

A Fresh Variational-Analysis Look at the Positive Semidefinite Matrices World

Jean-Baptiste Hiriart-Urruty · Jérôme Malick

Received: 11 October 2011 / Accepted: 12 December 2011 / Published online: 5 January 2012
© Springer Science+Business Media, LLC 2012

Abstract Engineering sciences and applications of mathematics show unambiguously that positive semidefiniteness of matrices is the most important generalization of non-negative real numbers. This notion of non-negativity for matrices has been well-studied in the literature; it has been the subject of review papers and entire chapters of books.

This paper reviews some of the nice, useful properties of positive (semi)definite matrices, and insists in particular on (i) characterizations of positive (semi)definiteness and (ii) the geometrical properties of the set of positive semidefinite matrices. Some properties that turn out to be less well-known have here a special treatment. The use of these properties in optimization, as well as various references to applications, is spread all the way through.

The “raison d’être” of this paper is essentially pedagogical; it adopts the viewpoint of variational analysis, shedding new light on the topic. Important, fruitful, and subtle, the positive semidefinite world is a good place to start with this domain of applied mathematics.

Keywords Positive semidefiniteness · Optimization · Convex analysis · Eigenvalues · Spectral functions · Riemannian geometry

Communicated by Michel Théra.

J.-B. Hiriart-Urruty
Institut de Mathématiques, Université Paul Sabatier, Toulouse, France
e-mail: jbhu@cict.fr

J. Malick (✉)
Lab. J. Kuntzmann, CNRS, Grenoble, France
e-mail: jerome.malick@inria.fr

1 Positive (Semi)Definiteness: A Useful Notion

Positive (semi)definite matrices are fundamental objects in applied mathematics and engineering. For example, they appear as covariance matrices in statistics, as Lyapunov functions in control, as kernels in machine learning, as diffusion tensors in medical imaging, as lifting matrices in combinatorics—just to name a few of the various uses of the concept.

Even if a first generalization of positive real numbers in matrix spaces would be the entry-wise positive matrices (see [1, Chap. 8]), the numerous applications mentioned above show that the most important generalization is the positive semidefinite matrices. This notion of non-negativity for matrices has been well-studied in the literature; it has been the topic of review papers and entire chapters of books, for example, [1, Chap. 7], [2, Chap. 6], [3, Chap. 8], [4, Chap. 1] and [5, Sect. 6.5]. Besides, a fruitful branch of mathematical optimization is devoted to problems involving positive semidefinite matrices; it is called *semidefinite programming* and has proved to be a powerful tool to model and solve many problems in engineering or science in general [6, 7].

In this article, we gather some results about positive (semi)definite matrices: some are well-known, some are more original. The special feature of this paper is that we adopt the viewpoint of variational analysis, shedding new light on these topics. The term “variational” is not to be understood in its old historical meaning (calculus of variations), but in the broadest possible sense (optimization, convex analysis, nonsmooth analysis, complementary problems, . . .). We insist in particular on the harmonious interplay between matrix analysis and optimization.

The set of positive semidefinite matrices is a closed, convex cone in the space of symmetric matrices, which enjoys a nice geometry, exploited in several applications. We review some of these useful and sometimes subtle geometrical properties—again with a variational eye. During this study, we will illustrate convex, semialgebraic and Riemannian geometries. As a nice, concrete example of advanced mathematics, positive semidefiniteness also has some pedagogical interest—on top of all its practical uses.

2 Positive (Semi)Definiteness: A Multi-facet Notion

We start with a bunch of simple ideas. The goal of this first section is, while sticking to basic properties, to show how different notions and various domains come into play when talking about positive semidefinite matrices. We also recall the background properties, introduce notation, and give a flavour of what comes next.

Let $\mathcal{S}^n(\mathbb{R})$ be the linear space of symmetric real matrices of size $n \times n$. A matrix $A \in \mathcal{S}^n(\mathbb{R})$ is said to be *positive semidefinite* (denoted $A \succeq 0$), if

$$x^\top Ax \geq 0 \quad \text{for all } x \in \mathbb{R}^n, \quad (1)$$

where the superscript \cdot^\top denotes the transposition of matrices. The matrix A is further called *positive definite* (denoted by $A \succ 0$) if the above inequality is strict for all nonzero x .

2.1 (Bi)Linear Algebra at Work

Let us recall the key-result for symmetric matrices: the n eigenvalues λ_i of a matrix $A \in \mathcal{S}^n(\mathbb{R})$ lie in \mathbb{R} , and there exists an orthogonal matrix U such that $U^\top AU$ is the diagonal matrix with the λ_i on the diagonal, denoted by $\text{diag}(\lambda_1, \dots, \lambda_n)$. This result can also be written as follows: there exist n unit eigenvectors x_i of A (the columns of U) such that we have the so-called *spectral decomposition*

$$A = \sum_{i=1}^n \lambda_i x_i x_i^\top.$$

This result mixes nicely linear and bilinear algebra: U allows both the diagonalization of A as $U^{-1}AU = \text{diag}(\lambda_1, \dots, \lambda_n)$ (linear algebra world), and the reduction of the quadratic form associated with A as $U^\top AU = \text{diag}(\lambda_1, \dots, \lambda_n)$, such that $\langle A(Uy), Uy \rangle = \langle \text{diag}(\lambda_1, \dots, \lambda_n)y, y \rangle$ (bilinear algebra world). There are several proofs of this result; one is optimization-based (see, e.g. [8, p. 222]).

This decomposition is used to define the square root matrix: if $A \geq 0$, the square root of A , denoted by $A^{1/2}$, is the unique $S \geq 0$ such that $S^2 = A$ (given explicitly by the decomposition $U^\top S U = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n})$). When $A > 0$, we also have $(A^{1/2})^{-1} = (A^{-1})^{1/2}$ so that the notation $A^{-1/2}$ (that we will often use) is not ambiguous.

We can also reduce two matrices at the same time: the so-called *simultaneous reduction* (see, e.g. [1, 4.6.12]) says that for a symmetric matrix A and a positive definite matrix B , there exists an invertible matrix P such that

$$P^\top A P = \text{diag}(\lambda_1, \dots, \lambda_n) \quad \text{and} \quad P^\top B P = \text{diag}(\mu_1, \dots, \mu_n).$$

We are not aware of simple results of that kind for more than two matrices; the reason might be deeper than one would think (see Problem 12 in [9]).

2.2 Inner Products

The space \mathbb{R}^n is equipped with the canonical inner product

$$\langle x, y \rangle := \sum_{i=1}^n x_i y_i = x^\top y.$$

We prefer to use the notation $\langle \cdot, \cdot \rangle$, often more handy for calculation involving the transpose. Up to a double sum, this inner product is the same for symmetric matrices: we equip $\mathcal{S}^n(\mathbb{R})$ with the canonical inner product

$$\langle\langle A, B \rangle\rangle := \sum_{i,j=1}^n A_{ij} B_{ij} = \text{trace}(A^\top B) = \text{trace}(AB).$$

A connection between these two scalar products is

$$\langle Ax, x \rangle = \langle\langle A, xx^\top \rangle\rangle.$$

This easy property has an important role in combinatorial optimization: the so-called “lifting” is the first step of the *semidefinite relaxation* of binary quadratic problems (see, e.g. [10]). Another interesting connection between the inner products is the Ky–Fan inequality: $A, B \in \mathcal{S}^n(\mathbb{R})$ satisfy the inequality

$$\langle\langle A, B \rangle\rangle \leq \langle \lambda(A), \lambda(B) \rangle,$$

where $\lambda(A) := (\lambda_1(A), \dots, \lambda_n(A)) \in \mathbb{R}^n$ is the vector of the eigenvalues of A ordered nonincreasingly. Moreover equality holds in the above inequality if and only if A and B admit a simultaneous eigendecomposition with ordered eigenvalues (see, e.g. [11] and references therein). This inequality is a key tool for the study of spectral functions and spectral sets [12]; we will come back to spectral sets from time to time in next sections.

2.3 Quadratic Forms and Differential Calculus

Naturally associated with $A \in \mathcal{S}^n(\mathbb{R})$ is the quadratic form

$$q: x \in \mathbb{R}^n \mapsto q(x) := \frac{1}{2} \langle Ax, x \rangle.$$

The definition of semidefiniteness (1) reads $q(x) \geq 0$ for all $x \in \mathbb{R}^n$. From an optimization point of view, we have

$$\inf_{x \in \mathbb{R}^n} q(x) > -\infty \iff A \succeq 0$$

and the set of minimizers is then $\ker A$. The quadratic form q is obviously of class C^∞ on \mathbb{R}^n , and we have for all $x \in \mathbb{R}^n$

$$\nabla q(x) = Ax \quad \text{and} \quad \nabla^2 q(x) = A. \quad (2)$$

This is easy to remember: we formally have the same formulae for $q(x) = ax^2/2$ with $x \in \mathbb{R}$. Note that the factor 1/2 in the definition of q aims at avoiding the factor 2 when differentiating q .

When A is positive definite, q defines a norm on \mathbb{R}^n . This can be seen, for example, by the simultaneous reduction: there exists P invertible such that $P^\top AP = I_n$, or in other words, up to the change of variables $y = P^{-1}x$, the quadratic form q is just the simpler form $\| \cdot \|^2$:

$$\|y\|^2 = \langle y, y \rangle = \langle P^\top APy, y \rangle = \langle APy, Py \rangle = \langle Ax, x \rangle.$$

2.4 When Convex Analysis Enters into the Picture

Recall that a function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is convex if

$$\forall x, y \in \mathbb{R}^n, \alpha \in [0, 1], \quad f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y).$$

When f is smooth, convexity is characterized by the positive semidefiniteness of its Hessian matrix $\nabla^2 f(x)$ for all $x \in \mathbb{R}^n$ (see, e.g. [13, B.4.3.1]). From (2), we get that

q is convex if and only if $A \succeq 0$. (Let us also mention that strict and strong convexities of q coincide and are equivalent to $A \succ 0$.)

During our visit to the world of the positive semidefinite matrices, we will meet several notions from convex analysis (polar cone, Moreau decomposition, tangent and normal cones, ...). The first notion is the Legendre–Fenchel transformation, which is an important transformation in convex analysis (see [13, Chap. E]) introducing a duality in convex optimization. The Legendre–Fenchel transformation defined by

$$q^* : s \in \mathbb{R}^n \mapsto q^*(s) := \sup_{x \in \mathbb{R}^n} \{ \langle s, x \rangle - q(x) \},$$

has a simple explicit expression for $q(x) = \langle Ax, x \rangle / 2$ (with $A \succeq 0$). In the case $A \succ 0$, we obtain

$$q^*(s) = \frac{1}{2} \langle A^{-1}s, s \rangle.$$

Incidentally, this expression allows us to establish $B^{-1} \preceq A^{-1}$ if $B \succeq A$, with a quick variational proof (first, use the fact that $q \leq p$ implies $q^* \geq p^*$, and then the fact that $B - A \succeq 0$). This latter result is the matrix counterpart of the decreasing of the inverse for positive numbers. Note, moreover, that the involution $A \mapsto A^{-1}$ of linear algebra corresponds to the involution $q \mapsto q^*$ of convex analysis.

If we just have $A \succeq 0$, then (see, e.g. [13, E.1.1.4])

$$q^*(s) = \begin{cases} \frac{1}{2} \langle x_0, s \rangle & \text{where } Ax_0 = s \text{ if } s \in \text{Im } A, \\ +\infty & \text{if } s \notin \text{Im } A. \end{cases}$$

Note, finally, that making the transformation twice, we come back home: $(q^*)^* = q$, as expected.

2.5 Do not Forget Geometry

A geometrical view is often used to introduce linear algebra (as in [5]). The geometrical object also associated with a matrix $A \succ 0$ is the ellipsoidal convex compact set

$$E_A := \{ x \in \mathbb{R}^n : \langle Ax, x \rangle \leq 1 \}.$$

The eigenvalues of A give the shape of E_A (see [5, Chap. 6]); it is the unit ball of the norm given by A . The linear function $A^{-1/2}$ maps the canonical unit ball B to E_A ; so the volume of E_A is $1/\sqrt{\det(A)}$ times the volume of B (see Problem 3 in [9] for a question on the volume of convex bodies). We mention that the inversion $A \mapsto A^{-1}$ coincides for E_A with the polarity transformation of convex compact bodies having 0 in the interior (see [13, Chap. C]).

2.6 Convex Cone, Visualization

The set of positive semidefinite matrices

$$\mathcal{S}_+^n(\mathbb{R}) = \{ A \in \mathcal{S}^n(\mathbb{R}) : x^\top Ax \geq 0 \text{ for all } x \in \mathbb{R}^n \} \tag{3}$$

enjoys a nice geometry. We give here a first flavour of its geometry when $n = 2$ and we will come back to this topic again in Sect. 4.

It is not difficult to show by means of definition (3) that $\mathcal{S}_+^n(\mathbb{R})$ is a (closed) convex cone. Recall that this means

$$\begin{aligned} (\alpha \in [0, 1], A, B \in \mathcal{S}_+^n(\mathbb{R})) &\implies \alpha A + (1 - \alpha)B \in \mathcal{S}_+^n(\mathbb{R}), \\ (\alpha \geq 0, A \in \mathcal{S}_+^n(\mathbb{R})) &\implies \alpha A \in \mathcal{S}_+^n(\mathbb{R}). \end{aligned}$$

Since $\mathcal{S}^2(\mathbb{R})$ is of dimension 3, we can visualize the cone $\mathcal{S}_+^2(\mathbb{R})$ as a subset of \mathbb{R}^3 . We identify $\mathcal{S}^2(\mathbb{R})$ and \mathbb{R}^3 by an isomorphism

$$\varphi: A = \begin{bmatrix} a & b \\ b & c \end{bmatrix} \in \mathcal{S}^2(\mathbb{R}) \longmapsto \varphi(A) \in \mathbb{R}^3.$$

We can choose the isometry $\varphi(A) = (a, \sqrt{2}b, c)$, but the best looking isomorphism is

$$\varphi(A) := \frac{1}{\sqrt{2}}(2b, c - a, c + a)$$

that allows us to identify the cone $\mathcal{S}^2(\mathbb{R})$

$$\left\{ \begin{bmatrix} a & b \\ b & c \end{bmatrix} : a \geq 0, c \geq 0, \text{ and } ac - b^2 \geq 0 \right\}$$

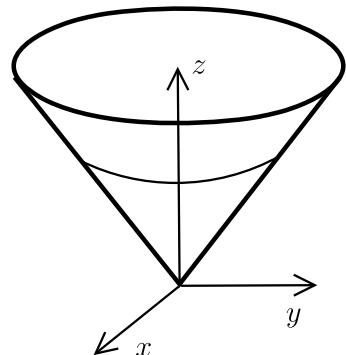
with $\varphi(\mathcal{S}^2(\mathbb{R}))$, which is

$$\{(x, y, z) \in \mathbb{R}^3, z \geq \sqrt{x^2 + y^2}\}.$$

Figure 1 shows that $\varphi(\mathcal{S}^2(\mathbb{R}))$ is the “usual” cone K of \mathbb{R}^3 , also called Lorentz cone or the second-order cone in \mathbb{R}^3 .

The boundary of K corresponds to the rank-one matrices, and the apex of K to the zero matrix. To avoid giving an incorrect intuition, we emphasize that even if the boundary of $\mathcal{S}_+^2(\mathbb{R})$ deprived of 0 appears to be a smooth manifold, this is not true in general when $n > 2$ ($\mathcal{S}_+^n(\mathbb{R})$ is, in fact, a union of smooth manifolds; see Sect 4.4).

Fig. 1 The usual cone in \mathbb{R}^3 , identified with $\mathcal{S}_+^2(\mathbb{R})$



3 Characterizations of Positive (Semi)Definiteness

3.1 Characterizations of Positive Definiteness

The characterizations of positive definiteness are numerous; each of them has its own usefulness, depending on the context. For example, the Gram formulation appears naturally in statistics, the characterization by invariants in mechanics, or the exponential in dynamical systems. We discuss some of these characterizations here: we start with a couple of characterizations that use decompositions; then, we put emphasis on the characterizations by the positivity of n real numbers; we finish with a more original characterization.

3.1.1 Factorization-Like Characterizations

One of the most useful and most basic characterizations is the following factorization: A is positive definite if and only if there exists an invertible B such that $A = BB^T$ (indeed, in this case $\langle Ax, x \rangle = \|Bx\|^2$ for all $x \in \mathbb{R}^n$). There exist in fact different factorizations of this form, among those: the Cholesky factorization with B upper-triangular with positive diagonal elements; and the square root factorization, with B symmetric positive definite (see, e.g. [2, Chap. 6]).

Other famous decomposition-like characterizations use exponential or Gram matrices: the property $A \succ 0$ is characterized by each of the two properties:

- Exponential form (see, e.g. [14, Examples 5–28]): there exists $B \in \mathcal{S}^n(\mathbb{R})$ such that $A = \exp(B)$.
- Gram form (see, e.g. [1, 7.2.40]): there exist n independent vectors $\{v_1, \dots, v_n\}$ of \mathbb{R}^n (hence forming a basis of \mathbb{R}^n) such that $A_{ij} = \langle v_i, v_j \rangle$ for all i, j .

3.1.2 Positivity of n Real Numbers

Important characterizations of positive definiteness rely on the positivity of n real numbers associated with symmetric matrices. For example, it is well-known that positive eigenvalues characterize positive definiteness; we gather in the next theorem three similar properties.

Theorem 3.1 (Positive definiteness by positivity of n real numbers) *Each of the following properties is equivalent to $A \succ 0$:*

- (i) *The eigenvalues $\lambda_1(A), \dots, \lambda_n(A)$ of A are positive;*
- (ii) *The leading principal minors $\Delta_1(A), \dots, \Delta_n(A)$ of A are positive;*
- (iii) *The principal invariants $i_1(A), \dots, i_n(A)$ are positive.*

We recall that the leading principal minors are defined as $\Delta_k(A) := \det A_k$ for $k = 1, \dots, n$, where A_k is the submatrix of A made of the first k rows and first k columns of A . Recall also that the principal invariants of A are (up to a change of sign) the coefficients of the monomials of the characteristic polynomial $P_A(x)$ of A ; more precisely

$$P_A(x) = (-x)^n + i_1(A)(-x)^{n-1} + \dots + i_k(A)(-x)^{n-k} + \dots + i_n(A),$$

whence

$$i_k(A) := \sum_{1 \leq n_1 < \dots < n_k \leq n} \lambda_{n_1}(A)\lambda_{n_2}(A) \cdots \lambda_{n_k}(A) \tag{4}$$

where the eigenvalues $\lambda_i(A)$ are the roots of P_A . For example,

$$i_1(A) = \text{trace}(A) = \lambda_1(A) + \dots + \lambda_n(A) \quad \text{and} \quad i_n(A) = \det(A) = \lambda_1(A) \cdots \lambda_n(A).$$

Another definition of $i_k(A)$ is

$$i_k(A) := \text{sum of the principal minors of order } k \text{ of } A.$$

A principal minor of order k is the determinant of the $k \times k$ matrix extracted from A by considering the rows and columns n_1, \dots, n_k with $1 \leq n_1 < \dots < n_k \leq n$.

Proof of Theorem 3.1 There are several ways to prove (ii): for instance, by purely linear algebra techniques, by induction [1, p. 404], or by quadratic optimization techniques [8, p. 220; 15, p. 409]. The result also follows from the factorization of a matrix into a product of triangular matrices; we sketch this proof (see more in [14]). Let $A \in \mathcal{S}^n(\mathbb{R})$ be such that $\Delta_k(A) = \det(A_k) \neq 0$ for all $k = 1, \dots, n$. Then there exists a (unique) lower triangular matrix with ones on the diagonal, denoted by S , such that

$$A = SDS^T \quad \text{where } D = \text{diag}(\Delta_1(A), \Delta_2(A)/\Delta_1(A), \dots, \Delta_n(A)/\Delta_{n-1}(A)).$$

We call $d_1 := \Delta_1(A)$, $d_k := \Delta_k(A)/\Delta_{k-1}(A)$ the so-called (principal) pivots of A . So we have

$$\begin{aligned} A > 0 &\iff (d_k > 0 \text{ for all } k = 1, \dots, n) \\ &\iff (\Delta_k(A) > 0 \text{ for all } k = 1, \dots, n). \end{aligned}$$

Then it is easy to establish characterization (ii).

Let us prove here characterization (iii) which is less usual. One implication comes easily: if $A > 0$, then all the eigenvalues of A are positive; each $i_k(A)$ is then positive as well (recall (4)). To establish the reverse implication, we assume that $i_k(A) > 0$ for all k , so in particular $\det(A) > 0$. Suppose now that there exists a negative eigenvalue of A , call it $\lambda_{i_0}(A) < 0$. Then observe that

$$0 < P_A(\lambda_{i_0}(A)) = (-\lambda_{i_0}(A))^n + \dots + i_k(A)(-\lambda_{i_0}(A))^{n-k} + \dots + i_n(A)$$

which contradicts the fact that $P_A(\lambda_{i_0}(A)) = 0$. Thus we have $\lambda_i(A) \geq 0$ for all i , and since $0 < \det(A) = \lambda_1(A) \cdots \lambda_n(A)$, we get that $\lambda_i(A) > 0$ for all i , which guarantees that A is positive definite. □

Example 3.1 (Invariants in dimension 2 and 3) Let us illustrate in small dimension the third point of the previous theorem. In the case $n = 2$, we have the simple formulation

$$P_A(x) = x^2 - \text{trace}(A)x + \det A.$$

We note, by the way, that this is a particular case of the development

$$\det(A + B) = \det A + \langle\langle \text{cof } A, B \rangle\rangle + \det B = \det A + \langle\langle A, \text{cof } B \rangle\rangle + \det B,$$

where $\text{cof } M$ stands for the cofactors matrix of the matrix M (in general, this development is obtained by differential calculus with the fact that the determinant is multilinear, see [16]). Let us come back to our situation: the two principal invariants of A are $\text{trace } A$ and $\det A$, so we have

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix} \succ 0 \iff \begin{cases} a > 0, \\ ac - b^2 > 0 \end{cases} \iff \begin{cases} a + c > 0, \\ ac - b^2 > 0. \end{cases}$$

For the case $n = 3$, we have

$$P_A(x) = -x^3 + \text{trace}(A)x^2 - \text{trace}(\text{cof } A)x - \det A$$

which is again a particular case of the nice formula

$$\det(A + B) = \det A + \langle\langle \text{cof } A, B \rangle\rangle + \langle\langle A, \text{cof } B \rangle\rangle + \det B.$$

The three principal invariants are

$$\begin{aligned} i_1(A) &= \text{trace } A = \lambda_1(A) + \lambda_2(A) + \lambda_3(A) \\ i_2(A) &= ((\text{trace } A)^2 - \text{trace}(A^2))/2 \\ &= \lambda_1(A)\lambda_2(A) + \lambda_2(A)\lambda_3(A) + \lambda_1(A)\lambda_3(A) = \text{trace}(\text{cof}(A)) \\ i_3(A) &= ((\text{trace } A)^3 - 3\text{trace}(A)\text{trace}(A^2) + 2\text{trace}(A^3))/6 \\ &= \lambda_1(A)\lambda_2(A)\lambda_3(A) = \det A. \end{aligned}$$

Thus we have

$$\begin{aligned} \begin{bmatrix} a & b & d \\ b & c & e \\ d & e & f \end{bmatrix} \succ 0 &\iff \begin{cases} a > 0, \\ ac - b^2 > 0, \\ \det A > 0 \end{cases} \\ &\iff \begin{cases} a + c + f > 0, \\ (cf - e^2) + (af - d^2) + (ac - b^2) > 0, \\ \det A > 0 \end{cases} \end{aligned}$$

which gives a handy characterization in dimension 3. These principal invariants are widely used for $n = 3$ in mechanics where they often have a physical interpretation (for example, as stresses and strains); see the textbook [17].

3.1.3 Characterization by Polar Cones

Here is now a more original characterization using convex cones in \mathbb{R}^n . We start by recalling some convex analysis that is needed; those notions (polar cones and Moreau

polar decomposition) are interesting on their own, and turn out to be less known than they deserve (for more details, see [13]).

If K is a closed, convex cone in a Euclidean space $(\mathcal{E}, \langle \cdot, \cdot \rangle)$, then the polar cone K° of K is the closed, convex cone of \mathcal{E} made of all the points whose projection onto K is 0; in other terms,

$$K^\circ := \{s \in \mathcal{E} : \langle s, x \rangle \leq 0, \text{ for all } x \in K\}. \tag{5}$$

Examples: for $K = \mathbb{R}_+^n$, we have $K^\circ = \mathbb{R}_-^n$; for a linear subspace $K = V$, we have $K^\circ = V^\perp$. We see that polarity is to (closed) convex cones what orthogonality is to linear subspaces. The fundamental result on this polarity is its reflexivity: doing the polar transformation twice leads back home, i.e. $(K^\circ)^\circ = K$. Another result that we will need later (and that is easy to prove from the definition) is that for an invertible matrix B , we have

$$[B(K)]^\circ = B^{-\top}(K^\circ). \tag{6}$$

The most important result on polar cones is without doubt the following Moreau decomposition. Let x, x_1 and x_2 be three elements of \mathcal{E} ; then the properties (i) and (ii) are equivalent:

- (i) $x = x_1 + x_2$ with $x_1 \in K, x_2 \in K^\circ$ and $\langle x_1, x_2 \rangle = 0$;
- (ii) $x_1 = \text{Proj}_K(x)$ and $x_2 = \text{Proj}_{K^\circ}(x)$.

Here Proj_C stands for the projection onto the closed, convex set C . The decomposition of \mathbb{R}^n into K and K° generalizes the decomposition of \mathbb{R}^n into the direct sum of a linear space L and its orthogonal complement L^\perp . Note that if we know how to project onto K , we get as a bonus the projection onto K° , which could be interesting in practical cases when K° is more complicated than K (or the other way around).

Let us come back to matrices and prove the following original characterization of positive definiteness.

Proposition 3.1 *Let $A \in \mathcal{S}^n(\mathbb{R})$ be invertible and K be a closed, convex cone of \mathbb{R}^n ; then*

$$A \succ 0 \iff \begin{cases} \langle Ax, x \rangle > 0 & \text{for all } x \in K \setminus \{0\} \text{ and} \\ \langle A^{-1}x, x \rangle > 0 & \text{for all } x \in K^\circ \setminus \{0\}. \end{cases}$$

Note the nice symmetry of the result since $(A^{-1})^{-1} = A$ and $(K^\circ)^\circ = K$. This characterization, the proof of which is not straightforward, is due to [18]. We present here a proof, suggested by X. Bonnefond, that uses the Moreau decomposition, as expected.

Proof The fact that the condition is necessary follows directly from the definition of positive definiteness; we focus on sufficiency. Let A be an invertible matrix satisfying the condition, and consider an eigendecomposition $A = UDU^\top$ (where $D = \text{diag}(\lambda_1, \dots, \lambda_n)$ with nonzero λ_i). Observe first that up to a change of cone ($K \leftarrow U^\top K$), we can assume that A is diagonal: the condition can indeed be written,

with the help of (6), as

$$\begin{cases} \langle Dx, x \rangle > 0 & \text{for all } x \in U^\top K \setminus \{0\}, \\ \langle D^{-1}x, x \rangle > 0 & \text{for all } x \in U^\top K^\circ \setminus \{0\} (= (U^\top K)^\circ \setminus \{0\}). \end{cases}$$

So we consider that $A = D$ is diagonal, and we just have to prove that the condition

$$\begin{cases} \sum_{i=1}^n \lambda_i x_i^2 > 0 & \text{for all } x \in K \setminus \{0\}, \\ \sum_{i=1}^n \frac{1}{\lambda_i} x_i^2 > 0 & \text{for all } x \in K^\circ \setminus \{0\} \end{cases}$$

yields that all the λ_i are strictly positive.

For the sake of contradiction, assume that only the k (with $k < n$) first eigenvalues $\lambda_1, \dots, \lambda_k$ are positive, while the $n - k$ last $\lambda_{k+1}, \dots, \lambda_n$ are negative. We do one more step of simplification of the expression by scaling the variables by $S = \text{diag}(\sqrt{|\lambda_1|}, \dots, \sqrt{|\lambda_n|})$: introducing $\tilde{K} := SK$, we observe that

$$\begin{cases} \sum_{i=1}^k y_i^2 > \sum_{i=k+1}^n y_i^2 & \text{for all } y \in \tilde{K} \setminus \{0\}, \\ \sum_{i=1}^k z_i^2 > \sum_{i=k+1}^n z_i^2 & \text{for all } z \in S^{-1}K^\circ \setminus \{0\} (= \tilde{K}^\circ \setminus \{0\}). \end{cases} \tag{7}$$

After multiplying these two inequalities together, the Cauchy–Schwarz inequality yields

$$\forall y \in \tilde{K} \setminus \{0\}, z \in \tilde{K}^\circ \setminus \{0\}, \quad \sum_{i=1}^k y_i^2 \sum_{i=1}^k z_i^2 > \left(\sum_{i=k+1}^n z_i y_i \right)^2. \tag{8}$$

We apply this inequality to a well-chosen couple of vectors y and z to get a contradiction. Consider the basis vector $e = (0, \dots, 0, 1)$ and note that e does not lie in \tilde{K} nor in \tilde{K}° . So its Moreau decomposition gives

$$e = y + z \quad \text{with } y \in \tilde{K} \setminus \{0\}, z \in \tilde{K}^\circ \setminus \{0\} \text{ and } \langle y, z \rangle = 0.$$

Observe that we have $z_i = -y_i$ for all $i = 1, \dots, k$ so that

$$\sum_{i=1}^k z_i^2 = \sum_{i=1}^k y_i^2.$$

Now the orthogonality of y and z gives

$$\sum_{i=k+1}^n z_i y_i = - \sum_{i=1}^k z_i y_i \left(= \sum_{i=1}^k y_i^2 \right).$$

This contradicts (8), so proves that $k = n$, which establishes the characterization. \square

3.2 Characterizations of Positive Semidefiniteness

Most of the characterizations of positive definiteness have their positive semidefiniteness counterparts. In this section, we briefly highlight some similarities and differences between the positive definite and semidefinitenesses.

The property $A \succeq 0$ is equivalent to each of the following statements:

- *There exists a matrix B such that $A = BB^T$.* This can also be read as: the quadratic form $q_A(x) = \|B^T x\|^2$ is zero on $\ker B = \ker A$.
- *The eigenvalues $\lambda_1, \dots, \lambda_n$ of A are non-negative.* The set $\mathcal{S}_+^n(\mathbb{R})$ can thus be seen as the inverse image of \mathbb{R}_+^n

$$\mathcal{S}_+^n(\mathbb{R}) = \lambda^{-1}(\mathbb{R}_+^n)$$

by the eigenvalue function $\lambda: \mathcal{S}^n(\mathbb{R}) \rightarrow \mathbb{R}^n$ assigning to a symmetric A its eigenvalues in a nonincreasing order. The convex set $\mathcal{S}_+^n(\mathbb{R})$ is then a convex *spectral set* in the sense of [12]. Note that, more generally, we can prove that a spectral set $\lambda^{-1}(C)$ is convex if and only if C is convex.

- *For $k = 1, \dots, n$, all the principal minors of order k of A are non-negative.* This means that the determinants of all the submatrices made of k rows and k columns should be non-negative—and not only the n leading principal minors $\Delta_k(A)$ of Theorem 3.1. Having only $\Delta_k(A) \geq 0$ does not guarantee semidefiniteness, as shown by the following counter-example where A is not positive semidefinite (it is actually negative semidefinite)

$$A = \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix} \quad \text{and} \quad \det A_1 = \det A_2 = 0.$$

In general, checking all the principal minors to draw a conclusion about the positive semidefiniteness would mean checking $2^n - 1$ polynomial relations—and not only n as for the positive definiteness, which is a surprising gap! In the next characterization though, only n polynomial relations come into play.

- *The principal invariants $i_1(A), \dots, i_n(A)$ are non-negative.* This is a sort of aggregate form of the previous condition, since for all k , the invariant $i_k(A)$ is the sum of $\binom{n}{k}$ principal minors. Checking the positivity of those n polynomial relations does yield positive semidefiniteness. The proof is conducted in the same way as for (iii) in Theorem 3.1 (see [1, p. 403]). Counting the number of principal minors involved in this characterization by the $i_k(A)$, we naturally retrieve $2^n - 1$ since

$$\binom{n}{1} + \binom{n}{2} + \dots + \binom{n}{n} = 2^n - 1.$$

4 Geometry of the Set of Positive (Semi)Definite Matrices

The set $\mathcal{S}_+^n(\mathbb{R})$ of positive semidefinite matrices is a closed, convex cone of $\mathcal{S}^n(\mathbb{R})$ that enjoys nice, subtle geometrical properties. Some aspects are exposed in this section; we see the cone $\mathcal{S}_+^n(\mathbb{R})$ as a concrete example to illustrate various notions of convex geometry and smooth geometry. We denote the set of positive definite matrices by $\mathcal{S}_{++}^n(\mathbb{R})$.

4.1 Closedness Under Matrix Operations

Closedness Under Addition As a convex cone, $\mathcal{S}_+^n(\mathbb{R})$ is closed under addition and multiplication by $\alpha \geq 0$. We can add moreover the following property (easy to see by definitions)

$$(A \geq 0, B > 0) \implies A + B > 0.$$

The cone $\mathcal{S}_+^n(\mathbb{R})$ is also closed under another addition, the so-called parallel sum. When $A > 0$ and $B > 0$, the parallel addition is defined by

$$A // B := (A^{-1} + B^{-1})^{-1}.$$

Since both the inversion and the sum preserve $\mathcal{S}_{++}^n(\mathbb{R})$, the parallel addition preserves it as well. This operation was introduced [19] by analogy with electrical networks: by the Kirchhoff law, two wires of resistances r_1 and r_2 connected in parallel have a total resistance r such that $1/r = 1/r_1 + 1/r_2$. When $A, B \geq 0$, we could define $A // B$ using pseudo-inversion (see, e.g. [4]). It turns out that this sum has also a natural variational-analysis definition (see, e.g. [20, Problem 25]) as

$$\langle (A // B)x, x \rangle = \inf_{y+z=x} \{ \langle Ay, y \rangle + \langle Bz, z \rangle \}.$$

We recognize this as the *infimal convolution* of the two quadratic forms q_A and q_B ; this operation is basic and important in convex analysis (see [13, Chap. B]).

Closedness Under Multiplication The product of two positive semidefinite matrices is not positive semidefinite, in general: the matrix product destroys symmetry. There is no way to get out of this since we have the following (rather surprising) result:

Any $n \times n$ -matrix A can be described as a product of two symmetric matrices.

This result is presented by [21] and [22] that give credit to G. Frobenius (1910). A quick way to prove it is to use the fact that a matrix A and its transpose are similar: there exists an invertible $S \in \mathcal{S}^n(\mathbb{R})$ (non-unique, though) such that $A^\top = S^{-1}AS$ [23]. As a consequence, we can decompose $A = S_1S_2$ with $S_1 = SA^\top$ (which is symmetric by choice of S) and $S_2 = S^{-1}$ which is symmetric.

Let us come back to (non)closedness under multiplication and give a positive result. Imposing a posteriori symmetry indeed solves the problem: if $A, B \geq 0$ then the product AB is positive semidefinite whenever it is symmetric—that is exactly when A and B commute. We add here two (not well-known) results of the same kind:

- Result of E. Wigner [24]. If $A_1, A_2, A_3 > 0$ and the product $A_1A_2A_3$ is symmetric, then it is positive definite.
- Result of C. Ballantine [25]. Except when $A = -\lambda I_n$ with $\lambda > 0$ and n even, any square matrix with positive determinant can be written as the product of four positive definite matrices. (For the case left aside, five positive definite matrices are needed.)

We also mention that, for another nonstandard matrix product, $\mathcal{S}_+^n(\mathbb{R})$ is closed without any further assumption. If $A = (a_{ij})$ and $B = (b_{ij})$, we define the matrix product $C = A \circ B = (c_{ij})$ by $c_{ij} = a_{ij}b_{ij}$ for all i, j . Result: if A and B are positive definite (resp., semidefinite) then so is the product $A \circ B$ (see [1, 7.5.3]).

4.2 Convex Sets Attached with $\mathcal{S}_+^n(\mathbb{R})$

4.2.1 Interior and Boundary

The (relative) interior of a convex set is always convex; here the interior of $\mathcal{S}_+^n(\mathbb{R})$ turns out to be exactly $\mathcal{S}_{++}^n(\mathbb{R})$. This is another appearance of the analogy with \mathbb{R}_+ , since the interior of $(\mathbb{R}_+)^n$ is $(\mathbb{R}_{++})^n$. The boundary of $\mathcal{S}_+^n(\mathbb{R})$ is exactly made up of the singular matrices $A \geq 0$. There we find the positive semidefinite matrices of rank k , for $1 \leq k \leq n - 1$, forming smooth manifolds fitting together nicely (see Sect. 4.4).

4.2.2 The Facial Structure of $\mathcal{S}_+^n(\mathbb{R})$

A convex subset of $F \subset C$ is a face of C if every segment $[a, b] \subset C$ such that there exists an element of F in $]a, b[$ is entirely contained in F . All the faces of $\mathcal{S}_+^n(\mathbb{R})$ are moreover *exposed* faces: a face F of $\mathcal{S}_+^n(\mathbb{R})$ is the intersection of $\mathcal{S}_+^n(\mathbb{R})$ with a *supporting* hyperplane (that is, a hyperplane H such that $\mathcal{S}_+^n(\mathbb{R})$ is entirely contained in one of the closed half-spaces delimited by H). More precisely, the result is the following (see [26] pointing to earlier references and discussing about the generalization to spectral sets): F is a face of $\mathcal{S}_+^n(\mathbb{R})$ if and only if F is the convex cone spanned by $\{vv^T : v \in V\}$ where V is a subspace of \mathbb{R}^n . The dimension of F is then $d(d + 1)/2$ if d is the dimension of V . For example, with $n = 2$: there is a face of dimension 0 (the apex of the cone $\mathcal{S}_+^n(\mathbb{R})$), faces of dimension 1 (extremal rays of $\mathcal{S}_+^2(\mathbb{R})$ directed by vectors xx^T with nonzero $x \in \mathbb{R}^n$), and a face of dimension 3 (the whole $\mathcal{S}_{++}^2(\mathbb{R})$).

There is another way to generate all the faces of $\mathcal{S}_+^n(\mathbb{R})$, as follows. Let L be a linear subspace of \mathbb{R}^n of dimension m , and set

$$F_L := \{A \in \mathcal{S}_+^n(\mathbb{R}) : L \subset \ker A\};$$

then F_L is a face of $\mathcal{S}_+^n(\mathbb{R})$ of dimension $r(r + 1)/2$ with $r = n - m$. When L ranges all the subspaces of dimension m , F_L visits all the faces of dimension $r(r + 1)/2$.

Let us mention that the knowledge of the faces of $\mathcal{S}_+^n(\mathbb{R})$ is useful for some optimization problems with semidefinite constraints. Indeed, the so-called “facial reduction technique” uses the explicit form of the faces to reformulate degenerate semidefinite optimization problems into problems in smaller dimension; see the initial article [27], and the recent [28] for an application to sensor network problems.

4.2.3 Polar Cone and Projection

In the space $\mathcal{S}^n(\mathbb{R})$ equipped with the inner product $\langle \cdot, \cdot \rangle$, the polar cone (5) of $K = \mathcal{S}_+^n(\mathbb{R})$ is simply $K^\circ = -\mathcal{S}_+^n(\mathbb{R})$. It is also easy to determine the Moreau decomposition of a matrix $A \in \mathcal{S}^n(\mathbb{R})$ onto $\mathcal{S}_+^n(\mathbb{R})$ and its polar. Indeed, consider an

orthogonal eigendecomposition $U^T A U = \text{diag}(\lambda_1, \dots, \lambda_n)$; then the two matrices

$$A_+ := U \text{diag}(\max\{0, \lambda_1\}, \dots, \max\{0, \lambda_n\}) U^T,$$

$$A_- := U \text{diag}(\min\{0, \lambda_1\}, \dots, \min\{0, \lambda_n\}) U^T$$

are such that $A = A_+ + A_-$ and $\langle A_+, A_- \rangle = 0$. Thus A_+ (resp., A_-) is the projection of A onto $\mathcal{S}_+^n(\mathbb{R})$ (resp., onto its polar $-\mathcal{S}_+^n(\mathbb{R})$). Said otherwise, to project A onto $\mathcal{S}_+^n(\mathbb{R})$, we just have to compute an eigendecomposition of A and cut off negative eigenvalues by 0. It is remarkable that we have an explicit expression of the projection onto $\mathcal{S}_+^n(\mathbb{R})$, and that this projection is easy to compute (essentially by computing an eigendecomposition). More sophisticated projections onto subsets of $\mathcal{S}_+^n(\mathbb{R})$ are also computable using standard tools of numerical optimization [29]; those projections have many applications in statistics and finance (see [30] for an early reference of the projection onto $\mathcal{S}_+^n(\mathbb{R})$; see also [29, Sect. 5]).

Finally, note that we can interpret the projection of A onto $\mathcal{S}_+^n(\mathbb{R}) = \lambda^{-1}(\mathbb{R}_+^n)$ with respect to the projection of its eigenvalues onto \mathbb{R}_+^n , namely

$$\text{Proj}_{\mathcal{S}_+^n(\mathbb{R})}(A) = U \text{diag}(\text{Proj}_{\mathbb{R}_+^n}(\lambda_1, \dots, \lambda_n)) U^T.$$

This result has a nice generalization: we can compute a projection onto a spectral set $\lambda^{-1}(C)$ as soon as we know how to project onto the underlying C (see [31]).

4.2.4 Tangent and Normal Cones

Normal cones in convex geometry play the role of normal spaces in smooth geometry: they give “orthogonal” directions to a set at a point of the set. Their “duals”, the tangent cones, then give a simple conic approximation of a convex set around a point.

The normal cone to $\mathcal{S}_+^n(\mathbb{R})$ at $X \in \mathcal{S}_+^n(\mathbb{R})$ can be defined as the set of directions $S \in \mathcal{S}^n(\mathbb{R})$ such that the projection of $X + S$ onto $\mathcal{S}_+^n(\mathbb{R})$ is X itself. In mathematical terms, this means

$$N_A := \{S \in \mathcal{S}^n(\mathbb{R}) : \langle S, B - A \rangle \leq 0 \text{ for all } B \in \mathcal{S}_+^n(\mathbb{R})\}.$$

The tangent cone is then defined as the polar of N_A and admits the characterization:

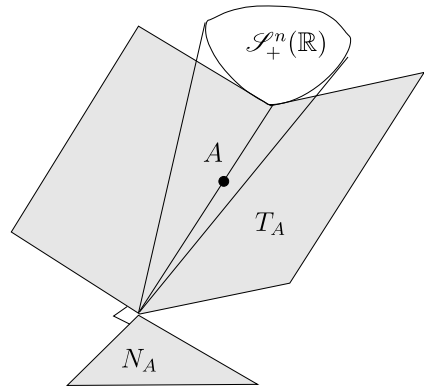
$$T_A := N_A^\circ = \overline{\mathbb{R}_+(\mathcal{S}_+^n(\mathbb{R}) - A)}.$$

To fix ideas, Fig. 2 gives an illustrative representation of the cone with its tangent and normal space at a point (beware of the fact that this is not a real representation, since for $n = 2$ the boundary of the cone is smooth).

For the special case $A = 0$, we simply have $N_0 = \mathcal{S}_+^n(\mathbb{R})^\circ = -\mathcal{S}_+^n(\mathbb{R})$ and $T_0 = \mathcal{S}_+^n(\mathbb{R})$. For the general case $A \in \mathcal{S}_+^n(\mathbb{R})$, we have

$$\begin{aligned} T_A &= \{M \in \mathcal{S}^n(\mathbb{R}) : \langle M u, u \rangle \geq 0 \text{ for all } u \in \ker A\}, \\ N_A &= \{M \leq 0 : \langle M, A \rangle = 0\} \\ &= \{M \leq 0 : M A = 0\} = \{M \leq 0 : A M = 0\} \\ &= \{M \leq 0 : \text{Im } M \subset \ker A\} = \{M \leq 0 : \text{Im } A \subset \ker M\}. \end{aligned} \tag{9}$$

Fig. 2 (Nonrealistic) illustration of the normal and tangent cones to $\mathcal{S}_+^n(\mathbb{R})$ at A



The first expression of N_A comes from the general result for closed, convex cones (see [13, A.5.2.6]). The second one follows from the (quite surprising) property:

$$\text{for } A, B \in \mathcal{S}_+^n(\mathbb{R}), \text{ we have: } \langle\langle A, B \rangle\rangle = 0 \iff AB = 0,$$

which has a one-line proof using a classical trick, as follows

$$\langle\langle A, B \rangle\rangle = \text{trace}(A^{1/2}A^{1/2}B^{1/2}B^{1/2}) = \text{trace}(B^{1/2}A^{1/2}A^{1/2}B^{1/2}) = \|A^{1/2}B^{1/2}\|^2.$$

The third expression comes easily from the second. We can also describe N_A one step further from the last expression. Take $A \in \mathcal{S}_+^n(\mathbb{R})$ of rank $r (< n)$. The dimension of ker A is $n - r$, let $\{u_{r+1}, \dots, u_n\}$ be a basis of it. Then

$$M \in N_A \iff M = \sum_{i=r+1}^n \alpha_i u_i u_i^\top \text{ with } \alpha_i \leq 0 \text{ (} i = r + 1, \dots, n\text{)}.$$

4.2.5 Spectahedron, the Spectral Simplex

A special subset of $\mathcal{S}_+^n(\mathbb{R})$ plays an important role in the variational analysis of the eigenvalues (see [32, 33]) and in some applications of semidefinite programming (as, for example, for sparse principal component analysis [34]). The so-called *spectahedron* is defined by

$$\Omega_1 := \{A \geq 0 : \text{trace}(A) = 1\}. \tag{10}$$

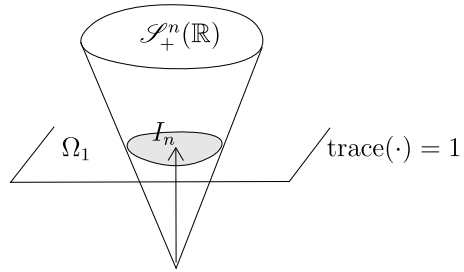
This set is convex and compact (see an illustration on Fig. 3). Its extreme points are exactly the matrices xx^\top with unit-vectors $x \in \mathbb{R}^n$. The spectahedron equivalently defined through the eigenvalues: $\Omega_1 = \lambda^{-1}(\Pi_1)$ is the spectral set of $\Pi_1 := \{(\lambda_1, \dots, \lambda_n) : \lambda_i \geq 0, \sum_{i=1}^n \lambda_i = 1\}$ the unit-simplex of \mathbb{R}^n .

This set and its properties generalize as follows. For an integer $m \in \{1, \dots, n\}$, set

$$\Omega_m := \{A \geq 0 : \text{trace}(A) = m \text{ and } \lambda_{\max}(A) \leq 1\}.$$

For example, Ω_n is obviously reduced to a single element, the identity matrix I_n . Note that we did not add $\lambda_{\max}(A) \leq 1$ in the definition of Ω_1 , since it was automatically satisfied. The general result is the following: the set Ω_m is convex and compact, and

Fig. 3 (Nonrealistic) representation of the spectrahedron



it is the convex hull of the matrices XX^T , where X is an $n \times m$ -matrix such that $X^T X = I_m$. In other words, the convex hull of the orthogonal projection matrices of rank m is exactly the set of symmetric matrices whose eigenvalues are between 0 and 1, and whose trace is m . The proof of this result starts with showing that Ω_m is a spectral $\Omega_m = \lambda^{-1}(\Pi_m)$ with the compact convex polyhedron $\Pi_m := \{(\lambda_1, \dots, \lambda_n) : 0 \leq \lambda_i \leq 1, \sum_{i=1}^n \lambda_i = m\}$. Then we can determine the extremal points of Π_m : the point $(\bar{\lambda}_1, \dots, \bar{\lambda}_n)$ is extremal in Π_m if and only if all the $\bar{\lambda}_i$ are zeros, expected m of them that are ones.

4.3 Representation as Inequality Constrained Set: Nonsmooth Viewpoint

A common, pleasant situation in optimization is when a constraint set C is represented with the help of inequalities

$$g_1(x) \leq 0, \quad \dots, \quad g_p(x) \leq 0,$$

and the interior of C with the help of strict inequalities with the same functions

$$g_1(x) < 0, \quad \dots, \quad g_p(x) < 0.$$

Does there exist such a representation for $C = \mathcal{S}_+^n(\mathbb{R})$ and its interior $\mathcal{S}_{++}^n(\mathbb{R})$? The answer is yes. We highlight two types of representing functions: a nonsmooth one in this section, and polynomial ones in the next section.

Consider the function $g : \mathcal{S}^n(\mathbb{R}) \rightarrow \mathbb{R}$ defined by

$$g(A) := \lambda_{\max}(-A) = -\lambda_{\min}(A). \tag{11}$$

This function is convex and positively homogeneous, as a maximum of linear functions. Indeed, we have the well-known Rayleigh quotient

$$\lambda_{\max}(A) = \max_{\|x\|=1} \langle Ax, x \rangle = \max_{\|x\|=1} \langle\langle A, xx^T \rangle\rangle (= \max_{X \in \Omega_1} \langle\langle A, X \rangle\rangle),$$

where Ω_1 is defined by (10). More precisely, the last expression says that λ_{\max} is the support function of Ω_1 (the notion of a support function is fundamental in convex analysis; see [13, Chap. C]). Let us come back to the representation of $\mathcal{S}_+^n(\mathbb{R})$ as an inequality constrained set: we have a first adequate representation as

$$\begin{aligned} \mathcal{S}_+^n(\mathbb{R}) &= \{A \in \mathcal{S}^n(\mathbb{R}) : g(A) \leq 0\}, \\ \mathcal{S}_{++}^n(\mathbb{R}) &= \{A \in \mathcal{S}^n(\mathbb{R}) : g(A) < 0\}. \end{aligned}$$

We can thus represent $\mathcal{S}_+^n(\mathbb{R})$ as a constrained set with a single inequality, but we have to use the nonsmooth function λ_{\max} . On the other hand, in Sect. 4.4, we will represent it with several smooth functions.

We mention briefly that smooth approximations of the nonsmooth representing function λ_{\max} give smooth approximations of $\mathcal{S}_+^n(\mathbb{R})$. We start from $\mathcal{S}_+^n(\mathbb{R}) = \{A \in \mathcal{S}^n(\mathbb{R}) : \lambda_{\max}(-A) \leq 0\}$ and, following [35], we define for all $\mu > 0$

$$X \in \mathcal{S}^n(\mathbb{R}) \mapsto f_\mu(X) := \mu \log \left(\sum_{i=1}^n \exp(\lambda_i(X)/\mu) \right)$$

where the $\lambda_i(X)$ are the eigenvalues of X . This expression of f_μ can be transformed into

$$f_\mu(X) = \lambda_{\max}(X) + \mu \log(\text{trace}(E(X))) \quad \text{with } E(X) = \exp\left(\frac{X - \lambda_{\max}(X)I_n}{\mu}\right).$$

In [35], it is shown that f_μ is globally Lipschitz, of class C^1 with gradient

$$\nabla f_\mu(X) = \frac{E(X)}{\text{trace}(E(X))},$$

and that we have the uniform approximation

$$\lambda_{\max}(X) \leq f_\mu(X) \leq \lambda_{\max}(X) + \mu \log n.$$

Therefore, $\mathcal{S}_+^n(\mathbb{R})$ could be uniformly approximated by $\{A \in \mathcal{S}^n(\mathbb{R}) : f_\mu(-A) \leq 0\}$ which is an inner $(\mu \log n)$ -approximation.

4.4 Representation as Inequality Constrained Sets: Semialgebraic Viewpoint

The adequate representation of the previous section used a nonsmooth function. The algebraic nature of $\mathcal{S}_+^n(\mathbb{R})$ gives a representation with polynomials, and it is tempting to consider the positively homogenous polynomial functions $\Delta_k(A)$. However, we cannot use them to have the desired decomposition since (recall Sect. 3.2)

$$\begin{aligned} \mathcal{S}_+^n(\mathbb{R}) &\subsetneq \{A \in \mathcal{S}^n(\mathbb{R}) : \Delta_k(A) \geq 0 \text{ for all } k = 1, \dots, n\}, \\ \mathcal{S}_{++}^n(\mathbb{R}) &= \{A \in \mathcal{S}^n(\mathbb{R}) : \Delta_k(A) > 0 \text{ for all } k = 1, \dots, n\}. \end{aligned}$$

We still have an adequate polynomial representation: setting $g_k = -i_k$, we have indeed

$$\begin{aligned} \mathcal{S}_+^n(\mathbb{R}) &= \{A \in \mathcal{S}^n(\mathbb{R}) : g_k(A) \leq 0 \text{ for all } k = 1, \dots, n\}, \\ \mathcal{S}_{++}^n(\mathbb{R}) &= \{A \in \mathcal{S}^n(\mathbb{R}) : g_k(A) < 0 \text{ for all } k = 1, \dots, n\}. \end{aligned} \tag{12}$$

The constraint functions g_k are polynomial, positively homogeneous of degree k ; obviously they are of class C^∞ , whereas g in (11) is nonsmooth. Note that $\mathcal{S}_+^n(\mathbb{R})$ represented by (12) is clearly a cone, but its convexity is less clear (and comes from subtle reasons: hyperbolic polynomials and the convexity theorem of Gårding; see [12]).

Also note that in the representation (12) we have the a priori knowledge of the *active constraints* at a matrix A (those such that $g_k(A) = 0$). Usually, we only know this a posteriori; here we directly know from the rank r of A that:

$$g_k(A) < 0 \text{ for } k = 1, \dots, r \quad \text{and} \quad g_k(A) = 0 \text{ for } k = r + 1, \dots, n.$$

We now briefly discuss the fact that $\mathcal{S}_+^n(\mathbb{R})$, as it appears in (12), is a semialgebraic set. A so-called *semialgebraic* set is a set defined by unions and intersections of a finite number of polynomial inequalities. It turns out that these sets enjoy very nice closedness properties (almost any “finite” operation preserves semialgebraicity; see [36]). These properties are useful tools for the analysis of structured nonsmooth optimization problems (see, e.g. [37, 38]). One of the main properties of semialgebraic sets is that they can be decomposed into a union of connected smooth manifolds (so-called *strata*) that fit together nicely. Denoting by \mathcal{R}_r the smooth submanifold made up from the positive semidefinite matrices of fixed rank r (for $r = 0, \dots, n$), we observe that the semialgebraic cone $\mathcal{S}_+^n(\mathbb{R})$ is the union of the $n + 1$ manifolds \mathcal{R}_r , which are themselves the union of their connected components, the strata of $\mathcal{S}_+^n(\mathbb{R})$. The two extreme cases are: $r = 0$ the apex of the cone and $r = n$ the interior of the cone. The decomposition as union of manifolds is explicit in the representation of $\mathcal{S}_+^2(\mathbb{R})$ in \mathbb{R}^3 : the point ($r = 0$), the boundary deprived of the origin ($r = 1$) and the interior of the cone ($r = 2$).

The tangent space to \mathcal{R}_r at A (denoted by \mathcal{T}_A) is connected to the tangent cone to $\mathcal{S}_+^n(\mathbb{R})$ at A . In fact, \mathcal{T}_A is the largest subspace included in T_A (recall (9)), namely

$$\mathcal{T}_A = T_A \cap -T_A = \{M \in \mathcal{S}^n(\mathbb{R}) : Mu = 0 \text{ for all } u \in \ker A\}.$$

4.5 Natural Riemannian Metric on $\mathcal{S}_{++}^n(\mathbb{R})$

In practice, the computations dealing with positive definite matrices are numerous and involve various operations: averaging, approximation, filtering, estimation, etc. It has been noticed that the Euclidean geometry is not always best suited for some of these operations (as, for example, in image processing; see, e.g. [39] and references therein). This section presents another geometry in $\mathcal{S}_{++}^n(\mathbb{R})$ having different operations (for example, for averaging). Next section connects this geometry to optimization. We refer to [4, Chap. 6] for more details and for the proofs of the results. Note that similar developments hold for \mathcal{R}_r , the fixed rank positive definite matrix manifold (see [40]), but they require more involved techniques. Though we use the language of Riemannian geometry, we stay here at a very basic level.

The cone $\mathcal{S}_{++}^n(\mathbb{R})$ is open, and has two natural geometries: the Euclidean geometry associated with the inner product $\langle\langle \cdot, \cdot \rangle\rangle$, where the distance between A and B is

$$d_2(A, B) := \|A - B\|_2 = \sqrt{\langle\langle A - B, A - B \rangle\rangle};$$

and a Riemannian metric defined from $ds = \|A^{-1/2}(dA)A^{-1/2}\|_2$ (the “infinitesimal length” at A) by the following standard way.

Let $A, B \in \mathcal{S}_{++}^n(\mathbb{R})$ and consider a piecewise C^1 path $\gamma : [a, b] \rightarrow \mathcal{S}_{++}^n(\mathbb{R})$ from $A = \gamma(a)$ to $B = \gamma(b)$; then the length of γ is

$$L(\gamma) := \int_a^b \|\gamma^{-1/2}(t)\gamma'(t)\gamma^{-1/2}(t)\|_2 dt,$$

and the Riemannian distance is

$$d_R(A, B) := \inf\{L(\gamma) : \gamma \text{ paths from } A \text{ to } B\}. \tag{13}$$

It turns out that we have nice explicit expressions of d_R and associated “geodesics” (locally, the shortest paths between two points). The unique geodesic that connects A and B is the following path

$$[0, 1] \mapsto \bar{\gamma}(t) := A^{1/2}(A^{-1/2}BA^{-1/2})^t A^{1/2}$$

that reaches the minimum in (13), so that

$$\begin{aligned} d_R(A, B)^2 &= \|\log(A^{-1/2}BA^{-1/2})\|^2 \\ &= \sum_{i=1}^n (\log \lambda_i(A^{-1/2}BA^{-1/2}))^2 = \sum_{i=1}^n (\log \lambda_i(A^{-1}B))^2. \end{aligned}$$

The last inequality comes from the fact that $A^{-1}B$ and $A^{1/2}BA^{1/2}$ are similar matrices, so they have the same eigenvalues. We can verify from this expression that we have indeed the desirable property $d_R(A, B) = d_R(B, A)$ (since $(A^{-1}B)^{-1} = B^{-1}A$, the squared log of eigenvalues are the same). We give a flavour on why this distance has a better behaviour for some applications.

Distance of Inverses We get easily that the inversion does not change the distance: for any A and B in $\mathcal{S}_{++}^n(\mathbb{R})$, we have

$$d_R(A, B) = d_R(A^{-1}, B^{-1}). \tag{14}$$

This property does not hold for the distance d_2 . Here it simply follows from the fact that $(A^{-1}B)^T = BA^{-1}$ have the same eigenvalues.

Geometric Average (and a Little More) Let A_1, \dots, A_m be m matrices of $\mathcal{S}_{++}^n(\mathbb{R})$. Writing the optimality conditions, we can prove that there exists a unique matrix $M_2(A_1, \dots, A_m)$ in $\mathcal{S}_{++}^n(\mathbb{R})$ that minimizes $X \mapsto \sum_{i=1}^m (\|X - A_i\|_2)^2$, and we get explicitly that M_2 is the usual (arithmetic) average

$$M_2(A_1, \dots, A_m) := \frac{A_1 + \dots + A_m}{m}.$$

For the Riemannian geometry, we have similarly that there exists a unique matrix X in $\mathcal{S}_{++}^n(\mathbb{R})$ that minimizes

$$X \mapsto \sum_{i=1}^m d_R(X, A_i)^2, \tag{15}$$

and this matrix, denoted by $M_R(A_1, \dots, A_m)$, is called the Riemannian average. An interesting property (not shared by M_2) comes from (14): we have

$$M_R(A_1^{-1}, \dots, A_m^{-1}) \text{ is the inverse of } M_R(A_1, \dots, A_m). \tag{16}$$

Finally, the optimality condition of (15) yields that $\sum_{i=1}^m \log(A_i^{-1}X) = 0$ characterizes the Riemannian average.

For $m = 2$, we see that $M_R(A_1, A_2) = A_1(A_1^{-1}A_2)^{1/2} = A_2(A_2^{-1}A_1)^{1/2}$ is the solution of the equation (see [4, Chap. 4] for the details). If furthermore A_1 and A_2 commute, we have

$$M_R(A_1, A_2) = A_1^{1/2}A_2^{1/2}.$$

This last property explains why the Riemannian average is sometimes called the geometric average.

For a general m , computing $M_R(A_1, \dots, A_m)$ might be less direct. As far as the symmetry property (16) is concerned, one could consider instead another matrix average, the so-called resolvent average $M_{\text{res}}(A_1, \dots, A_m)$ of [41]. It can be defined as the minimum of a function looking like (15) (with a Bregman distance replacing d_R ; see [41, Proposition 2.8]), but it also has a simple, easy-to-compute expression that can be written

$$(M_{\text{res}}(A_1, \dots, A_m) + I_n)^{-1} = ((A_1 + I_n)^{-1} + \dots + (A_m + I_n)^{-1})/n.$$

The above equality interprets that the resolvent of the average is the arithmetic average of the resolvents of the A_i 's, which give the name ‘‘resolvent average’’. The fact that the resolvent average satisfies (16) comes from techniques of variational analysis (see [41, Theorem 4.8]).

4.6 Barrier Function of $\mathcal{S}_{++}^n(\mathbb{R})$

As a function of the real variable $x > 0$, the logarithm $x \mapsto \log x$ has a matrix relative which turns out to have a central role in optimization. The celebrated (negative) log-function for matrices is

$$X > 0 \mapsto F(X) := -\log(\det(X)) = \log(\det X^{-1}). \tag{17}$$

Differential calculus for F in $\mathcal{S}_{++}^n(\mathbb{R})$ gives the ‘‘same’’ results as for the \log in \mathbb{R}_+ . The composite function F is of class C^∞ on $\mathcal{S}_{++}^n(\mathbb{R})$; its derivative $DF(X): H \in \mathcal{S}^n(\mathbb{R}) \mapsto DF(X)[H]$ is such that

$$DF(X)[H] = -\text{trace}(X^{-1}H) = \langle -X^{-1}, H \rangle = -\text{trace}(X^{-1/2}HX^{-1/2}),$$

and gives the gradient $\nabla F(X) = -X^{-1}$. We have furthermore

$$\begin{aligned} D^2F(X)[H, H] &= \text{trace}((X^{-1/2}HX^{-1/2})^2) = \langle X^{-1}HX^{-1}, H \rangle, \\ D^3F(X)[H, H, H] &= -2\text{trace}((X^{-1/2}HX^{-1/2})^3). \end{aligned}$$

We can also prove that F is strictly convex on $\mathcal{S}_{++}^n(\mathbb{R})$. This fundamental property is the topic of several exercises in [42].

As mentioned above, this function has a very special role in optimization, more precisely in semidefinite programming [6]. The function F is indeed a *self-concordant barrier-function* for $\mathcal{S}_{++}^n(\mathbb{R})$. The role of *barrier* is intuitive: $F(X) \rightarrow +\infty$ when $X > 0$ approaches the boundary of $\mathcal{S}_{++}^n(\mathbb{R})$, which consists of the singular matrices $A \succeq 0$. Self-concordance is a technical property, namely

$$|D^3 F(X)[H, H, H]| \leq \alpha (D^2 F(X)[H, H])^{3/2}$$

for a constant α (here $\alpha = 2$ is valid). In the 1990s, Y. Nesterov and A. Nemirovski developed a theory [43] that has had tremendous consequences in convex optimization. They showed that self-concordant barrier-functions allow us to design algorithms (called interior-point methods) for linear optimization problems with conic constraints, and that moreover the complexity of those algorithms is fully understood. In particular, the interior-point algorithms for semidefinite programming made a little revolution during the 1990s: they provided numerical methods to solve a wide range of engineering problems, for example, in control [44], or in combinatorial optimization [45].

The log-function F in (17) is at the heart of interior point methods for semidefinite programming; it also has a remarkable connection with the Riemannian metric of Sect. 4.5. Observe indeed that the norm in $\mathcal{S}_+^n(\mathbb{R})$ associated to $D^2 F(X)$ corresponds nicely with the infinitesimal norm defined by the Riemannian structure:

$$\|H\|_{D^2 F(X)} = (D^2 F(X)[H, H])^{1/2} = \|X^{-1/2} H X^{-1/2}\|.$$

It follows that the Riemannian distance naturally associated with the barrier is exactly the natural Riemannian distance (13). This geometrical interpretation of the barrier function then shows that interior-point methods have an intrinsic appeal, and may explain their strong complexity results (see more in [46]).

4.7 Unit Partition in $\mathcal{S}_{++}^n(\mathbb{R})$

In this last section, we discuss the so-called unit partition problem in $\mathcal{S}_{++}^n(\mathbb{R})$. Motivated by a problem originated in economics, I. Ekeland posed it as a challenge in a conference in 1997 [47]. Given $k + 1$ nonzero vectors x^1, \dots, x^k, y in \mathbb{R}^n , when is it possible to find positive definite matrices M_1, \dots, M_k such that

$$M_i y = x^i \quad \text{for all } i = 1, \dots, k \tag{18}$$

and

$$\sum_{i=1}^k M_i = I_n? \tag{19}$$

The first equation is of the quasi-Newton type [48]; the second equation gives the name of the problem.

A condition that guarantees the existence of such matrices should obviously depend on the vectors x^1, \dots, x^k, y . It is easy to get from (18), (19) that a necessary condition is

$$\langle x^i, y \rangle > 0 \quad \text{for all } i = 1, \dots, k \quad \text{and} \quad \sum_{i=1}^k x^i = y. \tag{20}$$

Unfortunately, this condition is not sufficient, as shown by the following counterexample. Consider in \mathbb{R}^2 the three vectors

$$x^1 = \begin{bmatrix} 1/2 \\ 1 \end{bmatrix}, \quad x^2 = \begin{bmatrix} 1/2 \\ -1 \end{bmatrix} \quad \text{and} \quad y = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Then observe that for both $i = 1, 2$, the property $M_i x^i = y$ with $M_i > 0$ yields

$$M_1 = \begin{bmatrix} 1/2 & 1 \\ 1 & \alpha \end{bmatrix} \quad \text{and} \quad M_2 = \begin{bmatrix} 1/2 & -1 \\ -1 & \beta \end{bmatrix},$$

with $\alpha, \beta > 2$. Thus it is impossible to have $M_1 + M_2 = I_2$. Note also that condition (20) is even not sufficient to get a weaker version of the partition (19) with $M_i \geq 0$.

We give here a constructive way to get the unit partition of $\mathcal{S}_+^n(\mathbb{R})$ (and later a variant of the result). The result is due to A. Inchakov, as noted by [49].

Theorem 4.1 (Condition for unit partition of $\mathcal{S}_{++}^n(\mathbb{R})$) *Consider $k + 1$ nonzero vectors x^1, \dots, x^k, y in \mathbb{R}^n , satisfying (20). Then a necessary and sufficient condition for the existence of $M_i > 0$ satisfying (18) and (19) is that*

$$A_0 := I_n - \sum_{i=1}^k \frac{x_i x_i^\top}{\langle x^i, y \rangle} \quad \text{be positive definite on } y^\perp. \tag{21}$$

Proof Let us prove first that the condition (21) is necessary. The Cauchy–Schwarz inequality gives for all i and for all $z \in \mathbb{R}^n$

$$\langle M_i y, z \rangle^2 \leq \langle M_i y, y \rangle \langle M_i z, z \rangle \quad \text{with equality if and only if } z \text{ and } y \text{ are collinear,}$$

which yields

$$\langle x^i, z \rangle^2 \leq \langle x^i, y \rangle \langle M_i z, z \rangle \quad \text{with equality if and only if } z \text{ and } y \text{ collinear.}$$

As a consequence, we have

$$\left\{ \begin{array}{l} \sum_{i=1}^k \left(\frac{\langle x^i, z \rangle^2}{\langle x^i, y \rangle} - \langle M_i z, z \rangle \right) = \sum_{i=1}^k \frac{\langle x^i, z \rangle^2}{\langle x^i, y \rangle} - \|z\|^2 \leq 0 \\ \text{with equality if and only if } z \text{ and } y \text{ are collinear,} \end{array} \right.$$

which means

$$q_{A_0}(z) \geq 0 \quad \text{with equality if and only if } z \text{ and } y \text{ are collinear.}$$

Thus A_0 is positive semidefinite and of kernel reduced to $\mathbb{R}y$, and we get (21).

To prove the sufficiency of condition (21), we propose the matrices

$$M_i := \frac{A_0}{k} + \frac{x^i x^{i\top}}{\langle x^i, y \rangle}.$$

By construction, we have $\sum_{i=1}^k M_i = I_n$; and moreover, since $A_0 y = 0$, we also have $M_i y = x_i$ for all $i = 1, \dots, k$. It remains to prove the positive definiteness of the M_i 's.

It is clear that $M_i \geq 0$, since it is a sum of two positive semidefinite matrices. Let us take $z \in \mathbb{R}^n$ such that $\langle M_i z, z \rangle = 0$, and let us prove that $z = 0$. We have

$$\frac{\langle A_0 z, z \rangle}{k} + \frac{\langle x^i, z \rangle^2}{\langle x^i, y \rangle} = 0,$$

which yields

$$\begin{cases} \langle A_0 z, z \rangle = 0 & \text{(hence } z \text{ and } y \text{ are collinear by (21))} \\ \langle x^i, z \rangle = 0. \end{cases}$$

The conclusion follows easily: there exists $\alpha \in \mathbb{R}$ such that $z = \alpha y$, and the condition $\langle x^i, y \rangle > 0$ implies $\alpha = 0$ and then $z = 0$, so M_i is positive definite. □

It is interesting to note that (21) is of the type

(E₀) $\langle Ax, x \rangle > 0$ for all $x \neq 0$ orthogonal to y , for $A \in \mathcal{S}^n(\mathbb{R})$ and $y \neq 0$ in \mathbb{R}^n , which is a frequently encountered property in matrix analysis and optimization.

Here are four formulations equivalent to (E₀) (see, e.g. [50]).

(E₁) Finsler–Debreu condition: there exists $\mu \geq 0$ such that $A + \mu y y^\top > 0$.

(E₂) Condition on the augmented matrix: the matrix $\bar{A} \in \mathcal{S}^{n+1}(\mathbb{R})$ defined by blocks as

$$\bar{A} = \begin{bmatrix} A & y \\ y^\top & 0 \end{bmatrix}$$

has exactly n positive eigenvalues.

(E₃) Condition on determinants (that the economists are keen of). For $I \subset \{1, \dots, n\}$, we denote by A_I the matrix extracted from A by taking only the columns and rows with indices in I . Similarly y_I is obtained from y by taking the entries y_i for $i \in I$. We set

$$\bar{A}_I = \begin{bmatrix} A_I & y_I \\ y_I^\top & 0 \end{bmatrix}$$

and the condition is $\det A_I < 0$ for all $I = \{1, 2\}, \{1, 2, 3\}, \dots, \{1, \dots, n\}$.

(E₄) Condition on the inverse of the augmented matrix. The matrix \bar{A} is invertible and the $n \times n$ -matrix extracted from \bar{A}^{-1} by taking the first n columns and rows is positive semidefinite.

We finish with a variant of Theorem 4.1 showing that starting from a weaker assumption, we get a weaker result.

Theorem 4.2 (Condition for unit partition of $\mathcal{S}_+^n(\mathbb{R})$) *Let x^1, \dots, x^k and $y \neq 0$ be $k + 1$ vectors of \mathbb{R}^n , satisfying*

$$\begin{cases} (\langle x^i, y \rangle > 0 \text{ or } x^i = 0) & \text{for all } i = 1, \dots, k, \\ \sum_{i=1}^k x^i = y. \end{cases} \tag{22}$$

Then a necessary and sufficient condition for the existence of $M_i \geq 0$ satisfying (18) and (19) is that

$$A_0 := I_n - \sum_{\{i: x^i \neq 0\}} \frac{x_i x_i^\top}{\langle x^i, y \rangle} \text{ is positive semidefinite.} \tag{23}$$

Proof Note that under the assumption of the theorem, we have $\langle x^i, y \rangle \geq 0$ for all $i = 1, \dots, k$ and $x^i \neq 0$ because $\sum_{i=1}^k x^i = y \neq 0$. The proof is similar to (and easier than) the previous one. Let us start with the necessity of the condition (23). We have $A_0 y = 0$ since $\sum_{i=1}^n x^i = \sum_{\{i: x^i \neq 0\}} x^i = y$. Using the Cauchy–Schwarz inequality, we write

$$\langle x^i, z \rangle^2 \leq \langle x^i, y \rangle \langle M_i z, z \rangle,$$

and then

$$\sum_{\{i: x^i \neq 0\}} \frac{\langle x^i, z \rangle^2}{\langle x^i, y \rangle} - \|z\|^2 \leq 0.$$

Thus we get $A_0 \geq 0$. As for sufficiency, we propose the semidefinite matrices

$$M_i := \begin{cases} \frac{A_0}{k} & \text{if } x^i = 0, \\ \frac{A_0}{k} + \frac{x^i x_i^\top}{\langle x^i, y \rangle} & \text{if } \langle x^i, y \rangle > 0. \end{cases}$$

We check that in both cases $M_i y = x^i$, and obviously we also have $\sum_{i=1}^k M_i = I_n$. \square

5 Concluding Remarks

This article highlights nice variational properties of (the set of the) positive semidefinite matrices, and gives pointers to some of their applications. Even if the notion of positive semidefiniteness is basic and taught everywhere, there are still many open questions related to it, and more generally to the interplay between matrix analysis and optimization; see the commented problems in [9] and [51]. The “variational” viewpoint adopted here also opens the way to another notion of positivity, the so-called “copositivity”, which also has rich and useful properties; for example, see the recent review [52].

Acknowledgements The authors thank J.-B. Lasserre for having made us aware of the characterization of the positive semidefiniteness by non-negativity of the principal invariants (see Sect. 3.2).

References

1. Horn, R.A., Johnson, C.R.: *Matrix Analysis*. Cambridge University Press, Cambridge (1989). (New edition, 1999)
2. Zhang, F.: *Matrix Theory*. Springer, Berlin (1999)
3. Bernstein, D.: *Matrix Mathematics*. Princeton University Press, Princeton (2005)
4. Bhatia, R.: *Positive Definite Matrices*. Series in Applied Mathematics. Princeton University Press, Princeton (2007)
5. Strang, G.: *Introduction to Linear Algebra*. Wellesley-Cambridge Press, Wellesley (2009). (4th international edition)
6. Saigal, R., Vandenberghe, L., Wolkowicz, H.: *Handbook of Semidefinite Programming*. Kluwer Academic, Norwell (2000)
7. Anjos, M., Lasserre, J.B.: *Handbook of Semidefinite, Conic and Polynomial Optimization*. Springer, Berlin (2012)
8. Alexeev, V., Galeev, E., Tikhomirov, V.: *Recueil de problèmes d'optimisation*. Mir, Moscow (1987)
9. Hiriart-Urruty, J.-B.: Potpourri of conjectures and open questions in nonlinear analysis and optimization. *SIAM Rev.* **49**(2), 255–273 (2007)
10. Poljak, S., Rendl, F., Wolkowicz, H.: A recipe for semidefinite relaxation for (0,1)-quadratic programming. *J. Glob. Optim.* **7**, 51–73 (1995)
11. Borwein, J.M., Lewis, A.S.: *Convex Analysis and Nonlinear Optimization*. Springer, New York (2000)
12. Lewis, A.S.: Nonsmooth analysis of eigenvalues. *Math. Program.* **84**(1), 1–24 (1999)
13. Hiriart-Urruty, J.-B., Lemaréchal, C.: *Fundamentals of Convex Analysis*. Springer, Heidelberg (2001)
14. Hiriart-Urruty, J.-B., Plusquellec, Y.: *Exercices d'algèbre linéaire et bilinéaire*. Cepadues, Toulouse (1988)
15. Brinkhuis, J., Tikhomirov, V.: *Optimization: Insights and Applications*. Series in Applied Mathematics. Princeton University Press, Princeton (2005)
16. Azé, D., Constans, G., Hiriart-Urruty, J.-B.: *Calcul différentiel et équations différentielles. Exercices corrigés*. EDP Sciences, Les Ulis Cedex A (2010)
17. Truesdell, C.: *A First Course in Rational Continuum Mechanics: General Concepts*. Academic Press, San Diego (1991) (2nd revised edition)
18. Han, S.-P., Mangasarian, O.: Conjugate cone characterization of positive definite and semidefinite matrices. *Linear Algebra Appl.* **56**, 89–103 (1984)
19. Anderson, W.N., Duffin, R.J.: Series and parallel addition of matrices. *J. Math. Anal. Appl.* **26**(3), 576–594 (1969)
20. Azé, D., Hiriart-Urruty, J.-B.: *Analyse variationnelle et optimisation*. Cepadues, Toulouse (2010)
21. Taussky, O.: The role of symmetric matrices in the study of general matrices. *Linear Algebra Appl.* **5**, 147–154 (1972)
22. Bosch, A.: The factorization of a square matrix into two symmetric matrices. *American Math. Monthly*, 462–464 (1986)
23. Taussky, O., Zassenhaus, H.: On the similarity transformation between a matrix and its transpose. *Pacific Journal of Mathematics*, 893–896 (1959)
24. Wigner, E.P.: On weakly positive matrices. *Can. J. Math.* **15**, 313–317 (1963)
25. Ballantine, C.S.: Products of positive definite matrices III. *J. Algebra* **10**, 174–182 (1968)
26. Lewis, A.S.: Eigenvalue-constrained faces. *Linear Algebra Appl.* **269**, 1–18 (1997)
27. Borwein, J.M., Wolkowicz, H.: Facial reduction for a cone-convex programming problem. *J. Aust. Math. Soc.* **30**(3), 369–380 (1981)
28. Krislock, N., Wolkowicz, H.: Explicit sensor network localization using semidefinite representations and facial reductions. *SIAM J. Optim.* **20**(5), 2679–2708 (2010)
29. Malick, J.: A dual approach to semidefinite least-squares problems. *SIAM J. Matrix Anal. Appl.* **26**(1), 272–284 (2004)
30. Schwertman, N., Allen, D.: Smoothing an indefinite variance–covariance matrix. *J. Stat. Comput. Simul.* **9**, 183–194 (1979)
31. Lewis, A.S., Malick, J.: Alternating projections on manifolds. *Math. Oper. Res.* **33**(1), 216–234 (2008)
32. Hiriart-Urruty, J.-B., Ye, D.: Sensitivity analysis of all eigenvalues of a symmetric matrix. *Numer. Math.* **70**(1), 45–72 (1995)
33. Overton, M., Womersley, R.: Optimality conditions and duality theory for minimizing sums of the largest eigenvalues of symmetric matrices. *Math. Program.* **62**(2), 321–358 (1993)

34. d'Aspremont, A., El Ghaoui, L., Jordan, M.I., Lanckriet, G.R.: A direct formulation for sparse PCA using SDP. *SIAM Rev.* **49**, 434–448 (2007)
35. Nesterov, Y.: Smoothing technique and its applications in semidefinite optimization. *Math. Program.* **110**(2), 245–259 (2007)
36. Bochnak, J., Coste, M., Roy, M.-F.: *Real Semialgebraic Geometry. Results in Mathematics and Related Areas*, vol. 36. Springer, Berlin (1998)
37. Bolte, J., Daniilidis, A., Lewis, A.: The Lojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. *SIAM J. Optim.* **17**(4), 1205–1223 (2006)
38. Ioffe, A.: An invitation to tame optimization. *SIAM J. Optim.* **19**(4), 1894–1917 (2009)
39. Moakher, M.: *Symmetric positive-definite matrices: From geometry to applications and visualization. In: Visualization and Processing of Tensor Fields* (2006)
40. Bonnabel, S., Sepulchre, R.: Riemannian metric and geometric mean for positive semidefinite matrices of fixed rank. *SIAM J. Matrix Anal. Appl.* **31**(3), 1055–1070 (2009)
41. Bauschke, H.H., Moffat, S.M., Wang, X.: The resolvent average for positive semidefinite matrices. *Linear Algebra Appl.* **432**(7), 1757–1771 (2010)
42. Hiriart-Urruty, J.-B.: *Optimisation et analyse convexe. Exercices corrigés.* EDP Sciences, Les Ulis Cedex A (2009)
43. Nesterov, Y., Nemirovski, A.S.: *Interior-Point Polynomial Algorithms in Convex Programming. SIAM Studies in Applied Mathematics*, vol. 13. SIAM, Philadelphia (1994)
44. Boyd, S., El Ghaoui, L., Feron, E., Balakrishnan, V.: *Linear Matrix Inequalities in System and Control Theory. Studies in Applied Mathematics.* SIAM, Philadelphia (1994)
45. Goemans, M., Williamson, D.: Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. ACM* **6**, 1115–1145 (1995)
46. Nesterov, Y., Todd, M.J.: On the Riemannian geometry defined by self-concordant barriers and interior-point methods. *Found. Comput. Math.* **2**, 333–361 (2002)
47. Ekeland, I.: Plenary talk at the workshop of the group MODE of the French applied mathematical society, Institut Henri Poincaré, Paris (1997)
48. Bonnans, J.F., Gilbert, J.Ch., Lemaréchal, C., Sagastizábal, C.: *Numerical Optimization.* Springer, Berlin (2003)
49. Ekeland, I.: Some applications of the Cartan–Kähler theorem to economic theory. In: Berndt, R., Riemenschneider, O. (eds.) *Erich Kähler: Mathematische Werke/Mathematical Works* (2003)
50. Chabrilac, Y., Crouzeix, J.-P.: Definiteness and semidefiniteness of quadratic forms revisited. *Linear Algebra Appl.* **63**, 283–292 (1984)
51. Hiriart-Urruty, J.-B.: A new series of conjectures and open questions in optimization and matrix analysis. *ESAIM Control Optim. Calc. Var.* **15**, 454–470 (2009)
52. Hiriart-Urruty, J.-B., Seeger, A.: A variational approach to copositive matrices. *SIAM Rev.* **52**(4), 593–629 (2010)