm{Sym}(n) &\equiv \text{symmetric (real) matrices of size } n \times n, \\ \Xi(\mathbb{R}^n) &\equiv \text{closed convex cones in } \mathbb{R}^n. \end{aligned}$$

We are concerned with the minimization of a quadratic form over the trace

$$K \cap \mathbb{S}_n = \{u \in K : \|u\| = 1\}$$

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

M. Torki
University of Avignon, I.U.P., 339 chemin des Meinajaries, Avignon 84911, France
e-mail: mounir.torki@univ-avignon.fr

that a cone K leaves on the unit sphere. More precisely, we want to identify the critical values and the local solutions to the variational problem

$$\begin{aligned} &\text{minimize } \langle u, Au\rangle \\ &u \in K \cap \mathbb{S}_n. \end{aligned} \tag{1}$$

One assumes that $A \in \mathrm{Sym}(n)$ and that $K \in \Xi(\mathbb{R}^n)$ is non-trivial in the sense that it contains at least one non-zero vector. Recall that $x \in \mathbb{R}^n$ is called a local solution to (1) if $x \in K \cap \mathbb{S}_n$ and there exists a neighborhood $\mathcal{N}$ of x such that $\langle x, Ax\rangle \leq \langle u, Au\rangle$ for all $u \in K \cap \mathbb{S}_n \cap \mathcal{N}$. The concept of critical value will be clarified in a moment.

Minimizing a quadratic form over the trace of a convex cone seems at first sight a narrow concern, but such particular type of variational problem arises in many practical situations. Interesting examples may be found, for instance, in the angular analysis of convex cones [6–8] and in the modeling of elastic mechanical systems with frictionless contacts [9] or with associated frictional contacts (prescribed normal reactions from the obstacle, a particular case of [10,11]). The variational problem (1) is also of interest in a Hilbert space setting [5,13], but we stick to finite dimensionality because in that way the connection to matrix spectral analysis is more transparent. The normalization constraint $\|u\| = 1$ appearing in (1) could be changed by something more general like $\langle u, Bu\rangle = 1$, but this would only obscure the presentation of our ideas.

The aim of the present work is threefold:

- To obtain upper bounds for the number of local minimal values of the variational problem (1). We shall discuss how these bounds depend on the geometric structure of the convex cone K.
- To explain why, under polyhedrality, it is possible to convert (1) into a finite collection of subspace-constrained eigenvalue problems. From this collection of eigenvalue problems, we shall draw all sort of information on the original variational problem (1). For instance, we shall derive some localization results for the local minimal values.
- To formulate a sufficient criterion for local minimality which is simple and computationally implementable. We also want to estimate how large is the neighborhood on which local minimality takes place.

1.2 Necessary optimality conditions

First and second-order necessary optimality conditions for the variational problem (1) are stated in the next theorem. In the sequel the notation

$$K^+ = \{y \in \mathbb{R}^n : \ \langle y, u\rangle \geq 0 \ \ \forall u \in K\}$$

refers to the dual cone of K, the symbol $y^\perp$ indicates the hyperplane orthogonal to $y \in \mathbb{R}^n$, and "cl" stands for topological closure.

Theorem 1 *Let $A \in \mathrm{Sym}(n)$ and $K \in \Xi(\mathbb{R}^n)$. Let $x \in \mathbb{R}^n$ be a local solution to (1). Then, x is a critical vector of (1) in the sense that*

$$x \in K \cap \mathbb{S}_n \quad \text{and} \quad Ax - \langle x, Ax\rangle x \in K^+. \tag{2}$$

Furthermore, x satisfies the second-order optimality condition

$$\langle h, [A - \langle x, Ax\rangle I]\, h\rangle \geq 0 \ \ \forall h \in \mathcal{C}_K(x), \tag{3}$$

where $\mathcal{C}_K(x) = \mathrm{cl}\{[Ax - \langle x, Ax\rangle x]^\perp \cap \mathbb{R}_+(K - x)\}$ is a non-trivial closed convex cone in $\mathbb{R}^n$.

Proof We pick up a displacement direction h in $K - x$ and examine the behavior of the quadratic form

$$u \in \mathbb{R}^n \mapsto q_A(u) = \langle u, Au \rangle$$

over the curve $\psi : [0, \varepsilon] \to \mathbb{R}^n$ given by

$$\psi(t) = \frac{x + th}{\|x + th\|} \,. \tag{4}$$

We ask $\varepsilon \in]0, 1]$ to be small enough so that the denominator in (4) doesn't vanish. Note that $\psi(t)$ corresponds to the normalization of $\gamma(t) = x + th \in K$. Hence, ψ is an admissible curve emanating from x in the sense that $\psi(0) = x$ and $\psi(t) \in K \cap \mathbb{S}_n$ for all $t \in [0, \varepsilon]$. Since x is a local solution to (1), the choice $t = 0$ yields a local minimum for the univariate function

$$t \in [0, \varepsilon] \mapsto g(t) = q_A(\psi(t)) = \frac{\langle \gamma(t), A\gamma(t) \rangle}{\|\gamma(t)\|^2},$$

and therefore the right-derivative

$$g'(0) = 2\ \langle Ax - \langle x, Ax \rangle x, h \rangle$$

is non-negative. But $h \in K - x$ is arbitrary, so the vector $y = Ax - \langle x, Ax \rangle x$ must be in the dual cone of K. This takes care of (2). For obtaining (3) we rely on the second-order Maclaurin expansion

$$g(t) = g(0) + tg'(0) + \frac{1}{2}t^2 g''(0) + t^2\delta(t),$$

where $\delta(t) \to 0$ as $t \to 0^+$. If the displacement direction $h \in K - x$ is orthogonal to y, then $g'(0) = 0$ and the second-order right-derivative

$$g''(0) = 2\ \langle h, [A - \langle x, Ax \rangle I]\, h \rangle$$

is non-negative. This proves that

$$\langle h, [A - \langle x, Ax \rangle I]\, h \rangle \geq 0 \quad \forall h \in y^\perp \cap (K - x).$$

The above inequality can be extended to $h \in y^\perp \cap \mathbb{R}_+(K - x)$ by using a positive homogeneity argument, and then to $h \in \mathcal{C}_K(x)$ by using a continuity argument. That $\mathcal{C}_K(x)$ is a non-trivial closed convex cone in $\mathbb{R}^n$ is clear. □

The above proof of Theorem 1 relies on the technique of admissible curves. As alternative proof method one could consider a Lagrange multiplier approach. Due to the special structure of the feasible set we don't have to worry here about constraint qualification assumptions.

It is not difficult to see that the second-order optimality condition (3) can be written in the equivalent form

$$\langle h, Ah \rangle \geq \langle x, Ax \rangle \quad \forall h \in \mathcal{C}_K(x),\ \|h\| = 1. \tag{5}$$

Since x is a unit vector belonging to $\mathcal{C}_K(x)$, the inequality (5) amounts to saying that x is a global solution to the minimization problem

$$\begin{aligned} &\text{minimize } \langle h, Ah \rangle \\ &h \in \mathcal{C}_K(x) \cap \mathbb{S}_n. \end{aligned} \tag{6}$$

By the way, if K were an arbitrary closed convex set, then (6) could be seen as a "conical linearization" around x of the original variational problem (1). Up to some technical details, this is in essence the conical linearization technique employed in [3].

We now fix the basic terminology that is employed in this paper. First of all we introduce the set of all critical values of the variational problem (1), that is to say,

$$\sigma(A, K) = \{\langle x, Ax\rangle : x \text{ is a critical vector of (1)}\}.$$

The above set is simply called the K-spectrum of A because

$$\lambda \in \sigma(A, K) \Longleftrightarrow \begin{cases} \text{there is a non-zero vector } x \in \mathbb{R}^n \text{ such that} \\ x \in K,\ Ax - \lambda x \in K^+,\ \langle x, Ax - \lambda x\rangle = 0. \end{cases} \tag{7}$$

Sometimes one refers to $\sigma(A, K)$ as the set of K-eigenvalues of the matrix A. In the same vein, a non-zero vector x as in (7) is said to be a K-eigenvector of A. The later terminology speaks by itself and doesn't need further justification. Just to make everything clear, we point out that

$$x \text{ is a critical vector of (1)} \iff x \text{ is a normalized } K\text{-eigenvector of } A.$$

The right-hand side of the equivalence (7) is used in [15] as the definition of the K-spectrum of an arbitrary matrix, be it symmetric or not. In this work, however, we stick to the symmetric case.

The K-spectrum of a symmetric matrix A is to be distinguished from the set of all local minimal values of (1), that is to say,

$$\sigma_{\text{locmin}}(A, K) = \{\langle x, Ax\rangle : x \text{ is a local solution to (1)}\}.$$

We are specially interested in the later set because we are concerned with the computation of local minima and not just with the identification of critical vectors. Recall that we are solving a minimization problem after all. In general one has the inclusion

$$\sigma_{\text{locmin}}(A, K) \subset \sigma(A, K)$$

and both sets contain the global minimal value

$$\lambda_{\min}(A, K) = \min_{u \in K \cap \mathbb{S}_n} \langle u, Au\rangle$$

of the variational problem (1). Moreover,

$$\lambda_{\min}(A, K) = \min\{\lambda : \lambda \in \sigma_{\text{locmin}}(A, K)\} = \min\{\lambda : \lambda \in \sigma(A, K)\}. \tag{8}$$

1.3 Dualization

Recall that a symmetric matrix E is said to be K-copositive if $\langle u, Eu\rangle \geq 0$ for all $u \in K$. In an n-dimensional context, the set of all such matrices is given by

$$\mathcal{P}_K = \{E \in \text{Sym}(n) : \lambda_{\min}(E, K) \geq 0\}.$$

The set $\mathcal{P}_K$ turns out to be a closed convex set in the linear space $\text{Sym}(n)$.

The minimal value of the variational problem (1) admits the min-max formulation

$$\lambda_{\min}(A, K) = \inf_{u \in K} \sup_{\lambda \in \mathbb{R}} \overbrace{\langle u, Au\rangle - \lambda(\langle u, u\rangle - 1)}^{L(u,\lambda)}. \tag{9}$$

By exchanging the order of the infimum and the supremum one gets

$$\beta(A, K) = \sup_{\lambda \in \mathbb{R}} \inf_{u \in K} L(u, \lambda),$$

which after a short simplification yields

$$\beta(A, K) = \sup\{\lambda \in \mathbb{R} : A - \lambda I \in \mathcal{P}_K\}. \tag{10}$$

One refers to (10) as the dual problem associated to (1).

Although the Lagrangean function $L : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ introduced in (9) fails to be convex with respect to the minimization variable u, there is no duality gap between the primal problem (1) and its dual (10). This and other facts are properly recorded in the next proposition.

Proposition 1 *Let $A \in \mathrm{Sym}(n)$ and $K \in \Xi(\mathbb{R}^n)$. Then,*

(a) The function $L : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ introduced in (9) admits saddle points over $K \times \mathbb{R}$.
(b) There is no duality gap between (1) and (10), i.e., $\lambda_{\min}(A, K) = \beta(A, K)$.
(c) The dual problem (10) has exactly one global solution, namely $\lambda = \lambda_{\min}(A, K)$.

Proof We cannot use standard minimax theorems relying on convexity assumptions. The fundamental ingredient of our proof is positive homogeneity. The dual problem associated to (1) concerns the maximization of a profit function $\Psi : \mathbb{R} \to \mathbb{R} \cup \{-\infty\}$ given by

$$\Psi(\lambda) = \inf_{u \in K} L(u, \lambda) = \lambda + \inf_{u \in K} \langle u, [A - \lambda I]u \rangle = \begin{cases} \lambda & \text{if } A - \lambda I \in \mathcal{P}_K, \\ -\infty & \text{otherwise.} \end{cases}$$

This explains why we are getting the formulation (10) for the number $\beta(A, K)$. The function Ψ can be rewritten in an entirely different manner, namely

$$\Psi(\lambda) = \lambda + \inf_{\rho \geq 0} \Gamma_\lambda(\rho)$$

with

$$\Gamma_\lambda(\rho) = \inf_{\substack{u \in K \\ \|u\| = \rho}} [\langle u, Au \rangle - \lambda \langle u, u \rangle] = \rho^2 [\lambda_{\min}(A, K) - \lambda].$$

Carrying out the minimization with respect to the auxiliary variable $\rho \geq 0$ one arrives at

$$\Psi(\lambda) = \begin{cases} \lambda & \text{if } \lambda \leq \lambda_{\min}(A, K), \\ -\infty & \text{otherwise .} \end{cases}$$

The later characterization of Ψ shows that the dual problem (10) has $\lambda = \lambda_{\min}(A, K)$ as unique global solution, in which case $\beta(A, K) = \lambda_{\min}(A, K)$. Finally, if x is a global solution to (1), then the pair $(x, \lambda_{\min}(A, K))$ is a saddle point of L over $K \times \mathbb{R}$. □

Remark 1 A key observation concerning the minimization problem (1) is that the cost function q_A is positively homogeneous (of degree 2) and the constraint function $\|\cdot\|$ is nonnegative and positively homogeneous (of degree 1). Proposition 1 could be obtained from a more general duality result on minimization problems with positively homogeneous data.

2 Local minima versus global minima

Needless to say, the main source of difficulties in the analysis of (1) is the "non-negativity constraint" induced by the cone K. The complexity of our minimization problem depends essentially on the structure of K.

Let us start with some words concerning the classical case of a subspace-constrained quadratic minimization problem. If

$$\lambda_1(A) \leq \lambda_2(A) \leq \cdots \leq \lambda_n(A)$$

are the eigenvalues of $A \in \mathrm{Sym}(n)$ arranged in non-decreasing order, then Fischer's famous max-min principle asserts that

$$\lambda_i(A) = \max_{\substack{K \in \mathcal{V}(\mathbb{R}^n) \\ \dim K = n-i+1}} \lambda_{\min}(A, K) \qquad \forall i \in \{1, 2, \ldots, n\}$$

with $\mathcal{V}(\mathbb{R}^n)$ denoting the set of all linear subspaces of $\mathbb{R}^n$. Notice that Fisher's principle involves the minimal value of the variational problem (1). The following proposition is probably known. We give its proof for the sake of completeness and for paving the way to the discussion of the general cone-constrained framework.

Proposition 2 *Let $A \in \mathrm{Sym}(n)$ and K be a d-dimensional linear subspace of $\mathbb{R}^n$. Then,*

(a) Any local solution $x \in \mathbb{R}^n$ to (1) is in fact a global solution to (1), i.e., $x \in K \cap \mathbb{S}_n$ and $\langle x, Ax\rangle = \lambda_{\min}(A, K)$.
(b) $\sigma_{\mathrm{locmin}}(A, K)$ contains $\lambda_{\min}(A, K)$ as unique element.
(c) $\lambda_{\min}(A, K)$ is equal to the smallest eigenvalue of the symmetric matrix $V^T AV$, where V stands for any matrix of size $n \times d$ whose columns form an orthonormal basis of K.

Proof If K is a linear subspace, then (2) becomes $Ax - \langle x, Ax\rangle x \in K^\perp$ with $K^\perp$ denoting the orthogonal of K. Linearity of K also implies that $\mathbb{R}_+(K - x) = K$. One gets $\mathcal{C}_K(x) = K$ and the second-order optimality condition (3) becomes

$$\langle h, Ah\rangle \geq \langle x, Ax\rangle \quad \forall h \in K \cap \mathbb{S}_n.$$

In other words, x is a global solution to (1). Part (b) is a direct consequence of (a). Finally, consider a representation of K in the form $K = \{Vz : z \in \mathbb{R}^d\}$ with V as indicated in (c). From [15,Example 2.2] one knows that $\sigma(A, K) = \mathrm{spec}(V^T AV)$, where the notation $\mathrm{spec}(E)$ refers to the usual spectrum of a symmetric matrix E. For completing the proof of the proposition it suffices to recall the general formula (8). □

The analysis of (1) is more interesting when K is not a linear subspace. In the truly conic case it is not possible to get rid of the non-negativity constraint $x \in K$ and transform (1) into an equivalent unconstrained eigenvalue problem. The situation is more involved as one may expect.

A striking feature of cone-constrained eigenvalue problems is that a local solution doesn't need to be a global one. It is natural then to address the following question: which region of K contains a local solution that is not a global one? The next proposition suggests that our attention is not to be directed toward the relative interior of the cone. A discrepancy between local and global optimality can occur only on the relative boundary of the cone.

In what follows one uses the notation $\mathrm{ri}(K)$ to indicate the relative interior of K. The linear space spanned by K is denoted by $\mathrm{span}K$.

Proposition 3 *Let $A \in \mathrm{Sym}(n)$ and $K \in \Xi(\mathbb{R}^n)$. Suppose that x is a local solution to (1) and that x belongs to $\mathrm{ri}(K)$. In such a case, x is a global solution to (1).*

Proof In view of Theorem 1 and the hypotheses made on x, one has

$$Ax - \langle x, Ax\rangle x \in [\mathrm{span}K]^{\perp}. \tag{11}$$

That x lies in the relative interior of K amounts to saying that $\mathbb{R}_+(K - x) = \mathrm{span}K$. The latter equality and (11) yield $\mathcal{C}_K(x) = \mathrm{span}K$. The second-order optimality condition (3) takes the form

$$\langle h, Ah\rangle \geq \langle x, Ax\rangle \quad \forall h \in \mathrm{span}K,\ \|h\| = 1, \tag{12}$$

proving in this way that x is a global solution to (1). □

Remark 2 Be aware that a local solution x as in Proposition 3 is not necessarily an eigenvector of A. What is true, however, is that x is an eigenvector of the symmetric matrix $A_K = P_K A P_K$, where $P_K : \mathbb{R}^n \to \mathbb{R}^n$ stands for the orthogonal projection onto $\mathrm{span}K$. The corresponding eigenvalue is $\langle x, A_K x\rangle = \langle x, Ax\rangle$.

Remark 3 The condition (12) proves not only that x is a global solution to (1), but a bit more than that. One deduces that x is also a global solution to the subspace-constrained eigenvalue problem

$$\lambda_{\min}(A, \mathrm{span}K) = \min_{\substack{u \in \mathrm{span}K \\ \|u\|=1}} \langle u, Au\rangle.$$

The later problem falls within the framework of Proposition 2. If $\mathrm{span}K$ is a d-dimensional subspace of $\mathbb{R}^n$, then

$$\lambda_{\min}(A, K) = \langle x, Ax\rangle = \lambda_{\min}(A, \mathrm{span}K) = \lambda_1(V_K^T A V_K),$$

where V_K stands for any matrix of size $n \times d$ whose columns form an orthonormal basis of $\mathrm{span}K$, and $\lambda_1(E)$ indicates the smallest eigenvalue of a symmetric matrix E.

3 How many local minimal values?

In this section we are interested in estimating the cardinality of $\sigma_{\mathrm{locmin}}(A, K)$ under the assumption that K is a polyhedral convex cone in $\mathbb{R}^n$. That a closed convex cone K is polyhedral simply means that it can be represented as intersection of finitely many half-spaces. This is equivalent to saying that K admits the representation

$$K = \mathrm{cone}\{g^1, \ldots, g^p\} = \left\{ \sum_{i=1}^{p} \alpha_i g^i : \alpha \in \mathbb{R}^p_+ \right\}$$

for some finite collection $\{g^1, \ldots, g^p\}$ of unit vectors in $\mathbb{R}^n$. There is no loss of generality in assuming that none of the g^i is a positive linear combination of the others. One usually refers to these vectors as the generators of the cone.

The polyhedrality hypothesis implies that $\sigma(A, K)$ is a finite set (cf. [14]), and so is therefore the smaller set $\sigma_{\mathrm{locmin}}(A, K)$. It is worthwhile noticing that the lack of polyhedrality may lead to a K-spectrum that is not even countable (cf. [7]).

To proceed further with the presentation we need to recall some basic facts from the theory of faces. By a *face* of a closed convex cone K one understands a convex cone F, subset of K, such that

$$u, v \in K \text{ and } u + v \in F \quad \Longrightarrow \quad u, v \in F.$$

A face is necessarily closed. The trivial set $\{0\}$ is a face of K if and only if K is pointed. A face F of K is said to be exposed[1] if it expressible in the form

$$F = K \cap y^{\perp} = \{u \in K : \langle y, u\rangle = 0\}$$

for a suitable vector $y \in K^{+}$. The dimension of a face F, denoted by $\dim F$, is simply the dimension of the linear space spanned by F. In the sequel one uses the notation

$$\begin{aligned}
\mathcal{F}(K) &\equiv \text{set of all faces of } K,\\
\mathcal{F}_{*}(K) &\equiv \text{set of all faces of } K \text{ excluding the trivial (or zero) face},\\
\mathcal{F}_{d}(K) &\equiv \text{set of all } d\text{-dimensional faces of } K.
\end{aligned}$$

For each non-zero vector x in K there is a unique $F \in \mathcal{F}_{*}(K)$ such that $x \in \mathrm{ri}(F)$. Such F is called the face of K associated to x. Sometimes it is convenient to express the condition $x \in \mathrm{ri}(F)$ by saying that F produces the vector x. Additional material from the theory of faces will be incorporated when the need arises.

Theorem 2 *Let $A \in \mathrm{Sym}(n)$ and $K \in \Xi(\mathbb{R}^n)$. Let x be a local solution to (1) and F be the associated face. Then,*

$$\langle x, Ax\rangle = \lambda_{\min}(A, \mathrm{span} F) = \lambda_1(V_F^T A V_F) \tag{13}$$

with V_F standing for any matrix of size $n \times \dim F$ whose columns form an orthonormal basis of $\mathrm{span} F$.

Proof Since x is a local solution to (1) and x belongs to F, it follows that x is a local solution to

$$\begin{aligned}
&\text{minimize } \langle u, Au\rangle\\
&u \in F \cap \mathbb{S}_n.
\end{aligned}$$

If one views the quadratic form q_A as a function on the linear space spanned by F, then we are back to context of Proposition 3, except that now one works with the closed convex cone F and not with K. Since x belongs to $\mathrm{ri}(F)$, it follows that x is a global solution to the subspace-constrained eigenvalue problem

$$\lambda_{\min}(A, \mathrm{span} F) = \min_{\substack{u \in \mathrm{span} F\\ \|u\|=1}} \langle u, Au\rangle. \tag{14}$$

This clearly yields the equalities stated in (13). □

Several consequences can be derived from Theorem 2. The next corollary shows that each face of K produces at most one local minimal value of (1), regardless of the dimension of that face.

Corollary 1 *Let $A \in \mathrm{Sym}(n)$ and $K \in \Xi(\mathbb{R}^n)$. If x and x' are two local solutions to (1) having the same associated face, then $\langle x, Ax\rangle = \langle x', Ax'\rangle$.*

[1] A non-polyhedral convex cone may well have a face that is not exposed, but this cannot occur in a polyhedral setting. Another advantage of working with polyhedral convex cones is that they have finitely many faces.

Proof Let F be the face shared by x and x'. Theorem 2 indicates that the local minimal values $\langle x, Ax\rangle$ and $\langle x', Ax'\rangle$ are both equal to $\lambda_{\min}(A, \text{span}F)$. □

A matrix of the form $V_F^T AV_F$ is called a truncation of A relative to the face F. Notice that A has many truncations relative to a prescribed face F because there are many orthonormal basis for span$F$. However, all the truncations of A relative to F have the same spectrum and therefore $\lambda_1(V_F^T AV_F)$ is defined unambiguously.

We mention in passing that the eigenvalues of A and those of $V_F^T AV_F$ satisfy the Poincaré interlacing property

$$\lambda_i(A) \leq \lambda_i(V_F^T AV_F) \leq \lambda_{n-d+i}(A) \qquad \forall i \in \{1, 2, \ldots, d\} \tag{15}$$

with d being the dimension of F. One gets in this way a localization result for the local minimal values associated to faces with prescribed dimension.

Corollary 2 *Let $A \in \text{Sym}(n)$ and $K \in \Xi(\mathbb{R}^n)$. Any local minimal value of (1) produced by a d-dimensional face of K lies in the interval $[\lambda_1(A), \lambda_{n-d+1}(A)]$.*

Proof Combine Theorem 2 and the interlacing property (15) for $i = 1$. □

An improved version of Corollary 2 is stated in the next proposition. Such improved version is helpful when the cone K is not full dimensional in $\mathbb{R}^n$, or when we have already computed the eigenvalues

$$\lambda_1(V_F^T AV_F) \leq \lambda_2(V_F^T AV_F) \leq \cdots \leq \lambda_k(V_F^T AV_F)$$

of a truncation of A relative to a given k-dimensional face F of K. One can take F as the cone K itself (in which case $k = \dim K$), but the choice of another face of K is also acceptable.

Proposition 4 *Let $A \in \text{Sym}(n)$ and $K \in \Xi(\mathbb{R}^n)$. Consider an integer $k \in \{1, 2, \ldots, \dim K\}$ and a k-dimensional face F of K. Then, any local minimal value of (1) produced by a d-dimensional subface of F lies in the interval $[\lambda_1(V_F^T AV_F), \lambda_{k-d+1}(V_F^T AV_F)]$.*

Proof Let $\lambda \in \mathbb{R}$ be a local minimal value associated to a d-dimensional subface G of F. By Theorem 2 we know that λ is representable in the form

$$\lambda = \lambda_1(V_G^T AV_G) \tag{16}$$

with V_G denoting any matrix of size $n \times d$ whose columns form an orthonormal basis of span$G$. Notice that the span of G is contained in the span of F because we are assuming that G is a subface of F. We can expand the matrix V_G with $r = k - d$ additional columns $\{w_1, \ldots, w_r\} \subset \mathbb{R}^n$ in such a way that the columns of the expanded matrix $V = [V_G, w_1, \ldots, w_r]$ form an orthonormal basis of span$F$. The eigenvalues of the $d \times d$ symmetric matrix $V_G^T AV_G$ and those of the larger $k \times k$ symmetric matrix $V^T AV$ are interlaced according to Poincaré's inequality

$$\lambda_i(V^T AV) \leq \lambda_i(V_G^T AV_G) \leq \lambda_{k-d+i}(V^T AV) \qquad \forall i \in \{1, 2, \ldots, d\}.$$

On the other hand, both truncations $V^T AV$ and $V_F^T AV_F$ of A relative to F have the same spectrum. One gets in particular

$$\lambda_1(V_F^T AV_F) \leq \lambda_1(V_G^T AV_G) \leq \lambda_{k-d+1}(V_F^T AV_F). \tag{17}$$

In view of (16), the proof of the proposition is complete. □

Remark 4 The localization result stated in Proposition 4 becomes sharper when the difference $k-d$ gets smaller. For instance, if d is just one dimension less than k, then the consecutive eigenvalues $\lambda_1(V_F^T A V_F)$ and $\lambda_2(V_F^T A V_F)$ serve to bound all the local minimal values of (1) associated to the d-dimensional subfaces of F. Another interesting remark is this: if the eigenvalue $\lambda_1(V_F^T A V_F)$ has algebraic multiplicity greater than or equal to $k-d+1$, then both sides of the sandwich (17) coincide, and therefore the d-dimensional subfaces of F can produce together at most one local minimal value of (1).

Enough has been said about local minimal values associated to faces of prescribed dimension. The next proposition is a containment result as well as a cardinality result for the set $\sigma_{\text{locmin}}(A, K)$. The first part doesn't requires polyhedrality, the second part, does.

Proposition 5 *Let $A \in \text{Sym}(n)$ and $K \in \Xi(\mathbb{R}^n)$. Then,*

$$\sigma_{\text{locmin}}(A, K) \subset \{\lambda_{\min}(A, \text{span}F) : F \in \mathcal{F}_*(K)\}, \tag{18}$$

i.e., each element of $\sigma_{\text{locmin}}(A, K)$ can be represented as the smallest eigenvalue of a truncation of A relative to some face of K. Furthermore, if K is polyhedral, then the cardinality of $\sigma_{\text{locmin}}(A, K)$ cannot exceed the number of non-trivial faces of K:

$$\text{card}[\sigma_{\text{locmin}}(A, K)] \leq \sum_{d=1}^{\dim K} f_K(d) \tag{19}$$

with $f_K(d) = \text{card}[\mathcal{F}_d(K)]$ standing for the number of d-dimensional faces of K.

Proof Everything is an easy consequence of Theorem 2. We shall prove later a more general version of this proposition (cf. Theorem 3). □

The upper bound (19) doesn't apply to the full K-spectrum of A. There are cases in which the general inclusion

$$\sigma(A, K) \subset \bigcup_{F \in \mathcal{F}_*(K)} \sigma(A, \text{span}F) \tag{20}$$

occurs as an equality and therefore the only thing one can say about the cardinality of $\sigma(A, K)$ when K is polyhedral is that

$$\text{card}[\sigma(A, K)] \leq \sum_{d=1}^{\dim K} d\, f_K(d). \tag{21}$$

By contrast, (20) and (21) hold for any matrix A, be it symmetric or not (cf. [15,Theorem 3.4]).

The next corollary shows that the bound (19) is sharp at least when closed half-spaces are concerned. Of course, the bound (19) is also sharp for linear subspaces and for half-lines.

Corollary 3 *Let K be a closed half-space in $\mathbb{R}^n$. Then,*

(a) For any $A \in \text{Sym}(n)$, the set $\sigma_{\text{locmin}}(A, K)$ has at most two elements.
(b) There exists a matrix $A \in \text{Sym}(n)$ such that $\text{card}[\sigma_{\text{locmin}}(A, K)] = 2$.

Proof A closed half-space in $\mathbb{R}^n$ is a convex cone of the form

$$K = \{u \in \mathbb{R}^n : \langle w, u \rangle \geq 0\} \tag{22}$$

with $w \in \mathbb{S}_n$. Note that K has only two faces: the cone K itself and the hyperplane $w^\perp$. Hence, the only possible local minimal values of (1) are $\lambda_1(A)$ and $\lambda_{\min}(A, w^\perp)$. This takes

care of part (a). For constructing a matrix A as in (b), consider first the particular half-space $\hat{K} = \{u \in \mathbb{R}^2 : u_1 \geq 0\}$. We claim that it is possible to find two different real numbers μ_1, μ_2 and a matrix

$$\hat{A} = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

such that

$$\sigma_{\text{locmin}}(\hat{A}, \hat{K}) = \{\mu_1, \mu_2\}. \tag{23}$$

For proving this claim we convert the minimization problem

$$\begin{aligned} &\text{minimize } a\,u_1^2 + 2b\,u_1u_2 + c\,u_2^2 \\ &u_1^2 + u_2^2 = 1 \\ &u_1 \geq 0 \end{aligned}$$

into that of minimizing the univariate function

$$t \in [-1, 1] \;\mapsto\; \varphi(t) = a\,(1 - t^2) + 2b\,t\sqrt{1 - t^2} + c\,t^2. \tag{24}$$

If one takes, for instance, $a = 1, b = 1$, and $c = 0$, then (24) attains a local minimum at $t = 1$, as well as a local minimum at

$$\hat{t} = -\sqrt{\frac{5 + \sqrt{5}}{10}} \approx -0.8507.$$

So, $\sigma_{\text{locmin}}(\hat{A}, \hat{K})$ is formed by $\mu_1 = \varphi(\hat{t}) = (1 - \sqrt{5})/2 \approx -0.618$ and by $\mu_2 = \varphi(1) = 0$. This completes the proof of (23). Parenthetically, note that $\mu_1 < \mu_2$ in consistency with Proposition 3. Indeed, the local solution $\hat{x} = (\sqrt{1 - \hat{t}^2}, \hat{t})$ lies in $\text{ri}(\hat{K})$ and therefore it must be a global solution. Consider now the general half-space (22). Let Q be an $n \times n$ orthonormal matrix such that $Qw = e_1 = (1, 0, 0, \ldots, 0)^T$. With this choice of Q one gets $Q(K) = \hat{K} \times \mathbb{R}^{n-2}$. If one defines

$$A = Q^T \begin{bmatrix} \hat{A} & 0 \\ 0 & 0 \end{bmatrix} Q,$$

then a matter of computation shows that

$$\sigma_{\text{locmin}}(A, K) = \sigma_{\text{locmin}}\left(\begin{bmatrix} \hat{A} & 0 \\ 0 & 0 \end{bmatrix}, \hat{K} \times \mathbb{R}^{n-2}\right) = \sigma_{\text{locmin}}(\hat{A}, \hat{K}).$$

This and (23) complete the proof of the corollary. □

4 The two-out-of-three rule

For some special classes of convex cones there is a bit of room for improvement in the upper bound (19). However, a complete revision of our counting strategy is needed in order to quantify a possible gain. While dealing with general polyhedral convex cones it is rather rare to have every face producing a different local minimal value of (1). In practice, plenty of faces are "idles" and don't contribute to the formation of the set $\sigma_{\text{locmin}}(A, K)$.

Next we develop a practical rule for removing idle faces. The lemma below gives us a feeling for what happens with (1) when one considers a cone generated by two vectors. Such a simple situation will give us a clue on how to treat more involved cases.

Lemma 1 *Let $A \in \mathrm{Sym}(n)$. Consider a convex cone*

$$K = \mathrm{cone}\{g^1, g^2\} = \{\alpha_1 g^1 + \alpha_2 g^2 : \alpha_1, \alpha_2 \geq 0\} \tag{25}$$

generated by two linearly independent unit vectors g^1, g^2 in $\mathbb{R}^n$. Then, $\sigma_{\mathrm{locmin}}(A, K)$ has at most two elements.

Proof Since g^1, g^2 are assumed to be linearly independent, the convex cone (25) has a span which is two-dimensional. We construct an orthonormal basis $\{v^1, v^2\}$ for span $K = \mathrm{span}\{g^1, g^2\}$ by using the Gram-Schmidt orthogonalization procedure: if $\theta \in]0, \pi[$ denotes the angle formed by g^1 and g^2, then we write

$$v^1 = g^1 \quad \text{and} \quad v^2 = -\left(\frac{\cos\theta}{\sin\theta}\right) g^1 + \left(\frac{1}{\sin\theta}\right) g^2.$$

Since $K \cap \mathbb{S}_n = \{(\cos t)v^1 + (\sin t)v^2 : t \in [0, \theta]\}$, one can write (1) as a one-dimensional minimization problem, namely

$$\begin{aligned} &\text{minimize } \overbrace{a\cos^2 t + 2b \sin t \cos t + c \sin^2 t}^{\varrho(t)} \\ &t \in [0, \theta]\,. \end{aligned} \tag{26}$$

We may suppose that the Gramian matrix

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix} = \begin{bmatrix} \langle v^1, Av^1\rangle & \langle v^1, Av^2\rangle \\ \langle v^2, Av^1\rangle & \langle v^2, Av^2\rangle \end{bmatrix}$$

is not equal to zero, otherwise the quadratic form q_A vanishes everywhere on K. We leave aside also the configuration $b = 0, a = c$ because in that case q_A would be constant over K. Counting the number of local minimal values of (26) is not a difficult task. The boundary points $t = 0$ and $t = \theta$ may be local minima or not. This will depend on the signs of the derivatives

$$\begin{aligned} \varrho'(0) &= 2b, \\ \varrho'(\theta) &= (c - a)\sin(2\theta) + 2b\cos(2\theta). \end{aligned}$$

As candidate for local minimality we must consider also any point $t_0 \in]0, \theta[$ such that $\varrho'(t_0) = 0$. A quick analysis of the behavior of ϱ' over $[0, \theta]$ shows that (26) admits two local minimal values at the most. One concludes in this way that $\sigma_{\mathrm{locmin}}(A, K)$ has one or two elements. □

The proof technique of Lemma 1 can be exploited a bit further. Among other things, we have proved that:

(i) If $b > 0$, then $\langle g^1, Ag^1\rangle \in \sigma_{\mathrm{locmin}}(A, K)$.
(ii) If $\varrho'(\theta) < 0$, then $\langle g^2, Ag^2\rangle \in \sigma_{\mathrm{locmin}}(A, K)$.
(iii) If $b > 0$ and $\varrho'(\theta) < 0$, then no local solution to (1) is to be found in $\mathrm{ri}(K)$.
(iv) If g^1 and $x \in \mathrm{ri}(K)$ are local solutions to (1) yielding different local minimal values, then g^1 and x form an angle greater than $\pi/2$ and $\langle x, Ax\rangle < \langle g^1, Ag^1\rangle$.

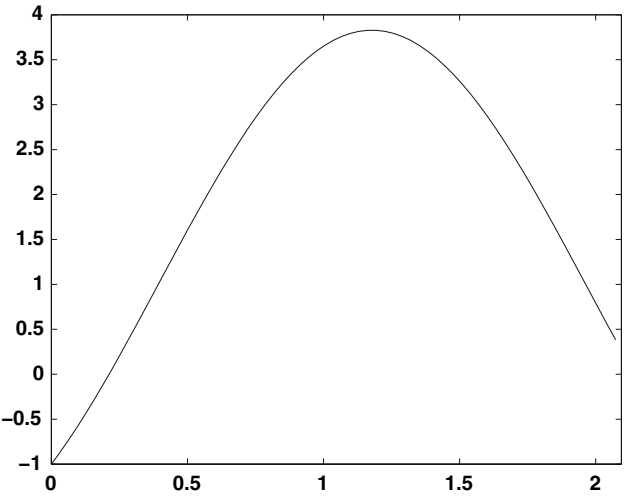

Fig. 1 Shape of ϱ when $a = -1, b = 2, c = 3$

The statement (iv) is quite subtle and requires perhaps an explanation. Let g^1 and $x \in \mathrm{ri}(K)$ be local solutions to (1) with $\langle g^1, Ag^1\rangle \neq \langle x, Ax\rangle$. The function ϱ increases over the interval $[0, \hat{t}]$, where $\hat{t}$ is the smallest positive real such that $\varrho'(\hat{t}) = 0$, i.e.,

$$\hat{t} = \frac{1}{2}\arctan\left(\frac{2b}{a-c}\right).$$

The function ϱ starts then to decrease until one reaches a second point $\tilde{t}$ at which ϱ' vanishes. This point $\tilde{t}$ is a local minimum of ϱ and x is in fact given by $x = (\cos\tilde{t})v^1 + (\sin\tilde{t})v^2$. A simple inspection at the function ϱ' shows that $\tilde{t} = \hat{t} + (\pi/2)$, proving in this way that the angle between g^1 and x is greater than $\pi/2$. By the way, since x is assumed to be in $\mathrm{ri}(K)$, we must have $\tilde{t} < \theta$. In other words, the situation described in (iv) can only occur if the generators g^1, g^2 form an angle greater than $\pi/2$. Finally, to see that $\langle x, Ax\rangle < \langle g^1, Ag^1\rangle$ one just needs to compare the values $\varrho(\hat{t})$ and $\varrho(0)$.

Figures 1 and 2 illustrate some of the possible shapes of ϱ depending on the parameters a, b, c. As angle between the generators we are taking $\theta = 2\pi/3$. In Fig. 1 the generators g^1, g^2 are producing two different local minimal values of (1). Hence, no local solution is to be found in the relative interior of the cone. In Fig. 2 only the generator g^1 produces a local minimal value. A second local minimal value is produced by a local solution x lying in the relative interior of the cone. As one can see from the graph of ϱ, the second local minimal value is in fact the global minimum of (1). This observation is consistent with Proposition 3. One can also see that the angle between the local solution g^1 and the global solution $x \in \mathrm{ri}(K)$ is greater than $\pi/2$. As explained some lines above, this phenomenon is not accidental.

The three non-trivial faces of the convex cone (25) can produce together at most two local minimal values of (1). This is what we call the "two-out-of-three rule". At least one of the three non-trivial faces of (25) is idle and this suggests that the upper bound (19) admits some sharpening. A more general formulation of the "two-out-of-three rule" reads as follows:

Proposition 6 *Let $A \in \mathrm{Sym}(n)$. Consider a convex cone $K = \mathrm{cone}\{g^1, \ldots, g^p\}$ generated by a collection $\{g^1, \ldots, g^p\}$ of p unit vectors in $\mathbb{R}^n$. Let $i, j \in \{1, \ldots, p\}$ be such that $K_{i,j} = \mathrm{cone}\{g^i, g^j\}$ is a two-dimensional face of K. Under these assumptions one has:*

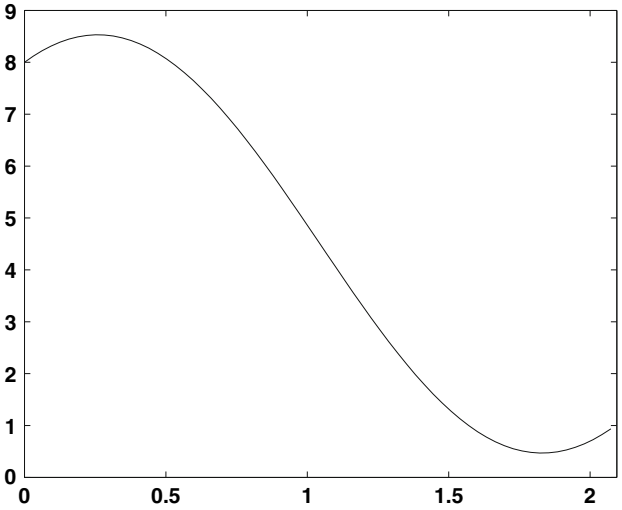

Fig. 2 Shape of ϱ when $a = 8, b = 2, c = 1$

(a) The three non-trivial faces of $K_{i,j}$ can produce together at most two local minimal values of (1).
(b) In case g^i and $x \in \mathrm{ri}(K_{i,j})$ are local solutions to (1) *producing different local minimal values, then g^i and x form an angle greater than $\pi/2$ and $\langle x, Ax\rangle < \langle g^i, Ag^i\rangle$.*

Proof For proving (a), suppose that g^i, g^j, and some $x \in \mathrm{ri}(K_{i,j})$, are local solutions of (1). In particular, g^i, g^j, x are local solutions of the variational problem which consists in minimizing the quadratic form q_A on the smaller set $K_{i,j} \cap \mathbb{S}_n$. We are then back to the framework of Lemma 1. Part (b) has been duely explained before. Of course, the situation described in (b) could occur only if g^i, g^j form an angle greater than $\pi/2$. □

As application of Proposition 6 consider a polyhedral cone K as in Fig. 3. Such cone is generated by p vectors in $\mathbb{R}^3$. As one can see, K has p one-dimensional faces, p two-dimensional faces, and 1 three-dimensional face, that is to say, $2p+1$ non-trivial faces in all. Hence, the general upper bound (19) yields the estimate

$$\mathrm{card}[\sigma_{\mathrm{locmin}}(A, K)] \;\leq\; 2p + 1.$$

On the other hand, if one applies the two-out-of-three rule to each one of the consecutive two-dimensional faces

$$\mathrm{cone}\{g^1, g^2\},\ \mathrm{cone}\{g^2, g^3\},\ \ldots,\ \mathrm{cone}\{g^{p-1}, g^p\},\ \mathrm{cone}\{g^p, g^1\},$$

then one ends up with the better estimate

$$\mathrm{card}[\sigma_{\mathrm{locmin}}(A, K)] \;\leq\; 2p + 1 - \lceil p/2 \rceil\,, \tag{27}$$

where $\lceil r \rceil$ stands for the upper integer part of r.

A suitable two-out-of-three rule can be stated for non-polyhedral convex cones as well. In fact, it is not necessary to refer to the faces and to the generators of the cone.

Proposition 7 *Let $A \in \mathrm{Sym}(n)$ and $K \in \Xi(\mathbb{R}^n)$. Suppose that x^1, x^2, x^3 are three different local solutions to* (1) *such that one of them is a positive linear combination of the others. Then, the three values $\langle x^1, Ax^1\rangle$, $\langle x^2, Ax^2\rangle$, $\langle x^3, Ax^3\rangle$ are the same.*

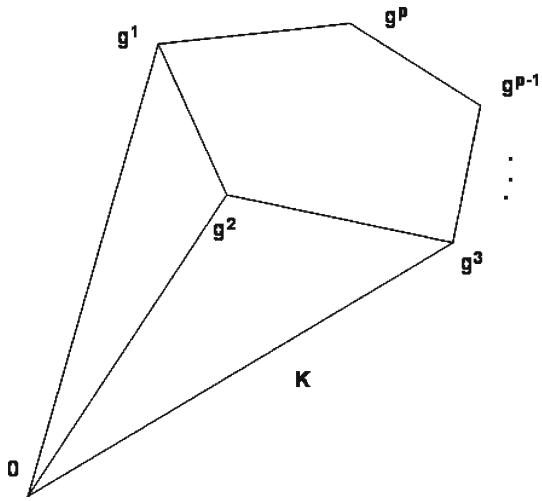

Fig. 3 A polyhedral cone generated by p vectors. No restrictions are imposed on the angles formed by the generatrors

Proof Suppose, for instance, that x^3 is a positive linear combination of x^1 and x^2. Since these three unit vectors are assumed to be different, it follows that x^1, x^2 are linearly independent and

$$x^3 \in \mathrm{ri}\left(\mathrm{cone}\{x^1, x^2\}\right).$$

For completing the proof of the proposition we apply Lemma 1 to the subcone $K_{1,2} = \mathrm{cone}\{x^1, x^2\}$. Since the vectors x^1, x^2, x^3 are local minima of the quadratic function q_A on $K_{1,2} \cap \mathbb{S}_n$ and they are placed in different faces of $K_{1,2}$, it follows that q_A must be constant over the arc $K_{1,2} \cap \mathbb{S}_n$. □

We state below another result in the same spirit but not comparable to Proposition 7.

Proposition 8 *Let $A \in \mathrm{Sym}(n)$ and $K \in \Xi(\mathbb{R}^n)$. If $\{x(t) : t \in [0,1]\}$ is an absolutely continuous curve formed by critical vectors of (1), then*

$$t \in [0,1] \mapsto \lambda(t) = \langle x(t), Ax(t)\rangle \tag{28}$$

is a constant function.

Proof We adapt a proof technique used for the analysis of critical angles in convex cones. A minor adjustment of [7,Proposition 2] is all what is needed. □

What Proposition 8 says is that two critical vectors of (1) yielding different critical values cannot be joined by an absolutely continuous curve formed by critical vectors of (1). It is not clear whether (28) remains constant if the curve under consideration is just continuous.

5 Pre-activity as relaxation of local minimality

The set on the right-hand side of (18) is obtained by ranging F over the full collection of non-trivial faces of K. The upper bound (19) is uniform in the sense that it applies to any matrix $A \in \mathrm{Sym}(n)$. This corresponds somehow to a worst case scenario. We next describe a more sophisticated method of selecting faces. This time we take into account not only the convex cone K but also the matrix $A \in \mathrm{Sym}(n)$.

Definition 1 *Let $A \in \mathrm{Sym}(n)$ and $K \in \Xi(\mathbb{R}^n)$. A face $F \in \mathcal{F}_*(K)$ is said to be pre-active for the variational problem (1) if there are a matrix V_F of size $n \times \dim F$ and a unit vector $z_F \in \mathbb{R}^{\dim F}$ such that:*

(i) the columns of V_F form an orthonormal basis of $\mathrm{span}F$,
(ii) z_F is an eigenvector of $V_F^T A V_F$ associated to the smallest eigenvalue $\lambda_1(V_F^T A V_F)$,
(iii) $V_F z_F \in \mathrm{ri}(F)$.

The collection of all pre-active faces for (1) is denoted by $\mathcal{F}_*(A, K)$. A face $F \in \mathcal{F}_*(K)$ is said to be active (respectively, critical) for the variational problem (1) if it is associated to a local solution (respectively, critical vector) of (1).

Although it is an abuse of language, we say sometimes that a face is pre-active (or active, or critical) without refering explicitly to the variational problem (1). The motivation behind Definition 1 should be clear by now. In fact, the concept of pre-activity is specially tailored for obtaining the following extension of Proposition 5.

Theorem 3 *Let $A \in \mathrm{Sym}(n)$ and $K \in \Xi(\mathbb{R}^n)$. Then,*

(a) Every active face $F \in \mathcal{F}_*(K)$ for (1) is pre-active for (1).
(b) One has the general inclusion

$$\sigma_{\mathrm{locmin}}(A, K) \subset \{\lambda_{\min}(A, \mathrm{span}F) : F \in \mathcal{F}_*(A, K)\}, \tag{29}$$

i.e., each element of $\sigma_{\mathrm{locmin}}(A, K)$ can be represented as the smallest eigenvalue of a truncation of A relative to some pre-active face of K.
(c) If K happens to be polyhedral, then

$$\lambda_{\min}(A, K) = \min_{F \in \mathcal{F}_*(A,K)} \lambda_1(V_F^T A V_F) \tag{30}$$

and the cardinality of $\sigma_{\mathrm{locmin}}(A, K)$ cannot exceed the number of pre-active faces:

$$\mathrm{card}[\sigma_{\mathrm{locmin}}(A, K)] \leq \mathrm{card}[\mathcal{F}_*(A, K)]. \tag{31}$$

Proof As in Proposition 5, this is again a matter of exploiting Theorem 2. Let us be more explicit this time. The proof of (b) runs as follows. Let $\lambda \in \sigma_{\mathrm{locmin}}(A, K)$. Hence, there is a local solution x to (1) such that $\lambda = \langle x, Ax\rangle$. Let F be the face associated to x. By definition, F is the unique face of K such that $x \in \mathrm{ri}(F)$. One can always construct a matrix V_F as in Definition 1(i). By (13) one knows that

$$\lambda = \lambda_{\min}(A, \mathrm{span}F) = \lambda_1(V_F^T A V_F).$$

We also know that x is a global solution to the subspace-constrained eigenvalue problem (14). Since the linear map $z \mapsto V_F z$ is a bijection between $\mathbb{R}^{\dim F}$ and $\mathrm{span}F$, there is a (unique) vector $z_F \in \mathbb{R}^{\dim F}$ such that $x = V_F z_F$. Such vector z_F depends of course on x. It is clear that z_F has unit length and satisfies the properties (ii) and (iii) in Definition 1. This shows that $F \in \mathcal{F}_*(A, K)$ and completes the proof of (b). The proof of (a) is similar. One starts with an active face F and then one takes any local solution x to (1) lying in the relative interior of F. As before, one concludes that F is pre-active. Suppose now that K is polyhedral. Formula (31) and the inequality

$$\lambda_{\min}(A, K) \geq \min_{F \in \mathcal{F}_*(A,K)} \lambda_1(V_F^T A V_F) \tag{32}$$

are consequences of (b). It remains to check that (32) occurs as an equality. Consider any $F \in \mathcal{F}_*(A, K)$ and construct

$$x_F = V_F z_F \tag{33}$$

with V_F and z_F as in Definition 1. Observe that x_F is feasible for (1) and

$$\lambda_{\min}(A, K) \leq \langle x_F, Ax_F \rangle = \langle z_F, V_F^T A V_F z_F \rangle = \lambda_1(V_F^T A V_F).$$

Since $F \in \mathcal{F}_*(A, K)$ was chosen arbitrarily, one gets

$$\lambda_{\min}(A, K) \leq \min_{F \in \mathcal{F}_*(A,K)} \lambda_1(V_F^T A V_F),$$

completing in this way the proof of (c). □

If K is polyhedral, then one can write

$$\lambda_{\min}(A, K) = \min_{F \in \mathcal{F}_*(K)} \lambda_{\min}(A, F) = \min_{F \in \mathcal{F}_*(A,K)} \lambda_{\min}(A, F).$$

These representation formulas are not helpful in practice because computing $\lambda_{\min}(A, F)$ is as difficult as computing the original expression $\lambda_{\min}(A, K)$. On the other hand, it is also possible to write

$$\lambda_{\min}(A, K) = \min_{\substack{F \in \mathcal{F}_*(K) \\ F \text{ active}}} \lambda_1(V_F^T A V_F),$$

but this formula is not helpful either. Although evaluating $\lambda_1(V_F^T A V_F)$ is a matter of classical linear algebra, identifying the active faces of K is a tough job. The general idea supporting the use of Theorem 3 is that the task of identifying the pre-active faces of K is much easier.

Example 1 By way of illustration we show how to identify the pre-active faces of the Pareto cone $K = \mathbb{R}^n_+$. The Pareto cone, also referred to as the non-negative orthant, is by far the most popular of all ordering cones in $\mathbb{R}^n$. It is not difficult to check that, for any $A \in \mathrm{Sym}(n)$, one has

$$F \in \mathcal{F}_*(A, \mathbb{R}^n_+) \Longleftrightarrow \begin{cases} \text{the smallest eigenvalue of } V_F^T A V_F \text{ admits} \\ \text{an eigenvector with positive components.} \end{cases} \tag{34}$$

The right-hand side of (34) is a test that can be treated with the standard tools of Perron-Frobenius eigenvalue analysis [2].

For the sake of convenience we refer to (33) as the "transfer equation" and to x_F as being a pre-active vector for the variational problem (1). If one adopts this terminology, then a pre-active face corresponds precisely to a face associated to a pre-active vector. This is consistent with other expressions like active face, critical face, etc.

A pre-active vector can be defined in a more direct manner without passing through the transfer equation. Also, a pre-active face can be "detected" without constructing explicitly the matrix V_F and the eigenvector z_F. The following proposition gives an alternative characterization of pre-activity. The leading role is now played by

$$A_F = P_F A P_F,$$

where the linear map $P_F : \mathbb{R}^n \to \mathbb{R}^n$ stands for the orthogonal projection onto $\mathrm{span} F$.

Proposition 9 *Let $A \in \mathrm{Sym}(n)$ and $K \in \Xi(\mathbb{R}^n)$. For a face $F \in \mathcal{F}_*(K)$ the following conditions are equivalent:*

(a) F is a pre-active for (1).
(b) F admits in its relative interior a unit vector x such that $A_F x = \lambda_{\min}(A, \mathrm{span}F)x$.

Proof Let F be pre-active. Construct V_F and z_F as in Definition 1. The linear map P_F clearly admits $V_F V_F^T$ as matrix representation. Left multiplication by V_F on both sides of

$$V_F^T A V_F z_F = \lambda_1(V_F^T A V_F) z_F$$

leads to the new equality

$$P_F A x_F = \lambda_{\min}(A, \mathrm{span}F) x_F,$$

where x_F is given by the transfer equation. Since x_F belongs to the relative interior of F, it follows that $P_F x_F = x_F$ and therefore $A_F x_F = \lambda_{\min}(A, \mathrm{span}F)x_F$. Conversely, suppose that $A_F x = \lambda_{\min}(A, \mathrm{span}F)x$ holds for some unit vector $x \in \mathrm{ri}(F)$. If V_F is defined as usual, one gets

$$V_F V_F^T A x = \lambda_{\min}(A, \mathrm{span}F)x.$$

One can write $x = V_F z$ for a suitable unit vector $z \in \mathbb{R}^{\dim F}$. This leads to

$$V_F \left[V_F^T A V_F z - \lambda_1(V_F^T A V_F) z \right] = 0.$$

The vector between square brackets must be zero and therefore z is an eigenvector associated to $\lambda_1(V_F^T A V_F)$. This proves that F is pre-active. □

One-dimensional faces are always pre-active as one can see from Proposition 9(b). The condition (b) in Proposition 9 seems a shorter and simpler way of introducing the concept of pre-activity. However, the simplicity of this formulation is only apparent. Most of the heavy work is hidden in the computation of the matrix A_F. Besides, the dimension of $V_F^T A V_F$ is smaller than the dimension of A_F and this fact counts when it comes to compute eigenvalues and eigenvectors.

The next example serves to illustrate the usefulness of the formula (30) stated in Theorem 3. We write down all the details for the sake of pedagogy.

Example 2 Consider the V-shaped cantilever $K = \{x \in \mathbb{R}^3 : x_3 \geq |x_2|\}$ and the symmetric matrix

$$A = \begin{bmatrix} 3 & -2 & 1 \\ -2 & 1 & 2 \\ 1 & 2 & 3 \end{bmatrix}.$$

The cantilever admits the bottom line $F_1 = \mathbb{R}(1, 0, 0)$ as one-dimensional face, the sets

$$F_2 = \{x \in \mathbb{R}^3 : x_2 \geq 0,\ x_2 - x_3 = 0\}$$
$$F_3 = \{x \in \mathbb{R}^3 : x_2 \leq 0,\ x_2 + x_3 = 0\}$$

as two-dimensional faces, and $F_4 = K$ as three-dimensional face. One has

$$V_{F_1} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix},\ V_{F_2} = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{2}/2 \\ 0 & \sqrt{2}/2 \end{bmatrix},\ V_{F_3} = \begin{bmatrix} 1 & 0 \\ 0 & -\sqrt{2}/2 \\ 0 & \sqrt{2}/2 \end{bmatrix},\ V_{F_4} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

These matrices depend only on K. A matter of computation shows that

$$\begin{aligned}
&\lambda_1(V_{F_1}^T A V_{F_1}) = 3, && z_{F_1} = 1,\\
&\lambda_1(V_{F_2}^T A V_{F_2}) = 2.6340, && z_{F_2} = (0.8881, 0.4597)^T,\\
&\lambda_1(V_{F_3}^T A V_{F_3}) = -1.0981, && z_{F_3} = (-0.4597, 0.8881)^T\\
&\lambda_1(V_{F_4}^T A V_{F_4}) = -1.3723, && z_{F_4} = (-0.4544, -0.7662, 0.4544)^T.
\end{aligned}$$

The pre-active faces are F_1, F_2 and F_3. By using formula (30) one gets $\lambda_{\min}(A, K) = -1.0981$. This global minimal value is achieved with the global solution $x_{F_3} = V_{F_3} z_{F_3} = (-0.4597, -0.6280, 0.6280)$.

6 Specific results for infra-dual cones

One can derive various refinements for the results established in Sect. 3, but this requires asking more structure to the cone K. An interesting situation occurs when K is *infra-dual* in the sense that it is contained in K^+. Such requirement can be formulated in the simpler form

$$\langle u, v\rangle \geq 0 \quad \forall u, v \in K.$$

Geometrically speaking, the infra-duality assumption amounts to saying that the maximal angle

$$\theta_{\max}(K) = \max_{u,v \in K \cap \mathbb{S}_n} \arccos\langle u, v\rangle$$

of K is less than or equal to $\pi/2$. None of the results presented in this section is true for a convex cone whose maximal angle exceeds this threshold.

Proposition 10 *Let $A \in \mathrm{Sym}(n)$ and $K \in \Xi(\mathbb{R}^n)$ be infra-dual. If $\hat{x}, x \in \mathbb{R}^n$ are non-zero vectors such that*

$$\begin{aligned}
\hat{x} \in \mathrm{ri}(K)\ ,\ A\hat{x} = \hat{\lambda}\hat{x},\\
x \in K\ ,\ Ax = \lambda x,
\end{aligned} \tag{35}$$

then $\hat{\lambda} = \lambda$.

The above proposition is a known result pertaining to the realm of classical eigenvalue analysis. More interesting to us is the following cone-constrained version.

Lemma 2 *Let $A \in \mathrm{Sym}(n)$ and $K \in \Xi(\mathbb{R}^n)$ be infra-dual. Let $\hat{\lambda}$ be an eigenvalue of A admitting an eigenvector in $\mathrm{ri}(K)$. Then,*

$$\hat{\lambda} = \max\{\lambda : \lambda \in \sigma(A, K)\}. \tag{36}$$

Furthermore, any K-eigenvector of A associated to $\hat{\lambda}$ is an eigenvector of A.

Proof The hypothesis made on $\hat{\lambda}$ refers to the existence of a non-zero vector $\hat{x} \in \mathbb{R}^n$ satisfying (35). Such $\hat{\lambda}$ clearly belongs to $\sigma(A, K)$. Consider any other λ in the set $\sigma(A, K)$. One can find then a non-zero vector $x \in \mathbb{R}^n$ and a vector $y \in \mathbb{R}^n$ (orthogonal to x) such that

$$x \in K, \quad y \in K^+, \quad y = Ax - \lambda x. \tag{37}$$

The equalities appearing in (35) and (37) yield respectively

$$\begin{aligned}
\langle A\hat{x}, x\rangle &= \hat{\lambda}\langle \hat{x}, x\rangle,\\
\langle Ax, \hat{x}\rangle &= \langle y, \hat{x}\rangle + \lambda\langle x, \hat{x}\rangle.
\end{aligned}$$

The symmetry of A allows us to conclude that $\langle y, \hat{x}\rangle = (\hat{\lambda} - \lambda)\langle \hat{x}, x\rangle$. Since $\langle y, \hat{x}\rangle \geq 0$ and $\langle \hat{x}, x\rangle > 0$, it follows that $\hat{\lambda} \geq \lambda$. This takes care of (36). The particular choice $\lambda = \hat{\lambda}$ yields $\langle y, \hat{x}\rangle = 0$, but the later equality can occur only if $y = 0$ (recall that $\hat{x} \in \mathrm{ri}(K)$ and $y \in K^+$). This shows that any K-eigenvector of A associated to $\hat{\lambda}$ is in fact an eigenvector of A. □

We mention in passing an easy consequence of Lemma 2. The proof of Corollary 4 is immediate and therefore omitted.

Corollary 4 *Let $A \in \mathrm{Sym}(n)$ and $K \in \Xi(\mathbb{R}^n)$ be infra-dual. Suppose that the smallest eigenvalue of A admits an eigenvector in $\mathrm{ri}(K)$. Then, $\sigma(A, K)$ contains $\lambda_1(A)$ as unique element.*

A quite bothering aspect of Lemma 2 is the assumption made on $\hat{\lambda}$. Such an eigenvalue $\hat{\lambda}$ may well not exist, in which case the lemma says absolutely nothing. Fortunately, this problem can be remediated. If A doesn't have eigenvectors in $\mathrm{ri}(K)$, then we still have the possibility of invoking the following alternative result.

Proposition 11 *Let $A \in \mathrm{Sym}(n)$ and $K \in \Xi(\mathbb{R}^n)$ be infra-dual. Let $\hat{x}$ be a normalized K-eigenvector of A and $\hat{F}$ be the associated face. Then,*

$$\langle x, Ax\rangle \leq \langle \hat{x}, A\hat{x}\rangle$$

for all normalized K-eigenvector x of A whose associated face is contained in $\hat{F}$.

Proof We know that $\hat{x}$ is an eigenvector of $A_{\hat{F}}$ lying in $\mathrm{ri}(\hat{F})$. Pick up any normalized K-eigenvector x of A whose associated face F is contained in $\hat{F}$. The combination of $x \in \mathrm{ri}(F)$ and $F \subset \hat{F}$ implies that x is an $\hat{F}$-eigenvector of A. In particular, $x \in \mathrm{span}\hat{F}$. In view of the general inclusion $P_{\hat{F}}[\hat{F}^+] \subset \hat{F}^+$, it follows that x is an $\hat{F}$-eigenvector of $A_{\hat{F}}$. By applying Lemma 2 to the pair $(A_{\hat{F}}, \hat{F})$, one gets

$$\langle x, Ax\rangle = \langle x, A_{\hat{F}}x\rangle \leq \langle \hat{x}, A_{\hat{F}}\hat{x}\rangle = \langle \hat{x}, A\hat{x}\rangle.$$

□

Remark 5 Strictly speaking, in Proposition 11 one doesn't need K to be infra-dual. It would suffice asking the face $\hat{F}$ (and hence, F) to have a maximal angle less than or equal to $\pi/2$. This observation is quite useful when it comes to deal with a convex cone that is not infra-dual but has a large dimensional face satisfying this angular restriction. A similar type of remark applies to several of the remaining results of this section. It is worthwhile noticing that Proposition 11 provides an alternative way of deriving Proposition 6(b).

Although the analysis of critical values is not main focus of this paper, we state below an upper bound for the cardinality of $\sigma(A, K)$.

Corollary 5 *Let $A \in \mathrm{Sym}(n)$. The following implications are true:*

(a) If $K \in \Xi(\mathbb{R}^n)$ is infra-dual, then each face of K produces at most one critical value of (1).
(b) If $K \in \Xi(\mathbb{R}^n)$ is polyhedral and infra-dual, then $\mathrm{card}[\sigma(A, K)] \leq \sum_{d=1}^{\dim K} f_K(d)$.

The proof of Corollary 5 is omitted because everything follows straightforwardly from Proposition 11. The bound given in part (b) is better than (21) but, of course, it requires the symmetry of A and the infra-duality of K.

The ground is now ready to state one of the fundamental theorems of this paper.

Theorem 4 *Let $A \in \mathrm{Sym}(n)$ and $K \in \Xi(\mathbb{R}^n)$ be infra-dual. Let x, x' be two different local solutions to (1). If the associated faces F, F' are "connected" in the sense that*

$$F \subset F' \text{ or } F' \subset F, \tag{38}$$

then

(a) $\langle x, Ax\rangle = \langle x', Ax'\rangle$.

(b) *x and x' are eigenvectors of $A_{F\cup F'}$. In particular, the common term in part (a) is a multiple eigenvalue of $A_{F\cup F'}$.*

Proof Suppose, for instance, that $F \subset F'$. In view of Theorem 1, x and x' are normalized K-eigenvectors of A. Proposition 11 yields then the inequality $\langle x, Ax\rangle \leq \langle x', Ax'\rangle$. On the other hand, Theorem 2 allows us to write

$$\langle x', Ax'\rangle = \lambda_{\min}(A, \mathrm{span}F') = \min_{\substack{u\in \mathrm{span}F' \\ \|u\|=1}} \langle u, Au\rangle. \tag{39}$$

Since $x \in \mathrm{ri}(F)$ and $F \subset F'$, it follows that $x \in \mathrm{span}F'$. Hence, the unit vector x is feasible for the minimization problem (39). This proves the reverse inequality $\langle x', Ax'\rangle \leq \langle x, Ax\rangle$. We now take care of part (b). We continue assuming that $F \subset F'$ so that $F \cup F' = F'$. Let λ' denote the common term in (a). Local minimality of $x' \in \mathrm{ri}(F')$ implies that $A_{F'}x' = \lambda' x'$. By proceeding as in the proof of Proposition 11 one deduces that x is an F'-eigenvector of $A_{F'}$. By applying the second part of Lemma 2 to the pair $(A_{F'}, F')$, one obtains $A_{F'}x = \lambda' x$. In short, x and x' are eigenvectors of $A_{F'}$ associated to the same eigenvalue. This common eigenvalue is necessarily multiple because the unit vectors x, x' are not collinear. □

Both conclusions of Theorem 4 remain true if the vector produced by the smallest face is just critical. The local minimality hypothesis is important only for the vector produced by the largest face.

Connectivity in the sense (38) is an essential assumption in Theorem 4. Let us elaborate a bit further on the connectivity structure of a family of faces of a polyhedral cone.

Definition 2 *Let $K \in \Xi(\mathbb{R}^n)$ be polyhedral and $\mathcal{F}$ be a non-empty family of faces of K. One says that:*

(i) *$\mathcal{F}$ is pairwise unconnected if any two different faces taken from $\mathcal{F}$ are not connected.*

(ii) *$\mathcal{F}$ captures K if every $F \in \mathcal{F}(K)$ is connected to some $F' \in \mathcal{F}$.*

Somewhat implicit in Definition 2(i) is the fact that $\mathcal{F}$ is formed by at least two faces. When $\mathcal{F}$ is formed by a unique face it is convenient to declare $\mathcal{F}$ as being pairwise unconnected.

To each polyhedral cone K one can associate a finite graph $\mathcal{G}_K$ whose nodes are the non-trivial faces of K. One draws an edge between two nodes if the corresponding faces are connected. For instance,[2] the graph associated to the Pareto cone $\mathbb{R}^3_+$ is as indicated in Fig. 4.

Recall that the *chromatic number* of a finite graph $\mathcal{G}$ is the minimum number of colors needed to paint the nodes of $\mathcal{G}$ such that no two adjacent nodes have the same color (cf. [4,Chapter 5]). In the definition below we introduce a sort of dual concept.

Definition 3 *The multichromatic number of a finite graph $\mathcal{G}$ is the largest number of colors that is possible to use if one paints all or some of the nodes of $\mathcal{G}$ in such a way that two adjacent nodes are painted with the same color, in case both are painted.*

[2] We label the $2^n - 1$ non-trivial faces $F_1, F_2, F_3, \ldots$ of the Pareto cone $\mathbb{R}^n_+$ by using the binary ordering method [12]. This labeling procedure has many advantages. For instance, a given face F_k is d-dimensional if and only if the number of "1" in binary representation of the integer k is equal to d.

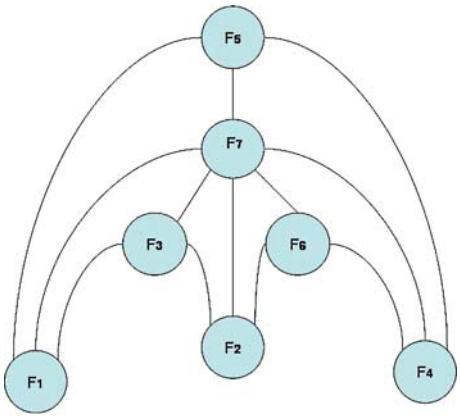

Fig. 4 The graph associated to the Pareto cone $\mathbb{R}^3_+$

The multichromatic number of a polyhedral cone K, denoted by $\nu(K)$, is simply the multichromatic number of the associated graph $\mathcal{G}_K$. It is straightforward to see that

$$\nu(K) = \max\{\text{card}[\mathcal{F}] : \mathcal{F} \in 2^{\mathcal{F}_*(K)} \text{ is pairwise unconnected}\} \tag{40}$$

with 2^S denoting the power set of S. The above integer can be interpreted as an index of pairwise unconnectivity of the polyhedral cone K.

Two comments regarding the definition of $\nu(K)$ are in order: firstly, the family $\mathcal{F}$ achieving the maximum in (40) is not necessarily unique; and, secondly, if $\mathcal{F}$ achieves the maximum in (40), then $\mathcal{F}$ captures K. Without further ado we state:

Proposition 12 *Let $K \in \Xi(\mathbb{R}^n)$ be polyhedral and infra-dual. Then, for all $A \in \text{Sym}(n)$, one has*

$$\text{card}[\sigma_{\text{locmin}}(A, K)] \leq \nu(K).$$

Proof Let $\mathcal{F}_{\text{locmin}}(A, K)$ denote the set of all active faces. If $\mathcal{F}_{\text{locmin}}(A, K)$ is pairwise unconnected, then $\text{card}[\sigma_{\text{locmin}}(A, K)] \leq \text{card}[\mathcal{F}_{\text{locmin}}(A, K)] \leq \nu(K)$, the first inequality being due to Corollary 5(a). If $\mathcal{F}_{\text{locmin}}(A, K)$ is not pairwise unconnected, then we drop from $\mathcal{F}_{\text{locmin}}(A, K)$ a face that is connected to another one in $\mathcal{F}_{\text{locmin}}(A, K)$. This operation is repeated until one ends with a subfamily $\mathcal{F} \subset \mathcal{F}_{\text{locmin}}(A, K)$ that is pairwise unconnected. In view of Theorem 4(a), $\mathcal{F}$ produces as many local minimal values as $\mathcal{F}_{\text{locmin}}(A, K)$. Hence, $\text{card}[\sigma_{\text{locmin}}(A, K)] \leq \text{card}[\mathcal{F}] \leq \nu(K)$. □

Example 3 Consider again a cone K generated by a finite collection $\{g^1, \ldots, g^p\}$ of unit vectors in $\mathbb{R}^3$, cf. Fig. 3. Assume that none of the generators can be expressed as positive linear combination of the others. The multichromatic number of this cone is $\nu(K) = p$. If

$$\langle g^i, g^j \rangle \geq 0 \quad \forall i, j \in \{1, \ldots, p\},$$

then K is infra-dual and Proposition 12 yields the upper bound $\text{card}[\sigma_{\text{locmin}}(A, K)] \leq p$ for all $A \in \text{Sym}(3)$. This bound is better than (27), but this is not surprising because we are now asking K to be infra-dual.

Corollary 6 *Let $A \in \text{Sym}(n)$ and $K \in \Xi(\mathbb{R}^n)$ be an infra-dual convex cone generated by p linearly independent unit vectors in $\mathbb{R}^n$. Then, the cardinality of $\sigma_{\text{locmin}}(A, K)$ cannot exceed*

$$\nu_p = \frac{p!}{\lfloor p/2 \rfloor! \, (p - \lfloor p/2 \rfloor)!}, \tag{41}$$

where $\lfloor r \rfloor$ stands for the lower integer part of r.

Table 1 ν_n versus $2^n - 1$

n	$2^n - 1$	ν_n
3	7	3
4	15	6
5	31	10
20	1.0×10^6	1.8×10^5
30	1.1×10^9	1.6×10^8
40	1.1×10^{12}	1.4×10^{11}

Figures are approximate for large values of n

Proof In view of the linear independence hypothesis, the graph associated to K is the same as the graph associated to the Pareto cone $\mathbb{R}^p_+$. The corollary is then a consequence of Proposition 12 and the fact that ν_p is the multichromatic number of $\mathbb{R}^p_+$. Note that (41) corresponds to the cardinality of

$$\mathcal{F} \equiv \lfloor p/2 \rfloor\text{-dimensional faces of } \mathbb{R}^p_+,$$

a family that achieves the maximum in the definition of $\nu(\mathbb{R}^p_+)$. □

Corollary 6 applies in particular to $p = n$, that is to say, when K is a simplicial cone in $\mathbb{R}^n$. For a large dimension n one can use Stirling's approximation formula $n! \approx \sqrt{2\pi n}\,(n/e)^n$ in order to obtain

$$\nu_n \approx \sqrt{\frac{2}{\pi}}\,\frac{2^n}{\sqrt{n}}\,.$$

Although ν_n grows less rapidly than 2^n, the damping factor $\sqrt{n}$ is not strong enough to prevent a growth of exponential type (Table 1).

The next corollary is somewhat different in spirit. The idea is getting something better than (41) in case we have already detected a particular active face. The next result is of special interest if the detected face has low dimension or, on the contrary, a dimension close to $\dim K$.

Corollary 7 *Let $A \in \mathrm{Sym}(n)$ and $K \in \Xi(\mathbb{R}^n)$ be an infra-dual convex cone generated by p linearly independent unit vectors in $\mathbb{R}^n$. If the variational problem (1) admits an active face of dimension $d \in \{1, \ldots, p\}$, then*

$$\mathrm{card}[\sigma_{\mathrm{locmin}}(A, K)] \;\leq\; 2^p - 2^d - 2^{p-d} + 2.$$

Proof Let F be an active face as indicated above. Given that F is d-dimensional, one clearly has

$$\begin{aligned} 2^d - 2 &= \text{number of non-trivial faces strictly contained in } F,\\ 2^{p-d} - 1 &= \text{number of faces strictly containing } F. \end{aligned}$$

These $(2^d - 2) + (2^{p-d} - 1)$ faces are connected to F. If a particular face in this group is active, then it produces the same local minimal value as F. So, when it comes to bound the cardinality of $\sigma_{\mathrm{locmin}}(A, K)$, we can throw away this redundant group and keep only

$$2^p - 1 - [(2^d - 2) + (2^{p-d} - 1)] = 2^p - 2^d - 2^{p-d} + 2$$

faces in all. □

7 Tightness of the cone constraint

Sufficient conditions for local optimality in cone-constrained optimization problems usually invoke an assumption called strict complementarity. The question that is being addressed in this section reads as follows:

> Under which assumption one can infer that a given normalized K-eigenvector of A is a local solution to the variational problem (1)?

In order to answer this question we distinguish between two different situations. The first, and simplest, situation occurs when the normalized K-eigenvector, say x, lies in the relative interior of K. In such a case, in view of Proposition 3 and Remark 3, one has

$$\begin{aligned} x \text{ is a local solution to (1)} &\Longleftrightarrow \langle x, Ax\rangle = \lambda_{\min}(A, \operatorname{span} K) \\ &\Longleftrightarrow x \text{ is a global solution to (1).} \end{aligned}$$

The second situation to be considered occurs when the K-eigenvector under examination is a unit vector in the relative boundary of K. In order to handle this more interesting case we introduce a suitable notion of strict complementarity that we call tightness. Our discussion takes place in a context where constraints are defined by a polyhedral convex cone K. A leading role is played by the canonical correspondence

$$\begin{aligned} \Delta_K : \mathcal{F}(K) &\to \mathcal{F}(K^+) \\ F \quad &\mapsto \Delta_K(F) = [\operatorname{span} F]^\perp \cap K^+ \end{aligned} \tag{42}$$

between the faces of K and those of K^+. For general results on the map (42) the reader can consult the survey paper by Barker [1] or the book by Ziegler [17]. The work by Tam [16] is also a good source of information.

Lemma 3 *If $K \in \Xi(\mathbb{R}^n)$ is polyhedral, then* $\operatorname{span}[\Delta_K(F)] = [\operatorname{span} F]^\perp$ *for all* $F \in \mathcal{F}(K)$.

Proof Observe that $\Delta_K(F) = (K + \operatorname{span} F)^+$. That K is polyhedral ensures the closedness of $K + \operatorname{span} F$. It is not difficult to check that

$$K + \operatorname{span} F = \mathcal{T}_K(x) \quad \forall x \in \operatorname{ri}(F), \tag{43}$$

where $\mathcal{T}_K(x) = \mathbb{R}_+(K - x)$. On the other hand, one has

$$\mathcal{T}_K(x) \cap -\mathcal{T}_K(x) = \operatorname{span} F \quad \forall x \in \operatorname{ri}(F). \tag{44}$$

We shall prove only $\mathcal{T}_K(x) \cap -\mathcal{T}_K(x) \subset \operatorname{span} F$, the reverse inclusion being trivial. Take a non-zero vector h in $\mathcal{T}_K(x) \cap -\mathcal{T}_K(x)$, i.e.,

$$\begin{aligned} h &= \alpha_1(u_1 - x) \\ -h &= \alpha_2(u_2 - x) \end{aligned}$$

with $\alpha_1, \alpha_2 > 0$ and $u_1, u_2 \in K$. This yields in particular

$$x = \left(\frac{\alpha_1}{\alpha_1 + \alpha_2}\right) u_1 + \left(\frac{\alpha_2}{\alpha_1 + \alpha_2}\right) u_2.$$

Since F is a face of K, it follows that $u_1, u_2 \in F$. Hence, $h \in \operatorname{span} F$. Finally, by combining (43) and (44), one obtains

$$\begin{aligned} \operatorname{span}[\Delta_K(F)] &= (K + \operatorname{span} F)^+ - (K + \operatorname{span} F)^+ = (\mathcal{T}_K(x))^+ - (\mathcal{T}_K(x))^+ \\ &= [\mathcal{T}_K(x) \cap -\mathcal{T}_K(x)]^\perp = [\operatorname{span} F]^\perp. \end{aligned}$$

This completes the proof of the lemma. □

That $x \in K \cap \mathbb{S}_n$ is a K-eigenvector of A can be expressed in the compact form

$$Ax - \langle x, Ax\rangle x \in \Delta_K(F) \tag{45}$$

with F being the face associated to x. Strict complementarity or tightness is an hypothesis expressing that $Ax - \langle x, Ax\rangle x$ belongs to the relative interior of $\Delta_K(F)$.

What must be added to (45) in order to conclude that x is a local solution to (1)? The next theorem answers this question and indicates the size of a ball

$$\mathbb{B}_n(x, r) = \{u \in \mathbb{R}^n : \|u - x\| \leq r\}$$

around x over which local minimality takes place.

Theorem 5 *Suppose that $A \in \mathrm{Sym}(n)$ is not a multiple of the identity matrix, that $K \in \Xi(\mathbb{R}^n)$ is polyhedral, and that x is a unit vector in K whose associated face F is not equal to K. Under the following hypotheses*

(i) $Ax - \langle x, Ax\rangle x \in \mathrm{ri}[\Delta_K(F)]$ *(tightness),*
(ii) $\langle x, Ax\rangle = \lambda_{\min}(A, \mathrm{span}F)$ *(minimality within the face),*

one can infer that x is a local solution to (1). More precisely,

$$\langle x, Ax\rangle \leq \langle u, Au\rangle \quad \forall u \in K \cap \mathbb{S}_n \cap \mathbb{B}_n(x, r),$$

where the radius r is given by

$$r = \frac{2\gamma}{\sqrt{4\alpha^2 + \beta^2}}. \tag{46}$$

Here

$$\gamma = \min_{\substack{v \in [\mathrm{span}F]^\perp \cap [\Delta_K(F)]^+ \\ \|v\|=1}} \langle Ax - \langle x, Ax\rangle x, v\rangle \tag{47}$$

is a positive real number that measures the "degree of tightness" of x, and

$$\alpha = \max_{\substack{h \in \mathrm{span}F \\ \|h\|=1}} \|(A - \langle x, Ax\rangle I)h\|, \tag{48}$$

$$\beta = \max_{\substack{h \in [\mathrm{span}F]^\perp \\ \|h\|=1}} \|(A - \langle x, Ax\rangle I)h\|, \tag{49}$$

are the operator norms of the restrictions of $A - \langle x, Ax\rangle I$ to the linear subspaces $\mathrm{span}F$ *and* $[\mathrm{span}F]^\perp$*, respectively.*

Proof Since F is different from K itself, one has $\Delta_K(F) \neq \{0\}$. In view of Lemma 3, a consequence of the polyhedrality of K is that the tightness assumption (i) is equivalent to the interiority condition

$$Ax - \langle x, Ax\rangle x \in \mathrm{int}_{[\mathrm{span}F]^\perp}[\Delta_K(F)] \tag{50}$$

with $\mathrm{int}_{[\mathrm{span}F]^\perp}[\Delta_K(F)]$ standing for the interior of $\Delta_K(F)$ relative to $[\mathrm{span}F]^\perp$. For the sake of convenience we introduce the notation

$$L = \mathrm{span}F, \quad M = [\mathrm{span}F]^\perp, \quad \lambda = \langle x, Ax\rangle, \quad A^x = A - \langle x, Ax\rangle I.$$

Consider a vector $u \in K \cap \mathbb{S}_n \cap \mathbb{B}_n(x, r)$ different from x. Write

$$u - x = d = d_L + d_M$$

with d_L and d_M denoting the orthogonal projections of d onto L and M, respectively. If $d_M = 0$, then u remains in span$F$ and, in view of the hypothesis (ii), one gets

$$\lambda = \lambda_{\min}(A, \text{span}F) \leq \langle u, Au \rangle.$$

So, there is no loss of generality in assuming that $d_M \neq 0$. One has

$$\begin{aligned} \langle u, Au \rangle &= \langle x, Ax \rangle + 2\langle Ax, d \rangle + \langle d, Ad \rangle \\ &= \lambda + 2\langle Ax, d_L \rangle + 2\langle Ax, d_M \rangle + \langle d, Ad \rangle. \end{aligned} \tag{51}$$

But

$$\langle Ax, d_L \rangle = \langle \underbrace{Ax - \lambda x}_{\text{in } M}, d_L \rangle + \lambda \langle x, d_L \rangle = \lambda \langle x, d_L \rangle = \lambda \langle x, d \rangle. \tag{52}$$

Plugging (52) into (51) one gets

$$\begin{aligned} \langle u, Au \rangle &= \lambda + 2\lambda \langle x, d \rangle + 2\langle Ax, d_M \rangle + \langle d, Ad \rangle \\ &= \lambda \|x + d\|^2 + 2\langle Ax, d_M \rangle + \langle d, (A - \lambda I)d \rangle \\ &= \lambda + 2\langle Ax, d_M \rangle + \langle d, A^x d \rangle. \end{aligned}$$

Let us examine separately the terms $2\langle Ax, d_M \rangle$ and $\langle d, A^x d \rangle$. We shall prove that their sum is non-negative. Our first observation is that $d_M \in M \cap [\Delta_K(F)]^+$. To see that $d_M \in [\Delta_K(F)]^+$, take $w \in \Delta_K(F) = M \cap K^+$ and write

$$\langle w, d_M \rangle = \langle w, d - d_L \rangle = \langle w, d \rangle = \langle w, u \rangle - \langle w, x \rangle = \langle w, u \rangle \geq 0.$$

Hence,

$$\langle Ax, d_M \rangle = \langle Ax - \lambda x, d_M \rangle \geq \gamma \, \|d_M\| \tag{53}$$

with γ being defined by (47). We claim that $\gamma > 0$. One can view $\Delta_K(F)$ and $M \cap [\Delta_K(F)]^+$ as mutually dual cones in the Hilbert space $(M, \langle \cdot, \cdot \rangle)$. Notice that $Ax - \lambda x$ belongs to the interior of $\Delta_K(F)$ (relative to the underlying space M). In particular, $Ax - \lambda x \neq 0$ and

$$\langle Ax - \lambda x, v \rangle > 0$$

for any unit vector v in $M \cap [\Delta_K(F)]^+$. A compactness argument shows that the infimum in (47) is attained and confirms the positivity of γ. Finally, let us examine the term

$$\langle d, A^x d \rangle = \langle d_L, A^x d_L \rangle + 2\langle d_M, A^x d_L \rangle + \langle d_M, A^x d_M \rangle.$$

As a consequence of the hypothesis (ii), the matrix A^x is positive semidefinite over the space L, i.e., $\langle d_L, A^x d_L \rangle \geq 0$. On the other hand,

$$\begin{aligned} |2\langle d_M, A^x d_L \rangle + \langle d_M, A^x d_M \rangle| &\leq 2|\langle d_M, A^x d_L \rangle| + |\langle d_M, A^x d_M \rangle| \\ &\leq \left(2\|A^x d_L\| + \|A^x d_M\|\right) \|d_M\| \\ &\leq (2\alpha \|d_L\| + \beta \|d_M\|) \, \|d_M\| \\ &\leq (\sqrt{4\alpha^2 + \beta^2} \, \|d\|) \|d_M\| \\ &\leq 2\gamma \, \|d_M\|. \end{aligned} \tag{54}$$

The combination of (53) and (54) shows that $2\langle Ax, d_M\rangle + \langle d, A^x d\rangle \geq 0$ and completes the proof. □

Remark 6 Theorem 5 can be stated also in a non-polyhedral setting but one must consider (50) as definition of tightness. We restricted our attention to the polyhedral case because the interiority hypothesis (50) has little chance to be realised if the boundary of K posseses some kind of "curvature". The best way of understanding this point is by considering the Lorentz cone

$$K = \{x \in \mathbb{R}^n : x_n \geq [x_1^2 + \cdots + x_{n-1}^2]^{1/2}\} \tag{55}$$

which is the prototype of a non-polyhedral convex cone. This cone is self-dual. Excepting for $\{0\}$ and the cone K itself, any other face F of (55) is one-dimensional. Note that $\Delta_K(F)$ is a half-line and therefore it has an empty interior relative to the hyperplane $[\text{span}F]^{\perp}$.

The tightness condition stated in Theorem 5 looks highly technical and difficult to check in practice. However, this is not always so. For instance, in the paretian case

$$\begin{aligned} &\text{minimize } \langle u, Au\rangle \\ &u \geq 0, \ \|u\| = 1 \end{aligned} \tag{56}$$

the tightness condition takes a very simple form and the tightness coefficient γ can be easily computed. The contraint $u \geq 0$ in (56) expresses the fact that each component of the vector u is non-negative.

The next corollary summarizes what we know about the paretian case. Corollary 8 is obtained by combining Theorems 1 and 5. Given a non-empty index set $J \subset \{1, \ldots, n\}$, the symbol A^J stands for the principal submatrix of A formed with the rows and columns of A indexed by J.

Corollary 8 *Let $A \in \text{Sym}(n)$. For $x \in \mathbb{R}^n$ to be a local solution to the variational problem (56) it is necessary (respectively, sufficient) that*

$$x_j = \begin{cases} z_j \ \ if \ \ j \in J, \\ 0 \ \ \ if \ \ j \notin J \end{cases} \tag{57}$$

for some non-empty index set $J \subset \{1, \ldots, n\}$ and some unit vector $z \in \mathbb{R}^{\text{card}J}$ satisfying the Perron-type eigenvalue problem

$$A^J z = \lambda_1(A^J)z, \quad z_j > 0 \ \forall j \in J \tag{58}$$

and

$$\sum_{j\in J} A_{ij} z_j \geq 0 \quad \forall i \notin J \tag{59}$$

$$(respectively, \quad \sum_{j\in J} A_{ij} z_j > 0 \quad \forall i \notin J). \tag{60}$$

Furthermore, the corresponding local minimal value is given by $\langle x, Ax\rangle = \langle z, A^J z\rangle = \lambda_1(A^J)$.

Proof Let $\{e_1, \ldots, e_n\}$ denote the canonical basis of the space $\mathbb{R}^n$. The index set J can be identified with the face $F = \text{cone}\{e_j : j \in J\}$ of the n-dimensional Pareto cone. By keeping in mind this identification, one sees that (57) corresponds to the transfer equation (33), the Perron-type eigenvalue problem (58) reflects pre-activity, and (59) is part of the criticality condition (2). This takes care of the necessary condition of local optimality. As far as sufficiency is concerned, observe that (60) corresponds to the tightness condition stated in Theorem 5. □

Remark 7 If $J = \{1, \ldots, n\}$, then (60) holds vacuosly and we are dealing in fact with a global solution. A more interesting situation occurs when $J \neq \{1, \ldots, n\}$. In such a case the tightness coefficient (47) is expressible in the form

$$\gamma = \min_{i \notin J} \sum_{j \in J} A_{ij} z_j = \min_{i \notin J} (Ax)_i \, .$$

The operator norms (48) and (49) are also easily computable. Hence, one can evaluate without troubles the radius (46) of the ball over which local minimality takes place.

References

1. Barker, G.P.: Theory of cones. Linear Algebra Appl. **39**, 263–291 (1981)
2. Berman, A., Plemmons, R.J.: Nonnegative Matrices in the Mathematical Sciences. SIAM Publications, Philadelphia (1994)
3. Conrad, F., Brauner, C.M., Issard-Roch, F., Nicolaenko, B.: Nonlinear eigenvalue problems in elliptic variational inequalities: a local study. Comm. Partial Differential Equations **10**, 151–190 (1985)
4. Diestel, R.: Graph Theory. Graduate Texts in Mathematics 173. Springer-Verlag, New York (1997)
5. Do, C.: Problèmes de valeurs propres pour une inéquation variationnelle sur un cône et application au flambement unilatéral d'une plaque mince. C. R. Acad. Sci. Paris **280**, 45–48 (1975)
6. Iusem, A., Seeger, A.: On vectors achieving the maximal angle of a convex cone. Math. Program. **104**, 501–523 (2005)
7. Iusem, A., Seeger, A.: On convex cones with infinitely many critical angles. Optimization **56**, 115–128 (2007)
8. Iusem, A., Seeger, A.: Searching for critical angles in a convex cone. Math. Program. (2008, in press) available online at doi:10.1007/s10107-007-0146-0
9. Klarbring, A.: On discrete and discretized nonlinear elastic structures in unilateral contact: stability, uniqueness and variational principles. Int. J. Solids Structures **24**, 459–479 (1988)
10. Pinto da Costa, A., Figueiredo, I.N., Judice, J.A., Martins, J.A.C.: A complementarity eigenproblem in the stability of finite dimensional elastic systems with frictional contact. Complementarity: Applications, Algorithms and Extensions, pp. 67–83. Applied Optimization Series 50. M. Ferris, O. Mangasarian and J.S. Pang (Eds), Kluwer Acad. Publ., Dordrecht, 1999
11. Pinto da Costa, A., Martins, J.A.C., Figueiredo, I.N., Judice, J.J.: The directional instability problem in systems with frictional contacts. Comput. Methods Appl. Mech. Engrg. **193**, 357–384 (2004)
12. Pinto da Costa, A., Seeger, A.: Cone-constrained eigenvalue problems. Part I: theory. Submitted.
13. Riddell, R.C.: Eigenvalue problems for nonlinear elliptic variational inequalities on a cone. J. Functional Analysis **26**, 333–355 (1977)
14. Seeger, A.: Eigenvalue analysis of equilibrium processes defined by linear complementarity conditions. Linear Algebra Appl. **292**, 1–14 (1999)
15. Seeger, A., Torki, M.: On eigenvalues induced by a cone constraint. Linear Algebra Appl. **372**, 181–206 (2003)
16. Tam, B.S.: On the duality operator of a convex cone. Linear Algebra Appl. **64**, 33–56 (1985)
17. Ziegler, G.M.: Lectures on Polytopesj. Springer-Verlag, New York (1995)