A copositivity probe

Wilfred Kaplan*\(^1\)

*Department of Mathematics, 3868 East Hall, University of Michigan, Ann Arbor, MI 48109-1109, USA

Received 7 August 2000; accepted 7 April 2001

Submitted by R.A. Brualdi

Abstract

The paper explores ways of determining whether a given symmetric matrix is copositive. In particular, a computational procedure is proposed for determining (if it exists) a representation of the matrix as a sum of a positive semidefinite matrix and a nonnegative matrix. The procedure is found to be successful in a significant number of cases. © 2001 Elsevier Science Inc. All rights reserved.

AMS classification: 15A18; 15A48; 15A63; 49N10; 65F30; 65K99; 90C22

Keywords: Copositive matrices; Eigenvalues; Optimization

1. Introduction

A real symmetric matrix \(A\) of order \(p\) is said to be copositive if \(x^T Ax \geq 0\) for \(x \geq 0\). It is termed strictly copositive if it is copositive and equality holds only for \(x = 0\). One also terms the quadratic form \(x^T Ax\) copositive or strictly copositive according as the matrix \(A\) has the corresponding property. There is an extensive literature on such matrices; see for example [1,5,11,13,19] and references cited therein.

Copositive matrices occur in optimization theory. See for example [6, Chapter 3], [8,9], [12, p. 133] and [17, Chapter 2].

In the paper [13] the author provided the following necessary and sufficient conditions for copositivity and strict copositivity.
Theorem 1.1. Matrix A is copositive (strictly copositive) if and only if every principal submatrix B of A has no eigenvector \( v > 0 \) with associated eigenvalue \( \lambda < 0 \) (\( \lambda \leq 0 \)).

This result has practical value only for matrices of low order, since the number of principal submatrices of a matrix of order \( p \) is roughly \( 2^p \). Other known methods for testing for copositivity (outside of the easy cases, such as positive semidefinite or nonnegative matrices) have similar disadvantages. A result of Murty and Kabadi [16] in fact shows that the general problem of testing for copositivity is NP-complete.

The present paper is motivated by the goal of revealing whether a symmetric matrix \( A \) is equal to the sum of a positive semidefinite matrix \( S \) and a nonnegative matrix \( P \). Every such matrix \( A \) is copositive and it is strictly copositive if \( S \) is positive definite. In the paper [10] Diananda conjectured that every copositive matrix was equal to such a sum and proved the conjecture to be true for matrices of order at most 4. However, Horn presented an example (see [10,11]) of a matrix of order 5 not representable as such a sum, and it is known that for all orders at least 5 the conjecture fails (see [2–4,11]).

Despite these negative results, we have found it worthwhile to pursue the goal stated. The matrices \( A \) representable as \( S + P \), as above, form a large class, which we denote by \( \mathcal{SP} \). If we can identify the members of \( \mathcal{SP} \) by a rapid procedure, then we have clearly made a useful contribution to the general problem of testing for copositivity.

In this paper, we present such a procedure in the form of a very simple algorithm. The idea behind the algorithm is a tool well known in quantum mechanics: the formulas for perturbation of eigenvalues and eigenvectors. We use only the simplest of these formulas; further study may show that additional ones are of value.

Much experimentation leads us to conjecture that the procedure is successful for those matrices \( A \) in \( \mathcal{SP} \) for which in the representation \( A = S + P \) the matrix \( S \) can be chosen to have a sufficiently large minimum eigenvalue. Obtaining a precise criterion appears to require a profound analysis of the way in which eigenvalues and eigenvectors vary as the entries in a matrix are varied.

If a matrix \( A \) being tested is not copositive, then of course the algorithm must fail; however, in many such cases the algorithm produces proof that \( A \) is not copositive by producing a positive vector \( v \) such that \( v^TAv < 0 \). This unexpected byproduct adds value to the procedure. Unfortunately, it does not always appear and thus far we have no proof that it must appear for a definite class of matrices.

We have described the procedure as “rapid”, without explaining the term. The algorithm iterates a simple step \( m \) times. At each step one computes all eigenvalues and eigenvectors of a symmetric matrix of order \( p \). The size of \( m \) cannot be guaranteed, but experiments show that, even for matrices of large order, \( m \) is normally about the size of \( p \). These statements give some idea of how much computation is involved. There are some ways in which the computation can be shortened. For example, one needs only the lowest eigenvalue of the matrix, not all eigenvalues, and
only the (normalized) eigenvector corresponding to that eigenvalue. Also, additional perturbation formulas could be used to reduce the need for computing eigenvalues and eigenvectors.

This discussion shows that much further study is required to make clear the benefits and liabilities of the algorithm. The present paper may nevertheless be of value for immediate applications. For one can simply try the algorithm; if it succeeds, one has a proof of copositivity; if it fails and has the byproduct mentioned, one has a proof of noncopositivity.

Convention. Throughout this paper all matrices will be assumed to be real and symmetric.

Remark 1. We offer no proof that the algorithm is successful for a certain class of matrices nor do we provide bounds on the number of iterations needed. Accordingly, this article is partly of heuristic nature.

Remark 2. The determination of representability of $A$ as $S + P$ can be treated as testing the feasibility of an appropriate semidefinite program. The results of Porkolab and Khachiyan [18] indicate that the computational difficulty of this approach may well be as great as that for testing copositivity.

2. Perturbation formulas

Let $A$, $B$ be matrices of order $p$. Let $z$ be a simple eigenvalue of $A$. Then for small $t$, $A + tB$ has a simple eigenvalue close to $z$; in fact, the eigenvalues of $A + tB$ can be represented by power series in powers of $t$ and there is such a series with constant term $z$. We seek the first order correction, the term in $t$. In many quantum mechanics texts this and higher order corrections are calculated; one finds that the first order correction gives

$$z' = z + tu^T Bu$$  \hspace{1cm} (2.1)

as the eigenvalue of $A + tB$; here $u$ is a normalized (that is, of Euclidean norm 1) eigenvector of $A$ for the eigenvalue $z$. There is a similar formula for the perturbed eigenvector (normalized up to the first order):

$$u' = u + t \sum \frac{v^T Bu}{w - z} v,$$  \hspace{1cm} (2.2)

where the sum is over the other (normalized) eigenvectors $v$ of $A$, with corresponding eigenvalues $w$. (See [7, pp. 140–151], [14, Chapter 2] and [15, Chapter XI] for the theory and formulas mentioned.)

Formula (2.1) can be easily derived as follows: consider the perturbed problem

$$(A + tB)(u + \delta u) = (z + \delta z)(u + \delta u).$$
Expand, neglect second-order terms and multiply by $u^T$ from the left; then use $Au = zu$ to get

$$\delta z = tu^T Bu.$$ 

3. Reduction to case of unit diagonal

**Theorem 3.1.** Let $A$ be a symmetric matrix of order $p$ with at least one diagonal entry $a_{ii}$ equal to 0.

(a) If all diagonal entries of $A$ are 0, then $A$ is copositive if and only if $A$ is nonnegative.

(b) If not all diagonal entries of $A$ are 0, then the quadratic form $x^T Ax$ can be written uniquely as the sum of two quadratic forms, of which the first consists of the terms in $x^T Ax$ in the variables $x_i$ for which the corresponding entry $a_{ii}$ is not 0, and the second consists of the remaining terms in $x^T Ax$. $A$ is copositive if and only if the first quadratic form is copositive and the second has a nonnegative matrix.

**Proof.** For (a), copositivity of $A$ implies, by Theorem 1.1, that each second-order principal submatrix of $A$ has no negative eigenvalue with positive eigenvector; since the diagonal elements of the principal submatrix are 0, it follows that the other two entries are nonnegative. Hence $A$ must be a nonnegative matrix. The converse is immediate.

For (b), the matrix of the first quadratic form is a principal submatrix of $A$. Let $A$ be copositive. Then it follows from Theorem 1.1 that this matrix is also copositive. Consideration of second-order principal submatrices, as in (a), shows that the matrix of the second quadratic form is nonnegative. The converse is again immediate. □

This result shows that, in testing for copositivity, one can restrict attention to matrices with no zero diagonal entries. Copositivity implies that no diagonal entry can be negative (a special case of Theorem 1.1). Hence one is led to consider only matrices with positive diagonal entries.

We can further restrict attention to matrices whose diagonal entries are all equal to 1; we say that the matrix has unit diagonal. To justify this step, we remark that because of the positivity of the diagonal elements, a simple change of scale in the coordinates $x_i$ of the quadratic form $x^T Ax$ reduces this quadratic form to one whose matrix has unit diagonal.

We remark that representability of $A$ as $S + P$, as in Section 1, is unaffected by such a change of scale. Furthermore, we lose no generality by requiring that $S$ also has unit diagonal.
Theorem 3.2. Let matrix $A$ have unit diagonal. Then $A$ can be represented as $S + P$, where $S$ is positive semidefinite and $P$ is nonnegative, if and only if $A$ can be so represented with $S$ having unit diagonal.

Proof. If $A = S + P$, then we replace each diagonal entry of $S$ by 1 and each diagonal entry of $P$ by 0. This increases or preserves the diagonal entries of $S$, so that the new matrix is still positive semidefinite, and the new $P$ is still nonnegative; the equality $A = S + P$ still holds. The converse is immediate. □

Theorem 3.3. Let the symmetric matrix $A = \{a_{ij}\}$ have unit diagonal.

(a) If $A$ is copositive, then $a_{ij} \geq -1$ for all $i$ and $j$.
(b) If $A$ is positive semidefinite, then $|a_{ij}| \leq 1$ for all $i$ and $j$.

Proof. Assertion (a) follows from Theorem 1.1 by the consideration of principal submatrices of order 2. Assertion (b) follows by a similar reasoning for positive semidefinite matrices, since for each principal submatrix $B$ of $A$ the quadratic form $u^T Bu$ is also positive semidefinite. □

Remark 3. For (a) the extreme case is the matrix $A$ of order $p$ whose off-diagonal entries are all $-1$. This is a circulant matrix and one verifies that for $p > 2$ it is not copositive; in particular, $A$ has the positive eigenvector $(1, \ldots, 1)^T$ with eigenvalue $2 - p$. For $p = 2$, $A$ is positive semidefinite.

Notation. For symmetric matrices of a given order $p$, $\mathcal{S}$ denotes the set of matrices with unit diagonal which are positive semidefinite, $\mathcal{C}$ denotes the set of matrices with unit diagonal which are copositive, $\mathcal{O}$ denotes those with zero diagonal, $\mathcal{P}$ denotes those which are nonnegative and have a zero diagonal. As in Section 1, $\mathcal{SP}$ denotes those representable as $S + P$, with $S$ in $\mathcal{S}$ and $P$ in $\mathcal{P}$. We observe that

$$\mathcal{S} \subset \mathcal{SP} \subset \mathcal{C},$$

where the last inclusion is a proper one if and only if $p \geq 5$.

In general $E(C)$ denotes the minimum eigenvalue of matrix $C$.

4. The algorithm

We are given a matrix $A$ of order $p$ with unit diagonal and our goal is to diminish the off-diagonal entries of $A$ to such an extent that the resulting matrix is in $\mathcal{S}$. To achieve this, we shall make many steps of the form $C = A - tB$, where $B \in \mathcal{P}$ and $t > 0$. By the perturbation formula (2.1), each simple eigenvalue $z$ of $A$ changes to $z - tu^T Bu$, where $u$ is the normalized eigenvector corresponding to $z$. If we can make the minimum eigenvalue $z$ increase by this step, then we have
made progress in moving to a matrix in $\mathcal{S}$. But the number $u^T Bu$ is a quadratic form whose general term is $b_{ij} u_i u_j$. We make each term negative or 0 if we make $b_{ij}$ equal to 1 when $u_i u_j < 0$ and equal to 0 otherwise. With this choice our algorithm is determined. We repeat the process, if possible, until the minimum eigenvalue has become nonnegative. In any case, we are forced to stop if the eigenvector $u$ is nonnegative or nonpositive; if the corresponding minimum eigenvalue is negative, then the algorithm has failed to yield the desired decomposition of the given matrix.

In accordance with Theorem 3.3, a copositive matrix can have no entry below $-1$. The iterations called for in the algorithm could lead to a matrix $C$ with such an entry and one would appear to be forced to stop when this occurs. However, instead of doing so, we modify the algorithm by raising each such entry to $-1$. One verifies that this is equivalent to replacing some of the entries of $B$ by smaller nonnegative numbers (possibly by 0). The algorithm can then proceed as before and typically achieves its goal of lowering the entries of the matrix while increasing the minimum eigenvalue.

The modification permits the algorithm to continue until an iteration produces no change. This eventuality would arise either because one has produced a matrix $C$ having a minimum eigenvalue with nonnegative eigenvector or because the modification itself prevented a change; the latter case could occur if the nonzero entries in $B$ were all replaced by 0 because of the modification. In practice the second alternative has not been found and there appears to be a theoretical basis for excluding it; see open question no. 4 in Section 8. However, since it has not yet been proved to be disallowed, we incorporate a corresponding stopping rule in Step 7 of the algorithm.

We state the algorithm formally:

Step 1. If all $a_{ij}$ for $i < j$ are nonnegative, STOP. $A$ is copositive.
Step 2. If $E(A) \geq 0$, STOP. $A$ is positive semidefinite and copositive.
Step 3. If $a_{ij} < -1$ for some $a_{ij}$, STOP. $A$ is not copositive.
Step 4. Set $R = A$ and choose a value of parameter $t$ (for example, $t = 0.05$).
Step 5. Let $z = E(R)$ and let $u$ be a corresponding normalized eigenvector for $z$. If $u$ is nonnegative or nonpositive, evaluate $q = u^T Au$ and STOP; if $q < 0$, $A$ is not copositive; otherwise, the method fails.
Step 6. Define the matrix $B = \{b_{ij}\}$ of order $p$ by the rules:

$$b_{ij} = 1 \text{ when } u_i u_j < 0; \quad b_{ij} = 0 \text{ otherwise.}$$

Step 7. Let $C_1 = R - tB$. Let $C$ be obtained from $C_1$ by replacing each entry which is less than $-1$ by $-1$. If $C = R$, STOP; the method fails.
Step 8. If $E(C) \geq 0$, STOP; $A$ is copositive and $A = S + P$, where $S = C$ and $P = A - C$. Otherwise, set $R = C$ and go to Step 5.

We give several examples; all except the third were found by random sampling and rounding. The first two are remarkably simple.
Example 1.

\[ A = \begin{bmatrix}
1 & .9 & -.54 & .21 \\
.9 & 1 & -.03 & .78 \\
-.54 & -.03 & 1 & .52 \\
.21 & .78 & .52 & 1
\end{bmatrix}. \]

Here \( E(A) = -0.0955 \) with corresponding eigenvector \( u = (0.5707, -0.7288, 0.0865, 0.3684)^T \).

Steps 1–3 do not lead to stopping. We choose \( t = 0.05 \) and \( R = A \) in Step 4. Step 5 does not lead to stopping. Step 6 gives \( B \) as the matrix

\[ B = \begin{bmatrix}
0 & 1 & 0 & 0 \\
1 & 0 & 1 & 1 \\
0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0
\end{bmatrix}. \]

Step 7 gives \( C = C_1 = R - 0.05B \) with no entry less than \(-1\) and \( E(C) = -0.0222 \) with eigenvector \( (0.5432, -0.7368, 0.0248, 0.4019)^T \). We set \( R = C \) and return to Step 5. This again leads to Steps 5 and 6 with the same \( B \) and Step 7 gives a new \( C \):

\[ C = \begin{bmatrix}
1 & .8 & -.54 & .21 \\
.8 & 1 & -.13 & .68 \\
-.54 & -.13 & 1 & .52 \\
.21 & .68 & .52 & 1
\end{bmatrix}. \]

We find that \( E(C) = 0.0451 \), so that \( C \) is positive definite. Hence our goal is achieved and this \( C \) is our \( S \); \( A = S + P \), where \( P \) is \( 0.1B \). Here \( A \) is strictly copositive and is in the set \( \mathcal{P} \).

Example 2. We take \( A \) to be the matrix

\[ A = \begin{bmatrix}
1 & -.72 & -.59 & -.6 \\
-.72 & 1 & .21 & -.46 \\
-.59 & .21 & 1 & -.6 \\
-.6 & -.46 & -.6 & 1
\end{bmatrix}. \]

and proceed as before with \( t = 0.05 \). We find that the process stops with the second iteration, since \( C \) has a negative minimum eigenvalue with positive eigenvector \( u = (0.5994, 0.4123, 0.4114, 0.5491)^T \). Thus we have failed to find a representation of \( A \) as \( S + P \). However, we observe that \( u^TAu = -0.4499 \) and conclude that \( A \) is not copositive. This example illustrates the ‘byproduct’ mentioned in Section 1.

Example 3. \( A \) is the matrix of order 4:
We observe that \( A = S + P \), where
\[
S = \begin{bmatrix}
1 & 1 & -1 & -1 \\
1 & 1 & -1 & -1 \\
1 & -1 & 1 & 1 \\
-1 & -1 & 1 & 1
\end{bmatrix}, \quad P = \begin{bmatrix}
0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{bmatrix}.
\]
We verify that \( S \) is positive semidefinite, with eigenvalues 0, 0, 0, 4. Thus \( A \) is copositive. However, the algorithm fails for many values of \( t \). It is clear that \( S \) is very special here. In the following section we discuss the geometry involved and the reasons for the failure of the algorithm in such a case.

**Example 4.** Here \( A \) is a matrix of order 10:
\[
\begin{bmatrix}
1 & -0.08 & 0.69 & 0.43 & 1.21 & 0.74 & 0.23 & 0.82 & 0.50 & -0.19 \\
-0.08 & 1 & 0.46 & 0.56 & 0.88 & 0.05 & 0.47 & 0.54 & 0.77 & 0.23 \\
0.69 & 0.46 & 1 & 0.53 & 0.77 & 0.33 & 0.70 & 0.12 & 0.25 & 0.66 \\
0.43 & 0.56 & 0.53 & 1 & -0.07 & 0.33 & 0.69 & 0.81 & 0.33 & 0.18 \\
1.21 & 0.88 & 0.77 & -0.07 & 1 & 0.10 & 0.21 & 0.13 & 0.37 & 0.37 \\
0.74 & 0.05 & 0.33 & 0.33 & 0.10 & 1 & -0.09 & 1.08 & 0.81 & 0.62 \\
0.23 & 0.47 & 0.70 & 0.69 & 0.21 & -0.09 & 1 & 0.22 & -0.05 & 0.20 \\
0.82 & 0.54 & 0.12 & 0.81 & 0.13 & 1.08 & 0.22 & 1 & 0.82 & 0.19 \\
0.50 & 0.77 & 0.25 & 0.33 & 0.37 & 0.81 & -0.05 & 0.82 & 1 & 0.82 \\
-0.19 & 0.23 & 0.66 & 0.18 & 0.37 & 0.62 & 0.20 & 0.19 & 0.82 & 1
\end{bmatrix}
\]
We again use \( t = 0.05 \) and find that after 12 iterations we obtain a positive definite matrix \( C \):
\[
\begin{bmatrix}
1 & -0.13 & 0.34 & -0.02 & 0.61 & 0.34 & -0.12 & 0.37 & 0.15 & -0.34 \\
-0.13 & 1 & 0.16 & 0.16 & 0.33 & -0.30 & 0.07 & 0.04 & 0.37 & 0.03 \\
0.34 & 0.16 & 1 & 0.13 & 0.52 & -0.02 & 0.30 & -0.08 & 0.05 & 0.26 \\
-0.02 & 0.16 & 0.13 & 1 & -0.22 & 0.28 & 0.49 & 0.51 & 0.03 & -0.22 \\
0.61 & 0.33 & 0.52 & -0.22 & 1 & -0.10 & -0.04 & -0.02 & 0.12 & -0.08 \\
0.34 & -0.30 & -0.02 & 0.28 & -0.10 & 1 & -0.24 & 0.73 & 0.56 & 0.17 \\
-0.12 & 0.07 & 0.30 & 0.49 & -0.04 & -0.24 & 1 & -0.08 & -0.35 & -0.10 \\
0.37 & 0.04 & -0.08 & 0.51 & -0.02 & 0.73 & -0.08 & 1 & 0.62 & -0.11 \\
0.15 & 0.37 & 0.05 & 0.03 & 0.12 & 0.56 & -0.35 & 0.62 & 1 & 0.32 \\
-0.34 & 0.03 & 0.26 & -0.22 & -0.08 & 0.17 & -0.10 & -0.11 & 0.32 & 1
\end{bmatrix}
\]
Accordingly, \( A \) is a strictly copositive matrix in \( S^P \).

In implementing the algorithm, one has to select the value of \( t \). The value 0.05, used in the preceding examples, is often successful. If it fails, one may be able to
achieve success with another value, usually a smaller one, which would normally mean more iterations. As in Example 3, it can happen that no value of \( t \) leads to success.

The number of iterations varies considerably. In practice, it is found that the process always stops when a minimum eigenvalue has been found which is nonnegative or which has a positive eigenvector.

We observe that, for a fixed choice of \( t \), if all entries are bounded by a positive number \( k \), then at most \( N = \lfloor (k + 1)/t \rfloor + 1 \) (brackets for integral part) iterations would be needed to decrease an entry to \(-1\), so that at most \( nN \) iterations would be required to bring all entries down to \(-1\), when the process must stop (see Remark 3). This reasoning shows that the algorithm would terminate even without the stopping rule in Step 7, but it does not provide an estimate of the number of iterations needed for success, since it does not tell how \( t \) must be chosen for that goal.

5. Geometrical considerations

We consider matrices of order \( p \) with unit diagonal. Each such matrix is specified by giving the above-diagonal entries, \( p(p - 1)/2 = n \) in number. We interpret these entries as coordinates in an \( n \)-dimensional space \( \mathbb{R}^n \), to which frequent reference will be made; the elements of \( \mathbb{R}^n \) will be considered as points or vectors. The correspondence between these points and the matrices of order \( p \) with unit diagonal is one-to-one. Accordingly, it will be convenient to identify sets of matrices having unit diagonal with the corresponding sets in \( \mathbb{R}^n \). In particular, we denote by \( \mathcal{S} \), \( \mathcal{C} \) and \( \mathcal{SP} \) the sets in \( \mathbb{R}^n \) corresponding to these sets of matrices. By Theorem 3.3, all these sets are contained in the set \( \{ x \mid x \in \mathbb{R}^n, x_i \geq -1 \text{ for all } i \} \), which will be denoted by \( \mathbb{R}^n_0 \). The sets \( \mathcal{C} \) and \( \mathcal{P} \) have no counterpart in \( \mathbb{R}^n \).

The sets \( \mathcal{C} \), \( \mathcal{S} \) and \( \mathcal{SP} \) are closed, \( n \)-dimensional and convex, as one easily verifies. The set \( \mathcal{C} \) is unbounded; the points of \( \mathcal{S} \) have coordinates between \(-1 \) and \(+1 \), by Theorem 3.3, so that \( \mathcal{S} \) is bounded. One shows easily that the interior of \( \mathcal{C} \) consists of the points representing strictly copositive matrices and the interior of \( \mathcal{S} \) consists of the points representing positive definite matrices.

Although the set \( \mathcal{C} \) is not bounded, in testing whether a matrix \( A \) is in \( \mathcal{SP} \), we can reduce the problem to one for a bounded set. For if \( A \) is in \( \mathcal{SP} \), then so is the matrix \( A_1 \) obtained from \( A \) by replacing each entry \( a_{ij} \) by \( \min(a_{ij}, 1) \); this follows from the fact that in the equation \( A = S + P \), each entry of \( S \) is at most equal to \( 1 \), so that \( A \geq A_1 \geq S \). Conversely, if \( A_1 \) is in \( \mathcal{SP} \), then so is \( A \), since \( A \geq A_1 \). It now follows that it suffices to consider matrices \( A \) whose entries are between \(-1 \) and \( 1 \), inclusive.

However, it should be remarked that the matrix \( A_1 \) might fail to be copositive, even when \( A \) is copositive; this might conceivably occur in cases when \( A \) is in \( \mathcal{C} \) but not in \( \mathcal{SP} \). We know of no example to illustrate this possibility. If one could prove that it never occurs, then one could conclude that for determining copositivity
it suffices to consider matrices with unit diagonal and entries of absolute value at most 1.

For each vector \( v = (v_1, \ldots, v_n)^T \) in our space \( \mathbb{R}^n \) we have a corresponding matrix \( A \) of order \( p \). The set of all matrices \( B \) with unit diagonal such that \( B \preceq A \) corresponds to the convex polyhedron in \( \mathbb{R}^n \) consisting of all \( w \) in \( \mathbb{R}^n \) such that \( w_i \leq v_i \) for \( i = 1, \ldots, n \). \( A \) has a representation as \( S + P \) precisely when this polyhedron intersects the convex set \( \mathcal{S} \). The intersection is itself a convex set, but of course it may be empty or consist of a single point or consist solely of boundary points of the two convex sets. This intersection is our ‘target’. The goal of the algorithm is to go from the starting point \( v \) to a point of the intersection. If the intersection is empty, the goal cannot be achieved and we would like to know when that happens (perhaps through the ‘byproduct’). If the intersection is just one point or is a ‘tiny’ set, an algorithm of extreme precision would be called for and the one proposed above falls far short of that; it is solely an approximation and cannot be expected to lead the way to a very small target.

The hope is that the algorithm would succeed when the intersection contains interior points of \( \mathcal{S} \), which as we know correspond to positive definite matrices. In that case, there is some room for error and an approximate method has a chance of succeeding. If successful, it would find a matrix \( S \) such that \( A = S + P \) and, \( S \) being far from unique, would find different such matrices \( S \) when the parameter \( t \) is varied. The success and the variety of choices of \( S \) would be enhanced as the size of the intersection increased. In general, one would expect the matrix \( S \) to be positive definite.

These remarks are illustrated by the examples of the preceding section. For Examples 1 and 4 the algorithm succeeds and in each case \( S \) is positive definite. In both cases the step size \( t \) was chosen as 0.05. One verifies that choices near that value also succeed, but produce different matrices \( S \). Thus in these examples the target is large enough. For Example 2 the intersection is empty and the algorithm must fail.

Example 3 is worth close examination. Here we know that there is a matrix \( S \) such that \( A = S + P \), and \( S \) is only positive semidefinite, corresponding to a boundary point of \( \mathcal{S} \). To make matters worse, we find that the \( S \) exhibited is the only matrix allowable; it comprises the whole intersection in question. To show this, we consider the matrices \( D \preceq A \) and determine how \( D \) can be chosen to be in \( \mathcal{S} \). Since no entry of \( D \) can be less than \(-1\), we are forced to make \( d_{13}, d_{14}, d_{23} \) and \( d_{24} \) equal to \(-1\). Further, we require \(-1 \leq d_{12} \leq 2 \) and \(-1 \leq d_{34} \leq 1 \). If \( D \) is to be positive semidefinite, then it must be copositive. Thus we can apply Theorem 1.1 to \( D \). If we do so, considering the principal submatrices of order 3 using the indices 1, 2, 3 and 1, 3, 4, we conclude that \( 1 \leq d_{12} \leq 2 \) and \( d_{34} = 1 \) (see [1, p. 22]). We now consider the characteristic equation of \( D \) and find that, unless \( d_{12} = 1 \), the equation has a negative root. Thus our assertion is verified: \( S \) is the only matrix such that \( A = P + S \). With a target reduced to a single point, the algorithm has little chance of success.

However, the algorithm failed for an unexpected reason, but one still related to the size of the target. A careful examination of the execution of the algorithm for
this example shows that at the first iteration the computer gave a wrong result for the matrix $B$; the minimum eigenvalue of $A$ is $-1$ and this has the normalized eigenvector $(-\sqrt{2}/2, \sqrt{2}/2, 0, 0)^T$. The computer wrongly interpreted the first zero as a negative number and hence made $b_{23} = b_{32} = 1$. So the difficulty lies in the well-known troubles related to the number 0. With the correct calculation, the algorithm would have succeeded with $t = 0.05$ in 20 iterations.

That is not the end of the story: If we use $t = 0.049$ or $t = 0.051$, even with a correct computation, the algorithm would fail. For there is no room for error! The freedom to adjust the size of $t$ depends in an essential way on the size of the target; the larger the target, the more the freedom in choosing the step size.

We denote by $M$ the set of matrices of order $p$ with unit diagonal whose minimum eigenvalue has a nonnegative eigenvector (or the corresponding subset of $\mathbb{R}^n$). We assert that $M$ contains all points of $\mathbb{R}^n$ whose coordinates are nonpositive. This assertion is equivalent to the assertion that for every matrix $A$ with unit diagonal and nonpositive above-diagonal entries the quadratic form $\phi(x) = x^T Ax$ takes its minimum on the unit sphere $\|x\| = 1$ at a nonnegative vector $u$. The assumptions imply that for each $x$ on the unit sphere $\phi(x) \geq \phi(|x|)$, where $|x|$ is the vector with components $|x_i|$. Hence $\phi$ takes on its minimum for some $u \geq 0$.

**Remark 4.** Matrices with nonpositive off-diagonal entries are termed Z-matrices; see [6, Section 3.11]. There are matrices in $M$ other than Z-matrices; for example, the copositive matrix

$$
\begin{bmatrix}
1 & -.2 & -.6 & -.3 \\
-.2 & 1 & -.1 & .1 \\
-.6 & -1 & 1 & -.2 \\
-.3 & .1 & -.2 & 1
\end{bmatrix}.
$$

**Remark 5.** Every matrix $A$ in $M$ is either positive semidefinite (and hence copositive) or not positive semidefinite and not copositive. The first case occurs when the minimum eigenvalue of $\phi$ is nonnegative, as in the preceding displayed matrix; the second when the minimum eigenvalue is negative. (See [6, p. 203].)

The set $M$ has a special significance for the algorithm. For, as remarked in Section 4, in practice the algorithm stops only when one has reached a matrix in $S$, when $A$ is copositive, or in $M$, when $A$ is not copositive or the algorithm has failed.

### 5.1. Vector field interpretation

Step 6 of the algorithm can be interpreted as the assignment to the point $v$ in $\mathbb{R}^n_0$, representing matrix $A$, of a vector in $\mathbb{R}^n$ representing the matrix $I + B$ whose off-diagonal entries are 0’s and 1’s; if $v$ is in $M$, the vector is the zero vector, and we can also assign the zero vector to each point of $S$. The algorithm corresponds to the
primitive Euler method of solving the corresponding differential equation; since one subtracts $tB$ at each step, where $t$ is positive, one is solving the differential equation in the direction opposite to that of the vector field. Of course the vector field is not even continuous; it is piecewise constant over appropriate subregions and is defined only if the minimum eigenvalue of $A$ is simple. Following this interpretation, one might attempt to improve the results by applying advanced techniques for solving differential equations numerically. In particular, one could try first smoothing the vector field by appropriate averaging. Since in practice all the ‘orbits’ of the differential equation end in $M \cup S$, this set appears to be an attractor for the flow.

A related vector field arises from a different approach. To the vector $v$ one can assign the number $E(A)$. This scalar field is continuous and even analytic, except at vectors corresponding to matrices whose minimum eigenvalue is multiple. The corresponding gradient field has the directions we are interested in: those in which the minimum eigenvalue is increasing most rapidly. Again there is a differential equation. However, its solutions may not be acceptable for testing copositivity. For in each case a matrix $A$ is given and the solution must remain in the polyhedron corresponding to the matrices $C \preceq A$. Finding the differential equation is itself a formidable task, whether an analytic expression is sought or a computational procedure is employed. Thus this approach is not promising.

6. Boundaries of $S$ and $C$

In this section, we take advantage of Theorem 1.1 and the other theorems proved above to obtain quite explicit representations of the boundaries of $S$ and $C$.

We first observe that each point of the space $\mathbb{R}^n$ represents a matrix $I + C$, where $C \in \mathcal{C}$, and each ray in the space with endpoint at the origin represents a family of matrices $I + rC$, $0 \leq r \leq \infty$. By elementary matrix algebra, the eigenvalues of each of these matrices are the numbers $1 + r\lambda_j$, where the $\lambda_j$ are the eigenvalues of $C$; the corresponding eigenvectors are the eigenvectors of $C$. For other lines in the space one usually does not have this simple linear variation of eigenvalues. (See [14, Chapter 2].) A typical principal submatrix of $I + rC$ has the form $I + rB$ (here $I$ has the order of $B$), where the matrix $B$ is an arbitrary principal submatrix of $C$; hence its eigenvalues and eigenvectors are also related in a simple way to those of the matrix $B$. These remarks are the basis of the theorems to follow.

We let $S^{n-1}$ denote the $(n-1)$-sphere as the subset of $\mathbb{R}^n$ consisting of vectors $v$ of norm 1.

**Theorem 6.1.** The subset $S$ of $\mathbb{R}^n$ can be represented as

$$\left\{ x = rv \mid v \in S^{n-1}, \ 0 \leq r \leq \sigma(v) \right\}, \tag{6.1}$$

where $\sigma$ is continuous on $S^{n-1}$ with positive real values and upper bound $\sqrt{n}$. Furthermore, for all $v$
\[ \sigma(v) = -1/E(C), \quad (6.2) \]

where \( v \) represents the matrix \( A = I + C \) with \( C \in \mathcal{O} \).

**Proof.** Since \( \mathcal{S} \) is a convex body containing the origin of \( \mathbb{R}^n \) in its interior, it has representation (6.1) with continuous positive \( \sigma \). The bound \( \sqrt{n} \) follows from Theorem 3.3. The matrix \( A = I + rC \) is in \( \mathcal{S} \) precisely when \( E(A) \geq 0 \). Since \( E(A) = 1 + rE(C) \), formula (6.2) follows. \( \square \)

**Theorem 6.2.** The subset \( \mathcal{C} \) of \( \mathbb{R}^n \) can be represented as

\[ \left\{ x = rv \mid v \in S^{n-1}, \quad 0 \leq r\# \kappa(v) \right\}, \quad (6.3) \]

where \( \kappa \) maps \( S^{n-1} \) into the interval \( 0 \leq k \leq \infty \) and \( \# \) denotes \( \leq \) if \( \kappa(v) \) is finite and denotes \( < \) otherwise; at each \( v \) for which \( \kappa(v) \) is finite the function \( \kappa \) is continuous and this occurs precisely when \( v \) has at least one negative component \( v_j \); for such a vector \( v \)

\[ 0 < \kappa(v) \leq -\left[ \min_{1 \leq i \leq n} v_i \right]^{-1}. \quad (6.4) \]

For such a \( v \) let \( A = I + C \) be the matrix represented by \( v \), with \( C \in \mathcal{O} \). Then

\[ \kappa(v) = -\left[ \min_{B \in \mathcal{B}} \lambda(B) \right]^{-1}, \quad (6.5) \]

where \( \mathcal{B} \) is the set of all principal submatrices of \( C \) which have a positive eigenvector with negative eigenvalue \( \lambda(B) \).

**Proof.** As a convex set in \( \mathbb{R}^n \) having the origin as interior point, \( \mathcal{C} \) has a representation (6.3), with \( \kappa \) having the range and continuity properties stated. If \( v \) is non-negative, then all the vectors \( rv \) with \( r \geq 0 \) represent copositive matrices, so that \( \kappa(v) = \infty \). Let \( rv \) represent a copositive matrix. If \( v \) has a negative component \( v_j \), then Theorem 3.3 implies that \( rv_j \geq -1 \), so that (6.4) follows. Theorem 1.1 implies that no principal submatrix of \( I + rC \) has a positive eigenvector with negative eigenvalue. But the principal submatrices of \( I + rC \) are the matrices \( I + rB \), where \( B \) is any principal submatrix of \( C \) (and \( I \) has the order of \( B \)). If \( B \) has a positive eigenvector with negative eigenvalue \( \lambda(B) \), then \( I + rB \) has the same eigenvector but no negative eigenvalue precisely when \( r \leq -\lambda(B) \). Accordingly, relation (6.5) follows. \( \square \)

7. **Numerical experiments**

The algorithm has been tried on a great variety of matrices of orders up to 50. It has been found to prove copositivity in many cases and to prove non-copositivity in many others. Failure has been uncommon.
In particular, a set of 76 examples has been studied and preserved in a computer file. Of these, the first 47 were chosen as random matrices with entries between \(-1\) and \(+1\) and unit diagonal. The remaining 29 matrices were similarly restricted but were chosen by addition of a random positive definite matrix and a random nonnegative matrix; thus each was in \(S\mathcal{P}\).

In the first 47 the orders were 4, 5 and 6. Failure occurred for a copositive matrix (as verified by Theorem 1.1) in 8 cases out of 21. For the non-copositive matrices the byproduct appeared in all but 2 out of 26 cases. Thus one could say that the algorithm was successful in 37 out of 47 cases; the value 0.05 for \(t\) was used for all but 3 of the successful cases of copositivity and all but 2 of the successful cases of non-copositivity. In every case of copositivity it was strict.

In the remaining 29 examples, the orders ranged from 6 to 20. The algorithm succeeded in all cases, using \(t = 0.05\), showing strict copositivity.

In the successful cases with \(t = 0.05\) the number of iterations varied from 2 to 6 for orders up to 6, with a few exceptions, and from 5 to 18 for orders 7 to 20. Thus the number of iterations appears to be roughly equal to the order of the matrix. When a smaller value of \(t\) was used, the number became much larger. For example, one matrix of order 4 was found to be copositive with \(t = 0.01\) and 55 iterations.

In the cases of failure for copositive matrices, the test results suggest that the matrices were not in \(S\mathcal{P}\), but this question merits further study.

Additional experiments have been carried out with the aid of Theorems 6.1 and 6.2. These permit one to find the cross-sections of \(\mathcal{S}\) and \(\mathcal{C}\) in an arbitrary plane (2-dimensional) through the origin. One can then apply the algorithm of Section 4 to attempt to find the cross-section of \(S\mathcal{P}\) in the chosen plane. Preliminary results indicate that the points of \(\mathcal{C}\) attainable by the algorithm form a convex subset of \(\mathcal{C}\) extending well beyond \(\mathcal{S}\).

8. Open questions

1. Find sufficient conditions for success of the algorithm for a matrix \(A\) in \(S\mathcal{P}\) and obtain bounds for the number of iterations required for appropriate choice of the parameter \(t\).
2. Find sufficient conditions for occurrence of the ‘byproduct’ for a matrix \(A\) which is not copositive.
3. Determine the validity of the assertion: a matrix \(A = \{a_{ij}\}\) with unit diagonal is copositive if the matrix \(A_1\) with entries \(\min(a_{ij}, 1)\) is copositive.
4. Experiments indicate that the vector field corresponding to the matrices \(B\) (as described in Section 5) assigns to each boundary point of \(\mathbb{R}_0^n\) a vector tangent to the boundary. Determine whether this property is generally valid. In more detail: demonstrate that if at a boundary point certain coordinates equal \(-1\), then the corresponding entries of \(B\) are 0 at the point. (Points of the boundary at which \(B\) is not unambiguously defined must be excluded.)
5. Show, if possible, that for each $p$ there is a parameter value $t_0$ such that for each parameter value $t \leq t_0$ the algorithm succeeds for a convex set of matrices including the set $\mathcal{S}$.

6. Apply the vector field interpretation of Section 5 and determine the structure of the corresponding flow in $\mathbb{R}_0^n$.

References


