Cambridge University Press
978-1-108-47871-7 — Matrix Positivity
Charles R. Johnson , Ronald L. Smith , Michael J. Tsatsomeros
Excerpt
More Information

1

Background

1.1 Purpose

Here, we review some well-known mathematical concepts that are helpful in developing the ideas of this work. These are primarily in the areas of matrix analysis, convexity, and an important theorem of the alternative (linear inequalities). In the latter two cases, we mention all that is needed. In the case of matrices, see the general reference [HJ13], or a good elementary linear algebra book, for facts or notation we use without further explanation.

1.2 Matrices

1.2.1 Matrix and Vector Notation

We use $\mathbb{R}^n$ ($\mathbb{C}^n$) to denote the set of all n-component real (complex) vectors, thought of as columns, and $M_{m,n}(\mathbb{F})$ to denote the m-by-n matrices over a general field $\mathbb{F}$. Skipping the field means $\mathbb{F} = \mathbb{C}$ and $M_{n,n}(\mathbb{F})$ is abbreviated to $M_n(\mathbb{F})$.

Inequalities, such as $>$, $\geq$, are to be interpreted entry-wise, so that $x > 0$ means that all entries of a vector x are positive. If all entries of x are nonnegative, but not all zero, we write $x \geq 0, \neq 0$.

The **transpose** of $A = [a_{ij}] \in M_{m,n}$ is denoted by A^T and the **conjugate transpose** (or **Hermitian adjoint**) by A^*. The **Hermitian** and **skew-Hermitian** parts of A are, respectively, denoted by

$$H(A) = \frac{A + A^*}{2} \quad \text{and} \quad S(A) = \frac{A - A^*}{2}.$$

The **spectrum**, or set of eigenvalues, of $A \in M_n$ is denoted by $\sigma(A)$ and the **spectral radius** by $\rho(A) = \max_{\lambda \in \sigma(A)} |\lambda|$.

Submatrices play an important role in analyzing matrix structure. For $\langle m\rangle = \{1, 2, \ldots, m\}$ and $\langle n\rangle = \{1, 2, \ldots, n\}$, denote by $A[\alpha, \beta]$ the **submatrix** of $A \in M_{m,n}(\mathbb{F})$ lying in rows $\alpha \subseteq \langle m\rangle$ and columns $\beta \subseteq \langle n\rangle$. If $m = n$ and $\alpha = \beta$, we abbreviate $A[\alpha, \alpha]$ to $A[\alpha]$ and refer to it as a **principal submatrix** of A; we call $A[\alpha]$ a **proper principal submatrix** if α is a proper subset of $\langle n\rangle$. A submatrix of $A \in M_n$ of the form $A[\langle k\rangle]$ for some $k \le n$ is called a **leading principal submatrix**. A submatrix may also be indicated by deletion of row and column indices, and for this round brackets are used. For example, $A(\alpha, \beta) = A[\alpha^c, \beta^c]$, in which the complementation is relative to $\langle m\rangle$ and $\langle n\rangle$, respectively. If an index set is a singleton i, we abbreviate $A(\{i\})$, for example, to $A(i)$ in case $m = n$.

Given a square submatrix $A[\alpha, \beta]$ of $A \in M_{m,n}$, we refer to $\det(A[\alpha, \beta])$ as a **minor** of A, or as a **principal minor** if $\alpha = \beta$. The determinant of a leading principal submatrix is referred to as a **leading principal minor**. By convention, if $\alpha = \emptyset$, then $\det(A[\alpha]) = 1$.

1.2.2 Gershgorin's Theorem

For $A = [a_{ij}] \in M_n(\mathbb{C})$, for each $i = 1, 2, \ldots, n$, define

$$R_i'(A) = \sum_{j=1, j\neq i}^{n} |a_{ij}| \quad \text{and the discs} \quad \Gamma_i(A) = \{z \in \mathbb{C} : |z - a_{ii}| \le R_i'(A)\}.$$

Gershgorin's Theorem then says that $\sigma(A) \subseteq \cup_{i=1}^{n} \Gamma_i(A)$. This has the implication that a **diagonally dominant matrix** $A \in M_n(\mathbb{C})$ (i.e., $|a_{ii}| > R_i'(A)$, $i = 1, 2, \ldots, n$) has nonzero determinant. In particular, the determinant is of the same sign as the product of the diagonal entries, in the case of real matrices.

1.2.3 Perron's Theorem

If $A \in M_n(\mathbb{R})$, $A > 0$, then very strong spectral properties follow as initially observed by Perron (**Perron's Theorem**). These include

- $\rho(A) \in \sigma(A)$; (1.2.1)
- the multiplicity of $\rho(A)$ is one; (1.2.2)
- $\lambda \in \sigma(A),\ |\lambda| = \rho(A) \implies \lambda = \rho(A)$; (1.2.3)
- there is a right (left) eigenvector of A with all entries positive; (1.2.4)
- $0 < B \le A$, $B \neq A$ implies $\rho(B) < \rho(A)$; (1.2.5)
- No eigenvector (right or left) of A has all nonnegative entries besides those associated with $\rho(A)$. (1.2.6)

There are slightly weaker statements for various sorts of nonnegative matrices; see [HJ91, chapter 8].

1.2.4 Schur Complements

If

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \in M_n$$

and A_{22} is square and invertible, the **Schur complement** (of A_{22} in A) is

$$A/A_{22} = A_{11} - A_{12}A_{22}^{-1}A_{21}.$$

More generally, if $\alpha \subseteq \langle n \rangle$ is an index set, then

$$A/A[\alpha] = A(\alpha) - A[\alpha^c, \alpha]A[\alpha]^{-1}A[\alpha, \alpha^c]$$

is the Schur complement of $A[\alpha]$ in A if $A[\alpha]$ is invertible. Schur complements enjoy many nice properties, such as

$$\det A = \det A[\alpha] \ \det(A/A[\alpha]),$$

which motivates the notation. Reference [Zha05] is a good reference on Schur complements and contains a detailed discussion of Schur complements in positivity classes.

1.3 Convexity

1.3.1 Convex Sets in $\mathbb{R}^n$ and $M_{m,n}(\mathbb{R})$

A **convex combination** of a collection of elements of $\mathbb{R}^n$ is a linear combination whose coefficients are nonnegative and sum to 1. A subset S of $\mathbb{R}^n$ or $M_{m,n}(\mathbb{R})$ is **convex** if it is closed under convex combinations. It suffices to know closure for pairs of elements, and geometrically this means that a set is convex if the line segment joining any two elements lies in the set.

An **extreme point** of a convex set is one that cannot be written as a convex combination of two distinct points in the set. The **generators** of a convex set are a minimal set of elements from which any element of the set may be written as a convex combination. For finite dimensions (our case), the extreme points are the generators. If there are finitely many, the convex set is called **polyhedral**.

A (convex) **cone** is just a convex set that is closed under linear combination whose coefficients are nonnegative (no constraint on the sum). A cone may also be finitely generated (polyhedral cross section) or not. The **dual** of a cone $\mathcal{C}$ is just the set

$$\mathcal{C}^D = \{x\colon x \cdot y \geq 0, \quad \text{whenever} \quad x \in \mathcal{C}\},$$

in which $\cdot$ denotes an inner product defined on the underlying space.

The **convex hull** of a collection of points is just the set of all possible convex combinations. This is a convex set, and any convex set is the convex hull of its generators.

1.3.2 Helly's Theorem

The intersection of finitely many convex sets is again a convex set (possibly empty). Given a collection of convex sets (possibly infinite) in a d dimensional space, the intersection of all of them is nonempty if and only if the intersection of any $d + 1$ of them is nonempty. This is known as **Helly's Theorem** [Hel1923] and can provide a powerful tool to show the existence of a solution to a system of equations.

1.3.3 Hyperplanes and Separation of Convex Sets

In an inner product space of dimension d, such as $\mathbb{R}^d$, a special kind of convex set is a **half-space**

$$H_a = \{x\colon a \cdot x \geq 0\},$$

in which $a \neq 0$ is a fixed element of the inner product space. Any intersection of half-spaces is a convex set, and any closed convex set is an intersection of half-spaces. The complement of a half-space,

$$H_a^c = \{x\colon a \cdot x < 0\},$$

is an open half-space. The set

$$\{x\colon a \cdot x = 0\}$$

is a $(d - 1)$-dimensional **hyperplane** that separates the two half-spaces H_a and H_a^c. If two convex sets S_1, S_2 in a d-dimensional space do not intersect, then they may be separated by a $(d - 1)$-dimensional hyperplane, so that $S_1 \subseteq H_a$ and $S_2 \subseteq H_{-a}$. If S_1 and S_2 are both closed or both open, we may use nonintersecting open half-spaces: $S_1 \subseteq H_a^c$ and $S_2 \subseteq H_{-a}^c$. If S_1 and S_2 do intersect but only in a subset of a hyperplane, then such a hyperplane may be used to separate them: $S_1 \subseteq H_a$ and $S_2 \subseteq H_{-a}$.

For further background on convex sets, see the general reference [Rock97].

1.4 Theorem of the Alternative

In the theory of linear inequalities (or optimization), there is a variety of statements saying that exactly one of two systems of linear inequalities has a solution. Such statements, and there are many variations, are called **Theorems of the Alternative** (and there are such theorems in even more general contexts). Such statements can be a powerful tool for showing that one system of inequalities has a solution, by ruling out the other; however, they generally are not able to provide any particular solution. The book [Man69] provides nice discussion of several theorems of the alternative and relations among them.

A particular version of the theorem of the alternative that is especially useful for us is the following:

Theorem 1.4.1 Let $A \in M_{m,n}(\mathbb{R})$. Then, either

(i) there is an $x \in \mathbb{R}^n$, $x \geq 0$, such that $Ax > 0$

or

(ii) there is a $y \in \mathbb{R}^m$, $y \geq 0$, $y \neq 0$, such that $y^T A \leq 0$,

but not both.

The "not both" portion is clear, as by (i), $y^T Ax > 0$ and because of (ii), $y^T Ax \leq 0$. In the event that $m = n$ and A is symmetric, we have

Corollary 1.4.2 Let $A \in M_n(\mathbb{R})$ and $A^T = A$. Then, either

(i) there is an $x \in \mathbb{R}^n$, $x \geq 0$, such that $Ax > 0$

or

(ii) there is a $y \in \mathbb{R}^m$, $y \geq 0$, $y \neq 0$, such that $Ay \leq 0$,

but not both.

2

Positivity Classes

2.1 Introduction

There are a remarkable number of ways that the notion of positivity for scalars has been generalized to matrices. Often, these represent the differing needs of applications, or differing natural aspects of the classical notion. Typically, the various ways involve the entries, transformational characteristics, the minors, the quadratic form, and combinations thereof. Our purpose here is to identify each of the generalizations and some of their basic characteristics. Usually there are natural variations that we also identify and relate. Several of these generalizations have been treated, in some depth, in book or survey form elsewhere; if so, we give some of the most prominent or accessible references. Our purpose in this work is to then treat in subsequent chapters those generalizations for which there seems not yet to be sufficiently general treatment in one place.

The order in which we give the generalizations is roughly grouped by type. We then summarize the containments among the positivity classes; one of them includes all the others.

2.2 (Entry-wise) Positive Matrices (EP)

One of the most natural generalizations of a positive number is reflected in the entries of a matrix. In general, adjectives, such as positive and nonnegative, refer to the entries of a matrix. The ways in which n-by-n positive matrices generalize the notion of a positive number are indicated in Perron's theorem; see Section 1.2.3.

Careful treatments of the theory of positive matrices may be found in several sources, such as [HJ13]. There are many further facts. Because of Frobenius's work on nonnegative matrices, the general theory is referred to as **Perron–Frobenius theory**.

There are important variants on positive matrices. For their closure, the nonnegative matrices, the spectral radius is still an eigenvalue, but all the other conclusions must be weakened somewhat. For nonnegative and **irreducible matrices** $\left(P^TAP \neq \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}\right.$, with A_{11} and A_{22} square and nonempty for any permutation $\left.P\right)$, only property (1.2.3) must be weakened to allow other eigenvalues on the spectral circle. And, if some power, A^q, is positive, the stronger conclusions remain valid. The following are simple examples that illustrate what might occur.

$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ The Perron root (spectral radius) is strictly dominant when the matrix is entry-wise positive.

$\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$ It remains so if the matrix is not positive but some power is positive (primitive matrix).

$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ The Perron root remains of multiplicity 1, but there may be ties for spectral radius when the matrix is irreducible but not primitive.

$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ The Perron root may be multiple and geometrically so, or not, when the matrix is reducible.

$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$ Or the Perron root may still have multiplicity 1, even when the matrix is reducible.

$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ The Perron root may be 0.

The **nonnegative orthant** in $\mathbb{R}^n$ is the closed cone that contains all entry-wise nonnegative vectors and is denoted by $\mathbb{R}^n_+$. The n-by-n **nonnegative matrices** ($A \geq 0$) are simply those that map the nonnegative orthant in $\mathbb{R}^n$ into itself ($A\mathbb{R}^n_+ \subseteq \mathbb{R}^n_+$). The n-by-n **positive matrices** ($A > 0$, or $A \in$ **EP** that stands for entry-wise positive) are those that map the nonzero elements of $\mathbb{R}^n_+$ into the interior of this cone. Matrices that map other cones of $\mathbb{R}^n$ with special structure into themselves emulate the spectral structure of nonnegative matrices, and this has been studied from several points of view in some detail.

Also studied have been matrices with some negative entries that still enjoy some or all of the above Perron conclusions or their Frobenius weakenings: $A \in M_n(\mathbb{R})$ is **eventually nonnegative (positive)** if there is an integer k such that $A^p \geq 0$ (> 0) for all $p \geq k$.

2.3 M-Matrices (M)

A matrix $A = [a_{ij}] \in M_n(\mathbb{R})$ is called a **Z-matrix** ($A \in$ **Z**) if $a_{ij} \leq 0$ for all $i \neq j$. Thus, a Z-matrix A may be written $A = \alpha I - P$ in which P is a nonnegative matrix. If $\alpha > \rho(P)$, then A is called a (non-singular) **M-matrix** ($A \in \mathbf{M}$). In several sources ([BP94, HJ91, FP62, NP79, NP80]) there are long lists of rather diverse-appearing conditions that are equivalent to A being an M-matrix, provided A is a Z-matrix. A modest list is the following:

1. A is **positive stable**, i.e., all eigenvalues of A are in the open right half-plane.
2. The leading principal minors of A are positive;
3. All principal minors of A are positive, i.e., A is a **P-matrix**;
4. A^{-1} exists and is a nonnegative matrix;
5. A is a semipositive matrix;
6. There exists a positive diagonal matrix D such that AD is row diagonally dominant;
7. There exist positive diagonal matrices D and E such that EAD is row and column diagonally dominant;
8. For each $k = 1, \ldots, n$, the sum of the k-by-k principal minors of A is positive;
9. A has an L–U factorization in which L and U have positive diagonal entries.

   In addition, M-matrices are closed under positive scalar multiplication, extraction of principal submatrices, and of Schur complements, and the so-called **Fan product** (which is the entry-wise or Hadamard product, except that the off-diagonal signs are retained). They are not closed under either addition or matrix multiplication.

   M-matrices, $A = [a_{ij}] \in M_n(\mathbb{R})$, also satisfy classical **determinantal inequalities**, such as

10. **Hadamard's inequality**: $\det A \le \prod_{i=1}^{n} a_{ii}$;
11. **Fischer's inequality**: $\det A \le \det A[\alpha] \det A[\alpha^c],\ \alpha \subseteq \{1, \ldots, n\}$;
12. **Koteljanskii's inequality**: $\det A[\alpha \cup \beta]\ \det A[\alpha \cap \beta] \le \det A[\alpha] \det A[\beta]$, $\alpha, \beta \subseteq \{1, \ldots, n\}$.

A complete description of all such principal minor inequalities is given in [Joh98].

M-matrices are not only positive stable, but, among Z-matrices, when all real eigenvalues are in the right half-plane, all eigenvalues are as well. They also have **positive diagonal Lyapunov solutions**, i.e., there is a positive diagonal matrix D such that $DA + A^T D$ is positive definite.

If $A = I - P$ is an irreducible M-matrix and we assume that $\rho(P) < 1$, then A is invertible and $A^{-1} = I + P + P^2 + \cdots$ Because P is irreducible, A^{-1} is positive. If A had been reducible, A^{-1} would be nonnegative. A matrix is **inverse M** (**IM**) if it is the inverse of an M-matrix, or if it is a nonnegative invertible matrix whose inverse is a Z-matrix. Among nonnegative matrices, the inverse M-matrices have a great deal of structure, and Chapter 5 is devoted to developing that structure for the first time in one place.

2.4 Totally Positive Matrices (TP)

A much stronger positivity requirement than that a matrix be positive is that all of its minors be positive. Such a matrix is called **totally positive** (**TP**). Such matrices arise in a remarkable variety of ways, and, of course, they also have very strong properties. The eigenvalues are positive and distinct, and the eigenvectors are highly structured in terms of the signs of their entries relative to the order in which the eigenvalue lies relative to the other eigenvalues. They always have an L–U factorization in which every minor of L and U is positive, unless it is identically 0. The determinantal inequalities of Hadamard, Fischer, and Koteljanskii (in Section 2.3) are satisfied, as well as many different ones. Transformationally, if A is **TP**, Ax cannot have more sign changes than x, and, of course, the **TP** matrices are closed under matrix multiplication. Though the definition requires many minors to be positive, because of Sylvester's determinantal identity, relatively few need be checked; the **contiguous minors** (both index sets are consecutive) suffice, and even the **initial minors** (those contiguous minors of which at least one index set begins with index 1) suffice. The initial minors are as numerous as the entries.

There are several comprehensive sources available for **TP** (and related) matrices, including the most recent book [FJ11]. Prior sources include the book [GK1935] and the survey [And80].

There are a number of natural variants on **TP** matrices. A **totally nonnegative** matrix, **TN**, is one in which all minors are nonnegative. **TN** is the topological closure of **TP**, and the properties are generally weaker. Additional variants are $\mathbf{TP}_k$ and $\mathbf{TN}_k$ in which each k-by-k submatrix is **TP** (respectively, **TN**).

2.5 Positive Definite Matrices (PD)

Perhaps the most prominent positivity class is defined by the quadratic form for Hermitian matrices. Matrix $A \in M_n(\mathbb{R})$ is called **positive definite** ($A \in$ **PD**) if A is Hermitian ($A^* = A$) and, for all $0 \neq x \in \mathbb{C}^n$, $x^T Ax > 0$.

There are several good sources on **PD** matrices, including [HJ13, Joh70, Bha07].

The **PD** matrices are closed under addition and positive scalar multiplication (they form a cone in $M_n(\mathbb{C})$), and under the Hadamard product, but not under conventional matrix multiplication. They are closed under extraction of principal submatrices and Schur complements. Among Hermitian matrices, all positive eigenvalues, positive leading principal minors, and all principal minors positive are each equivalent to being **PD**. A matrix $A \in M_n(\mathbb{C})$, is **PD** if and only if $A = B^*B$, with $B \in M_n(\mathbb{C})$, nonsingular; B may always be taken to be upper triangular (**Cholesky factorization**).

The standard variation on **PD** is **positive semidefinite** matrices, **PSD**, in which the quadratic form is only required to be nonnegative. **PSD** is the closure of **PD**, and most of the weakened properties follow from this. Of course, **negative definite** (**ND** $= -$**PD**) and **negative semidefinite** (**NSD** $= -$**PSD**) are not generalizations of positivity.

Another important variation is, of course, that the Hermitian part $H(A) = \frac{A+A^*}{2}$ is **PD**. In the case of $M_n(\mathbb{R})$ this just means that the quadratic form is positive, but the matrix is not required to be symmetric. Another variation is that $H(A)$ be **PSD**. There are some references that consider the former class, e.g., [Joh70, Joh72, Joh73, Joh75a, Joh75b, Joh75c, BaJoh76].

2.6 Strictly Copositive Matrices (SC)

An important generalization of **PSD** matrices is the copositive matrices (**C**) for which it is only required that the quadratic form be nonnegative on nonnegative vectors: $A \in M_n(\mathbb{R})$ is **copositive** if $A^T = A$ and $x^TAx \geq 0$ for all $x \geq 0$.

Much theory has been developed about the subtle class of copositive matrices. Because there is no comprehensive reference, Chapter 6 is devoted to this class. Variations include the **strictly copositive** matrices (**SC**) for which positivity of the quadratic form is required on nonnegative, nonzero vectors, and **copositive** $+$, the copositive matrices for which $x \geq 0$ and $x^TAx = 0$ imply $Ax = 0$. Also, the real matrices with Hermitian part in **C** or **SC** could be considered.

2.7 Doubly Nonnegative Matrices (DN)

The intersection of the cone of (symmetric) nonnegative matrices and the cone of **PD** matrices in $M_n(\mathbb{R})$ is the cone of **doubly nonnegative matrices** (**DN**). Natural variations include the closure of **DN** (nonnegative and **PSD**) and the **doubly positive matrices** (**DP**), which are positive and **PD**. The most natural