

Scaling of Symmetric Matrices by Positive Diagonal Congruence

Charles R. Johnson ^a and Robert Reams ^b

^a*Department of Mathematics, College of William and Mary, P.O. Box 8795,
Williamsburg, VA 23187-8795*

^b*Department of Mathematics, Virginia Commonwealth University,
1001 West Main Street, Richmond, VA 23284*

Abstract

We consider the problem of characterizing n -by- n real symmetric matrices A for which there is an n -by- n diagonal matrix D , with positive diagonal entries, so that DAD has row (and column) sums 1. Under certain conditions we provide necessary and sufficient conditions for the existence of a scaling for A , based upon both the positive definiteness of A on a cone lying in the nonnegative orthant and the semipositivity of A . This generalizes known results for strictly copositive matrices. Also given are (1) a condition sufficient for a unique scaling; (2) a characterization of those positive semidefinite matrices that are scalable; and (3) a new condition equivalent to strict copositivity, which we call total scalability. When A has positive entries, a simple iterative algorithm (different from Sinkhorn's) is given to calculate the unique scaling.

Key words: Sinkhorn's Theorem diagonal scaling semipositive copositive totally scalable doubly stochastic stochastic symmetric

MSC: 15A48 15A51 15A57 15A63

1 Introduction

For $x = (x_1, \dots, x_n)^T \in \mathbf{R}^n$ we will use the notation that $x \geq 0$ when $x_i \geq 0$ for $1 \leq i \leq n$, and $x > 0$ when $x_i > 0$ for all $1 \leq i \leq n$. We will say that a vector (or matrix) is *nonnegative* in the event that all of its entries are nonnegative. Let $e = (1, 1, \dots, 1)^T \in \mathbf{R}^n$. The *nonnegative orthant* is the set of all $x \in \mathbf{R}^n$ such that $x \geq 0$. We say that $A \in \mathbf{R}^{n \times n}$ is *semipositive* if there exists $x > 0$ so that $Ax > 0$. A symmetric matrix $A \in \mathbf{R}^{n \times n}$ is *positive*

semidefinite (positive definite) on a subset S of \mathbf{R}^n when $x^T Ax \geq 0$ ($x^T Ax > 0$) for all $x \in S$ ($x \in S, x \neq 0$). A symmetric matrix $A \in \mathbf{R}^{n \times n}$ is *copositive* (strictly copositive) when $x^T Ax \geq 0$ ($x^T Ax > 0$) for all $x \geq 0$ ($x \geq 0$ and $x \neq 0$). So, a copositive (strictly copositive) matrix A is positive semidefinite (positive definite) on the nonnegative orthant. A *cone* S is a nonempty subset of \mathbf{R}^n such that if $a, b \in S$ then $\mu a + \nu b \in S$, for all $\mu, \nu \geq 0$. We call a cone an *n-cone* in the event that the cone contains n linearly independent vectors. We will say that a matrix A is *row scalable* if there exists a diagonal matrix $D = \text{diag}(d_1, \dots, d_n) \in \mathbf{R}^{n \times n}$, with $d_i > 0$ for $1 \leq i \leq n$, so that $DADe = e$, i.e. DAD has row sums 1. In such a case we shall also say that D *row scales* A , or D is a *row scaling* for A . We will be primarily interested in the case when the matrix A to be row scaled is symmetric, and so shall drop off the adjective “row”. For $A \in \mathbf{R}^{n \times n}$, we will sometimes say that DAD is a *positive diagonal congruence* of A if DAD does not necessarily have row sums 1, but D is a diagonal matrix with positive diagonal entries. For convenience, we name both a vector and its components in lower case, such as $x = (x_1, \dots, x_n)^T \in \mathbf{R}^n$, and use upper case with the same letter for the diagonal matrix whose diagonal entries are the components of the vector in lower case. For example, we write $X = \text{diag}(x_1, \dots, x_n)$ for the diagonal matrix whose diagonal entries are the components of x . A matrix $A \in \mathbf{R}^{n \times n}$ with nonnegative entries is said to be *row stochastic* if $Ae = e$, and *doubly stochastic* if $Ae = e$ and $e^T A = e^T$. Obviously, a symmetric row stochastic matrix is doubly stochastic, and so we shall sometimes just describe such matrices as *stochastic*. A matrix $A \in \mathbf{R}^{n \times n}$ is said to be *irreducible* if there does not exist a permutation matrix $Q \in \mathbf{R}^{n \times n}$ so that QAQ^T has the form $QAQ^T = \begin{pmatrix} A_1 & 0 \\ A_2 & A_3 \end{pmatrix}$, where $A_1 \in \mathbf{R}^{k \times k}$, $A_2 \in \mathbf{R}^{l \times k}$, $A_3 \in \mathbf{R}^{l \times l}$, $k + l = n$, $k > 0$, $l > 0$. A nonnegative matrix is said to be *primitive* (or *aperiodic*) if A^m has all positive entries, for some integer $m \geq 1$. The primitivity of A depends only on the location of its zero entries [9], and so A is primitive if and only if $D_1 A D_2$ is primitive, where D_1 and D_2 are diagonal matrices with all diagonal entries positive.

Sinkhorn [32] showed that for a matrix $A \in \mathbf{R}^{n \times n}$ with all entries positive, there are unique (up to a scalar factor) diagonal matrices D_1 and D_2 , each with all diagonal entries positive, such that $D_1 A D_2$ is doubly stochastic. We shall refer to this result as Sinkhorn’s Theorem. His proof uses an iterative process, and some convexity arguments, to show that the process converges. His result was later proved in other ways by Menon [21], Djoković [7], London [18], Letac [17], Bapat [1], and Borobia and Cantó [2]. Sinkhorn also shows in [32] that if $A \in \mathbf{R}^{n \times n}$ is symmetric, with all entries positive, then $D_1 = D_2$.

A matrix $A \in \mathbf{R}^{n \times n}$ being *fully indecomposable* means that there do not exist permutation matrices P and Q so that $PAQ = \begin{pmatrix} A_1 & 0 \\ A_2 & A_3 \end{pmatrix}$, with A_1 and A_3 square and $A_2 \neq 0$. Brualdi, Parter and Schneider [4], and independently Sinkhorn and Knopp [34], showed that for a matrix $A \in \mathbf{R}^{n \times n}$ which is nonnegative and fully indecomposable, there are unique (up to a scalar factor) diagonal matrices D_1 and D_2 , each with positive diagonal entries such that $D_1 A D_2$ is doubly stochastic.

We wish to emphasize that we will be primarily concerned with the case when the matrix to be scaled is symmetric, and will use the word “scaling” when $D_1 = D_2$, although elsewhere “scaling” can refer to the case in which D_1 is not necessarily equal to D_2 . In Sinkhorn’s first paper on the subject [32], he credits Marcus and Newman with being the first to show that when a symmetric matrix A has positive entries then A can be scaled [24] (also referred to in [25]), although their proof has not appeared in the literature. Sinkhorn provided his own proof of this result for symmetric matrices in [33] (Sinkhorn’s proof in [32] does not preserve symmetry at each iteration). Maxfield and Minc claimed in [28] that A is scalable by an iterative algorithm when A is positive definite (so necessarily symmetric) with nonnegative entries, although their proof has also not appeared in the literature. Marshall and Olkin proved in [27] that when A is strictly copositive then A is scalable. Strict copositivity encompasses positive definiteness. Pereira [30] has extended Marshall and Olkin’s proof [27] on the existence of a scaling in the real positive definite case to positive definite matrices with complex entries. Sinkhorn’s Theorem has been extended to positive operators by Hobby and Pyke [8]. Sinkhorn’s Theorem has also been extended in a different direction by Sinkhorn [35], [36], for nonnegative matrices and by Brualdi [3] for symmetric nonnegative matrices, to achieve prescribed row and column sums not necessarily 1.

Brualdi, Parter and Schneider [4] showed that if A is a symmetric nonnegative matrix with positive diagonal entries then A is scalable. When A is a symmetric nonnegative matrix, Csima and Datta [5] showed that A is scalable if and only if there is a symmetric stochastic matrix with the same zero pattern as A . See [4] and [5] for other results similar to these.

Marshall and Olkin [27] proved that if A is strictly copositive then A is scalable. This result not only encompasses the case that A is positive definite, but also the case that A is nonnegative and has positive diagonal entries; since a matrix A being nonnegative with positive diagonal entries implies A is strictly copositive. This last statement is true since $x \geq 0$ and $0 = x^T Ax = \sum_{i=1}^n x_i (Ax)_i$ implies $x_i (Ax)_i = 0$ for each i . But if some $x_i \neq 0$ then $(Ax)_i = a_{i1}x_1 + \cdots + a_{ii}x_i + \cdots + a_{in}x_n = 0$, which is not possible since $a_{ii} > 0$. The converse is not true, since for example, the matrix $\begin{pmatrix} 1 & -1/2 \\ -1/2 & 1 \end{pmatrix}$ is strictly copositive.

The computational complexity of algorithms which perform the scaling, usually in polynomial time, is also of considerable interest because of its connections with linear programming. See for instance Linial, Samorodnitsky and Wigderson [19], Khachiyan and Kalantari [14], [15], Nemerovski and Rothblum [29], and Rothblum and Schneider [31].

The brief survey above serves only to provide background for our results. In Section 2, we give a simple iterative algorithm that row scales a matrix $A \in \mathbf{R}^{n \times n}$ with positive entries. We show in Section 3 that, depending on the matrix, we may have finitely many or infinitely many scalings, and we provide a condition for a unique scaling. In Section 4, we

show the equivalence of scalability to A being positive definite on a certain n -cone, which satisfies stated conditions. Since testing whether a matrix is copositive is NP-complete [23], we do not expect that deciding in general whether a matrix is scalable would be any easier. Section 5 discusses the scalability of positive semidefinite matrices, which was also considered by Marcus and Newman [25]. Finally, Section 6 provides an equivalent condition to strict copositivity, using scalability criteria.

2 A DAD scaling algorithm

Theorem 1 below provides another proof of the fact that a symmetric matrix $A \in \mathbf{R}^{n \times n}$ with positive entries can be scaled by an iterative process, where symmetry is preserved after each iteration. This proof is somewhat simpler than that given by Sinkhorn in [33]. In fact the proof works, as does Sinkhorn's, when A is not necessarily symmetric, so we shall not assume symmetry (either).

Theorem 1 *Let $A \in \mathbf{R}^{n \times n}$ have all positive entries. Then there is a diagonal matrix $D \in \mathbf{R}^{n \times n}$ with positive diagonal entries so that $DADe = e$.*

Proof Let $Ae = s = (s_1, \dots, s_n)^T$. Then we can write $(S^{-\frac{1}{2}}AS^{-\frac{1}{2}})S^{\frac{1}{2}}e = S^{\frac{1}{2}}e$, where $S = \text{diag}(s_1, \dots, s_n)$. Let $S^{-\frac{1}{2}}AS^{-\frac{1}{2}}e = t = (t_1, \dots, t_n)^T$. Then in the same way $(T^{-\frac{1}{2}}S^{-\frac{1}{2}}AS^{-\frac{1}{2}}T^{-\frac{1}{2}})T^{\frac{1}{2}}e = T^{\frac{1}{2}}e$, etc. Iterating like this constitutes the algorithm. Note that if A is symmetric then the matrices $S^{-\frac{1}{2}}AS^{-\frac{1}{2}}$, $T^{-\frac{1}{2}}S^{-\frac{1}{2}}AS^{-\frac{1}{2}}T^{-\frac{1}{2}}$, ..., at each iteration stay symmetric. We now prove the algorithm converges.

If A has equal row sums then $Ae = \lambda e = s$, and $S = \lambda I$, so that $S^{-\frac{1}{2}}AS^{-\frac{1}{2}} = \lambda^{-\frac{1}{2}}A\lambda^{-\frac{1}{2}}$ and $\lambda^{-\frac{1}{2}}A\lambda^{-\frac{1}{2}}\lambda^{\frac{1}{2}}e = \lambda^{\frac{1}{2}}e$, in which case $\lambda^{-\frac{1}{2}}A\lambda^{-\frac{1}{2}}e = e$, and we can take $D = \lambda^{-\frac{1}{2}}I$.

For the rest of the proof suppose that A does not have all row sums equal.

Note that

$$t_i = \frac{a_{i1}}{s_i^{\frac{1}{2}}s_1^{\frac{1}{2}}} + \frac{a_{i2}}{s_i^{\frac{1}{2}}s_2^{\frac{1}{2}}} + \dots + \frac{a_{in}}{s_i^{\frac{1}{2}}s_n^{\frac{1}{2}}} = \frac{1}{s_i^{\frac{1}{2}}} \left(\frac{a_{i1}}{s_1^{\frac{1}{2}}} + \frac{a_{i2}}{s_2^{\frac{1}{2}}} + \dots + \frac{a_{in}}{s_n^{\frac{1}{2}}} \right). \quad (1)$$

Then $\min_{1 \leq j \leq n} s_j^{\frac{1}{2}} < s_k^{\frac{1}{2}}$ for at least one k , so

$$t_i < \frac{1}{s_i^{\frac{1}{2}} \min_{1 \leq j \leq n} s_j^{\frac{1}{2}}} (a_{i1} + a_{i2} + \dots + a_{in}) = \frac{s_i^{\frac{1}{2}}}{\min_{1 \leq j \leq n} s_j^{\frac{1}{2}}}, \text{ for all } i, \quad (2)$$

and similarly

$$t_i > \frac{1}{s_i^{\frac{1}{2}} \max_{1 \leq j \leq n} s_j^{\frac{1}{2}}} (a_{i1} + a_{i2} + \dots + a_{in}) = \frac{s_i^{\frac{1}{2}}}{\max_{1 \leq j \leq n} s_j^{\frac{1}{2}}}, \text{ for all } i. \quad (3)$$

Thus we have

$$\frac{s_i^{\frac{1}{2}}}{\max_{1 \leq j \leq n} s_j^{\frac{1}{2}}} < t_i < \frac{s_i^{\frac{1}{2}}}{\min_{1 \leq j \leq n} s_j^{\frac{1}{2}}}, \text{ for all } i, 1 \leq i \leq n. \quad (4)$$

When i is such that we have a minimal s_i then $\frac{\min_{1 \leq j \leq n} s_j^{\frac{1}{2}}}{\max_{1 \leq j \leq n} s_j^{\frac{1}{2}}} < t_i < 1$, and when i is such that we have a maximal s_i then $1 < t_i < \frac{\max_{1 \leq j \leq n} s_j^{\frac{1}{2}}}{\min_{1 \leq j \leq n} s_j^{\frac{1}{2}}}$.

Also from inequality (4), when i is such that we have a minimal t_i then

$$\frac{s_i^{\frac{1}{2}}}{\max_{1 \leq j \leq n} s_j^{\frac{1}{2}}} < \min_{1 \leq j \leq n} t_j < \frac{s_i^{\frac{1}{2}}}{\min_{1 \leq j \leq n} s_j^{\frac{1}{2}}}, \quad (5)$$

so

$$\frac{\min_{1 \leq j \leq n} s_j^{\frac{1}{2}}}{\max_{1 \leq j \leq n} s_j^{\frac{1}{2}}} < \min_{1 \leq j \leq n} t_j, \text{ and similarly } \max_{1 \leq j \leq n} t_j < \frac{\max_{1 \leq j \leq n} s_j^{\frac{1}{2}}}{\min_{1 \leq j \leq n} s_j^{\frac{1}{2}}}, \quad (6)$$

and we can conclude that

$$\frac{\min_{1 \leq j \leq n} s_j^{\frac{1}{2}}}{\max_{1 \leq j \leq n} s_j^{\frac{1}{2}}} < \min_{1 \leq j \leq n} t_j < 1 < \max_{1 \leq j \leq n} t_j < \frac{\max_{1 \leq j \leq n} s_j^{\frac{1}{2}}}{\min_{1 \leq j \leq n} s_j^{\frac{1}{2}}}. \quad (7)$$

Multiplying the inequalities in (6) we have

$$\max_{1 \leq j \leq n} t_j \frac{\min_{1 \leq j \leq n} s_j^{\frac{1}{2}}}{\max_{1 \leq j \leq n} s_j^{\frac{1}{2}}} < \frac{\max_{1 \leq j \leq n} s_j^{\frac{1}{2}}}{\min_{1 \leq j \leq n} s_j^{\frac{1}{2}}} \min_{1 \leq j \leq n} t_j, \quad (8)$$

and cross-multiplying we have

$$\frac{\min_{1 \leq j \leq n} s_j}{\max_{1 \leq j \leq n} s_j} < \frac{\min_{1 \leq j \leq n} t_j}{\max_{1 \leq j \leq n} t_j}, \text{ and } \frac{\max_{1 \leq j \leq n} t_j}{\min_{1 \leq j \leq n} t_j} < \frac{\max_{1 \leq j \leq n} s_j}{\min_{1 \leq j \leq n} s_j}. \quad (9)$$

Consider now the sequence $(S_i)_{i=1}^{\infty}$, where $S_i = \text{diag}(s^{(i)})$, and $s^{(i)}$ denotes the vector of row sums at the i th iteration (so $S_1 = S$, and $S_2 = T$). The sequence $(S_i)_{i=1}^{\infty}$ generates the monotone strictly decreasing sequence of real numbers $\frac{\max_{1 \leq j \leq n} s_j^{(i)}}{\min_{1 \leq j \leq n} s_j^{(i)}}$. In terms of the spectral norm $\|\cdot\|$ and the condition number (see e.g. [16]) $\text{cond}(S_i) = \|S_i\| \|S_i^{-1}\| = \frac{\max_{1 \leq j \leq n} s_j^{(i)}}{\min_{1 \leq j \leq n} s_j^{(i)}}$,

we can see that the sequence $\text{cond}(S_i)$ is bounded below by 1. So the sequence $\text{cond}(S_i)$ is convergent by the Monotone Convergence Theorem to $\inf_i \text{cond}(S_i)$.

We now have an infinite sequence of diagonal matrices S_i , all of which have positive diagonal because from inequality (7) the minimal diagonal entry is bounded below, for $i \geq 2$, by $\text{cond}(S^{\frac{1}{2}})$. Since $(S_i)_{i=1}^{\infty}$ is bounded it has a convergent subsequence $(S_{i_k})_{k=1}^{\infty}$ and let us define $R = \text{diag}(r)$ to be the limit. Then $\frac{\max_{1 \leq j \leq n} r_j}{\min_{1 \leq j \leq n} r_j} = \inf_i \text{cond}(S_i)$, since the $\text{cond}(S_{i_k})$'s must have the same limit as the $\text{cond}(S_i)$'s.

Then $r = Re = DADe$ for some positive diagonal matrix D , where D is the product of all the matrices $S_i^{-\frac{1}{2}}$ in the sequence S_i (and not just in the sequence S_{i_k} , because of how the S_i 's are defined). Suppose that DAD has unequal row sums, so $\frac{\max_{1 \leq j \leq n} r_j}{\min_{1 \leq j \leq n} r_j} > 1$. Now performing another iteration of the algorithm, so that $(R^{-\frac{1}{2}}DADR^{-\frac{1}{2}})R^{\frac{1}{2}}e = R^{\frac{1}{2}}e$, we can define $q = Qe = (R^{-\frac{1}{2}}DADR^{-\frac{1}{2}})e$. Then from the rightmost inequality (9) we have $\frac{\max_{1 \leq j \leq n} q_j}{\min_{1 \leq j \leq n} q_j} < \frac{\max_{1 \leq j \leq n} r_j}{\min_{1 \leq j \leq n} r_j} = \inf_i \frac{\max_{1 \leq j \leq n} s_j^{(i)}}{\min_{1 \leq j \leq n} s_j^{(i)}}$, making $\frac{\max_{1 \leq j \leq n} q_j}{\min_{1 \leq j \leq n} q_j}$ smaller than the infimum of such ratios, which is not possible. So we must have $\frac{\max_{1 \leq j \leq n} r_j}{\min_{1 \leq j \leq n} r_j} = 1$, and $r = e$. ■

3 Counting scalings

When A is positive definite, or A has positive entries, the matrix D that scales A is unique (The former fact was proved in [27] and the latter in [32]). These facts are consequences of our next theorem.

Theorem 2 *Let $A \in \mathbf{R}^{n \times n}$. If A has two or more distinct scalings then A is positively diagonally congruent to a matrix that has both 1 and -1 as eigenvalues.*

Proof Let X and Y be two unequal diagonal matrices, with positive diagonal entries, so that $XAx = e$ and $YAy = e$, then

$$XYAx = y \quad \text{and} \quad XYAy = x. \quad (10)$$

Subtracting the two equations in (10) we have $XYA(x - y) = y - x$, so

$$(XY)^{\frac{1}{2}}A(XY)^{\frac{1}{2}}(XY)^{-\frac{1}{2}}(x - y) = -(XY)^{-\frac{1}{2}}(x - y), \quad (11)$$

while adding the two equations in (10) leads to

$$(XY)^{\frac{1}{2}}A(XY)^{\frac{1}{2}}(XY)^{-\frac{1}{2}}(x + y) = (XY)^{-\frac{1}{2}}(x + y). \quad (12)$$

But then $(XY)^{\frac{1}{2}}A(XY)^{\frac{1}{2}}$ is positively diagonally congruent to A , and has eigenvalues 1 and -1 , since $x > 0$, $y > 0$, so $x + y > 0$, and $x \neq y$. ■

Theorem 2 implies that there is a unique scaling for some broad subcategories of matrices, when A is scalable.

Corollary 3 *If $A \in \mathbf{R}^{n \times n}$ is positive definite, or positive semidefinite and scalable, or A is primitive (this includes the case of A having all positive entries), or A has nonnegative entries, is scalable, irreducible, and has the property that there does not exist a permutation matrix P so that $PAP^T = \begin{pmatrix} 0 & B \\ B^T & 0 \end{pmatrix}$, then A has a unique scaling.*

Proof If A is positive definite or positive semidefinite then clearly A cannot be positively diagonally congruent to a matrix with a negative eigenvalue. Then, from Theorem 2, the scaling of A is unique. If A has positive entries, then ZAZ still has positive entries (where Z is any diagonal matrix with positive diagonal entries), and can't have eigenvalues 1 and -1 from Perron's Theorem [9], as 1 would be the Perron root. If A is primitive, then ZAZ is primitive, and ZAZ can't have eigenvalues 1 and -1 , because if it did then $(ZAZ)^m$ would have eigenvalues 1 and -1 for m odd, and have all positive entries, which is not possible from Perron's Theorem. If A has nonnegative entries, is irreducible, and has the property that there does not exist a permutation matrix P so that $PAP^T = \begin{pmatrix} 0 & B \\ B^T & 0 \end{pmatrix}$, then ZAZ also has this property, and can't have eigenvalues 1 and -1 from the Perron-Frobenius Theorem, as 1 would be the Perron root, and PAP^T would have the Frobenius form for an irreducible matrix [22]. ■

Example To illustrate that we do not always have a unique scaling, when there exists a

scaling, consider the matrix $A = \begin{pmatrix} 1 & a & b \\ a & 1 & b \\ b & b & 1 \end{pmatrix}$. Then with $D = \text{diag}(d_1, d_2, d_3)$ we see that

for $DAD = \begin{pmatrix} d_1^2 & ad_1d_2 & bd_1d_3 \\ ad_1d_2 & d_2^2 & bd_2d_3 \\ bd_1d_3 & bd_2d_3 & d_3^2 \end{pmatrix}$ to have the first two row sums equal, we must have

$d_1^2 + bd_1d_3 = d_2^2 + bd_2d_3$ which can be rearranged as $(d_1 - d_2)(d_1 + d_2 + bd_3) = 0$. If $d_1 = d_2$ then dividing DAD by d_1^2 and taking $D = \text{diag}(1, 1, x)$ we must have $1 + a + bx = 2bx + x^2$

to have equal row sums, in which case $x^2 + bx - (a + 1) = 0$, so $x = \frac{-b \pm \sqrt{b^2 + 4(a+1)}}{2}$. Suppose

$b > 0$. If we are to have a scaling at all, we must have $b^2 + 4(a + 1) > 0$. So there is only one positive value for x , or none. Suppose $b < 0$. If we are to have positive row sums

we need $x(2b + x) > 0$, so $x > -2b$, which implies $a + 1 > 0$. So again we have just one positive value for x , or none. If $d_1 \neq d_2$ then $d_1 + d_2 + bd_3 = 0$. This is only possible

when $b < 0$. So suppose $b < 0$. We already have the first two row sums equal. The first (or second) row sum, after substituting $d_1 + d_2 + bd_3 = 0$ becomes $d_1d_2(a - 1) = 1$. So we must have $a - 1 > 0$. The third row summing to 1 means that $-b^2d_3^2 + d_3^2 = 1$. So we must have $1 - b^2 > 0$. Then $d_3 = \frac{1}{\sqrt{1-b^2}}$. We can eliminate d_2 and d_3 from $d_1 + d_2 + bd_3 = 0$

leaving us with the quadratic equation $d_1 + \frac{1}{d_1(a-1)} + \frac{b}{\sqrt{1-b^2}} = 0$. So, $d_1 = \frac{-b \pm \sqrt{b^2 - 4\frac{1-b^2}{a-1}}}{2\sqrt{1-b^2}}$,

$d_2 = \frac{1}{d_1(a-1)}$ and $d_3 = \frac{1}{\sqrt{1-b^2}}$. Where, again, to have a scaling we must have $b^2 - 4\frac{1-b^2}{a-1} \geq 0$, i.e. $a \geq \frac{4}{b^2} - 3$. So we obtain zero, one (when $a = \frac{4}{b^2} - 3$), or two possible positive values

for d_1 (the latter when $a > \frac{4}{b^2} - 3$ since $1 - b^2 > 0$). Thus our 3×3 matrix A may have zero, one, or two scalings, depending on the magnitudes of a and b .

Example It is also possible for a matrix to have infinitely many scalings. Consider the matrix $A = \begin{pmatrix} 0 & a \\ a & 0 \end{pmatrix}$, where $a > 0$. Let $D = \text{diag}(d_1, \frac{1}{ad_1})$. Then DAD is stochastic for any $d_1 > 0$. Of course any $n \times n$ matrix with this 2×2 matrix A as a direct summand, and the remaining block scalable, will also have infinitely many scalings.

4 Conditions for scalability

Theorem 4 provides four conditions on A that are almost, but not quite, equivalent to A being scalable. Thus, in order to show one direction of the theorem, namely that scalability implies the four conditions, we assume that A , when scalable, has only finitely many scalings.

Theorem 4 *Let $A \in \mathbf{R}^{n \times n}$ be symmetric. Consider the following four conditions:*

(i) *There exists some closed n -cone S , a nonempty subset of the nonnegative orthant, on which A is positive definite;*

(ii) *For the S in (i), define $T = S \cap \{x \mid Ax \geq 0, x^T Ax = n\}$. Then there is a $u > 0$ in the interior of T such that $Au > 0$ (in particular A is semipositive).*

(iii) *For any vector u , for which condition (ii) holds, we have for all x on the boundary of T that $u_1 \cdots u_n (Au)_1 \cdots (Au)_n > x_1 \cdots x_n (Ax)_1 \cdots (Ax)_n$.*

(iv) *For at least one vector x in the interior of T , at which a local maximum is achieved by $f(x_1, \dots, x_n) = x_1 (Ax)_1 \cdots x_n (Ax)_n$ on T , a local minimum is also achieved by the function $h(x_1, \dots, x_n) = \frac{1}{x_1 \cdots x_n} - (Ax)_1 \cdots (Ax)_n$ on T .*

If there is an S for which conditions (i), (ii), (iii) and (iv) hold, then A is scalable. Conversely, if A is scalable and has only finitely many scalings, then there is an S for which conditions (i), (ii), (iii) and (iv) hold.

Proof Before proving the theorem, we make a few comments about conditions (i)-(iv).

When showing that conditions (i)-(iv) imply scalability, we do not identify the set S other than to say that it is a closed n -cone S in the nonnegative orthant. However, for the converse, when A is scalable we will identify S in the course of the proof.

Conditions (ii) and (iv) mention the interior of $T \subset \mathbf{R}^n$. For any subset T of \mathbf{R}^n , a subset V of T is said to be open in T if $V = U \cap T$, for some subset U which is open in \mathbf{R}^n . By the interior of T we shall mean the union of all subsets of T which are open in T . Condition (iii) mentions the boundary of T . We will show that T is bounded.

Regarding condition (ii), note that semipositivity is a necessary condition for scalability. Condition (ii) states that the matrix A needs to be semipositive with only a certain type of vector u . Also, since $Au > 0$, for $u > 0$, and $u^T Au = n$, we see that A is positive

definite on an open subset U of \mathbf{R}^n , where $u \in U$. Moreover, if we write $u^T A u$ in the form $u^T A u = \sum_{i=1}^n \lambda_i v_i^2$, for $\lambda_i \in \mathbf{R}$, $1 \leq i \leq n$, and $v = (v_1, v_2, \dots, v_n)^T \in \mathbf{R}^n$, then we see that we can perturb u to other vectors u' close to u and (still) have $u'^T A u' = n$, and $A u' > 0$. Since A is positive definite on S from condition (i), and $u \in S$, A is positive definite on $U \cap S \cap \{x | Ax \geq 0, x^T Ax = n\} = U \cap T$, i.e. A is positive definite on the subset $U \cap T$, a set which is open in T .

Although condition (iii) does not state that a vector u , which satisfies a condition (ii), is a local maximum in T for the function f of condition (iv), it is a consequence of condition (iii) that if f does achieve a local maximum on T (in fact it must since we'll show that T is compact), then it will do so in the interior of T .

Finally, we prove the theorem. Suppose (i),(ii),(iii) and (iv) hold. Then $S \cap \{x | \|x\| = 1\}$ is compact, and $\min_{x \in S \text{ and } \|x\|=1} x^T A x = \alpha > 0$, since A is positive definite on S . This implies $y^T A y \geq \alpha y^T y$, for all $y \in S$. Now consider the function $f(x_1, \dots, x_n) = x_1 \cdots x_n (Ax)_1 \cdots (Ax)_n$ on the set $T = S \cap \{x | Ax \geq 0, x^T A x = n\}$. If $y \in T$ then $y^T A y = n$ and we must have $n \geq \alpha y^T y$, so T is bounded, and evidently closed. Taking a vector $u \in T$ that satisfies condition (ii), and then using (iii), $f(u_1, \dots, u_n)$ is strictly greater than $f(x_1, \dots, x_n)$, where x is on the boundary of T , so the maximum of f is achieved on the interior of T (Note that z in the interior of T implies that $z_i > 0$ and $(Az)_i > 0$ for all i). Let us use Lagrange's Multipliers to maximize f on T , that is, maximize f subject to $g(x_1, \dots, x_n) = x^T A x - n = x_1(Ax)_1 + \cdots + x_n(Ax)_n - n = 0$. Then setting $\nabla f = \lambda \nabla g$ we have

$$\begin{pmatrix} \frac{p}{x_1} & 0 & \cdots & \\ 0 & \frac{p}{x_2} & \cdots & \\ \vdots & \vdots & \ddots & \\ 0 & 0 & & \frac{p}{x_n} \end{pmatrix} e + A \begin{pmatrix} \frac{p}{(Ax)_1} & 0 & \cdots & \\ 0 & \frac{p}{(Ax)_2} & \cdots & \\ \vdots & \vdots & \ddots & \\ 0 & 0 & & \frac{p}{(Ax)_n} \end{pmatrix} e = 2\lambda A x, \quad (13)$$

with $p = x_1(Ax)_1 x_2(Ax)_2 \cdots x_n(Ax)_n$, where x is a point where the maximum is achieved. Now multiplying on the left of this equation by x^T we deduce that $\lambda = p$, then dividing across by p and multiplying across by $X = \text{diag}(x_1, \dots, x_n)$ we are left with the equation

$$e + X A \begin{pmatrix} \frac{1}{(Ax)_1} \\ \frac{1}{(Ax)_2} \\ \vdots \\ \frac{1}{(Ax)_n} \end{pmatrix} = 2X A x. \quad (14)$$

From (iv) we can again use Lagrange Multipliers, and find the minimum of h on T . We remark first that from the Arithmetic-Geometric Mean inequality since $x^T A x = x_1(Ax)_1 + \cdots + x_n(Ax)_n = n$ for $x \in T$, we know that $x_1(Ax)_1 \cdots x_n(Ax)_n \leq 1$, that is to

say $0 \leq h$ on the interior of T . Setting $\nabla h = \lambda \nabla g$ and multiplying across by X we are left with

$$\frac{-1}{(x_1 \cdots x_n)} e - (Ax)_1 (Ax)_2 \cdots (Ax)_n X A \begin{pmatrix} \frac{1}{(Ax)_1} \\ \frac{1}{(Ax)_2} \\ \vdots \\ \frac{1}{(Ax)_n} \end{pmatrix} = 2\lambda X A x, \quad (15)$$

and multiplying on the left by e^T we have $2\lambda = \frac{-1}{x_1 \cdots x_n} - (Ax)_1 \cdots (Ax)_n$. Substituting for λ we have

$$\frac{1}{x_1 \cdots x_n} e + (Ax)_1 \cdots (Ax)_n X A \begin{pmatrix} \frac{1}{(Ax)_1} \\ \frac{1}{(Ax)_2} \\ \vdots \\ \frac{1}{(Ax)_n} \end{pmatrix} = \left[\frac{1}{x_1 \cdots x_n} + (Ax)_1 \cdots (Ax)_n \right] X A x. \quad (16)$$

From (iv), since the maximum for f coincides with the minimum for h at some x in the interior of T we can substitute from (14) into (16) to obtain

$$\frac{1}{x_1 \cdots x_n} e + (Ax)_1 \cdots (Ax)_n [2X A x - e] = \left[\frac{1}{x_1 \cdots x_n} + (Ax)_1 \cdots (Ax)_n \right] X A x, \quad (17)$$

and after cancelling the $(Ax)_1 \cdots (Ax)_n X A x$ term from both sides we have

$$\left[\frac{1}{x_1 \cdots x_n} - (Ax)_1 \cdots (Ax)_n \right] e = \left[\frac{1}{x_1 \cdots x_n} - (Ax)_1 \cdots (Ax)_n \right] X A x. \quad (18)$$

Finally, if $\frac{1}{x_1 \cdots x_n} - (Ax)_1 \cdots (Ax)_n = 0$ then we have equality in the Arithmetic-Geometric Mean inequality, so $x_1 (Ax)_1 = \cdots = x_n (Ax)_n > 0$, which means we have scaled A . On the other hand, if $\frac{1}{x_1 \cdots x_n} - (Ax)_1 \cdots (Ax)_n \neq 0$, then $e = X A x$, which also means that we have scaled A , completing the proof of the first implication.

Conversely, suppose that A is scalable and A has only finitely many scalings. Then $X A x = e$, for some $x > 0$. The idea behind proving (i) is simple enough, namely, since $x^T A x = n$ then there must be a cone of vectors around x , in the nonnegative orthant and as narrow as is needed, on which A is positive definite. We present the details.

Let $\beta = \min_{(Ax)^T v = 0, \|v\|=1} v^T A v$. Note that for each $v \neq 0$ such that $(Ax)^T v = x^T (A v) = 0$, each v and each $A v$ must have at least one negative component and at least one positive component, since $x > 0$ and $A x > 0$. Let $\gamma_i = \min_{\|v\|=1} (A v)_i$, for $1 \leq i \leq n$, and $\gamma = \min_{1 \leq j \leq n} \gamma_j$. Then it follows from the preceding sentence that $\gamma < 0$.

Define S by $S = \{\lambda(x + tv) \mid (Ax)^T v = 0, \|v\| = 1, \lambda \geq 0\}$, where $0 \leq t \leq t_m$ and t_m is the maximum value that satisfies each of the following requirements: t_m is small enough that S only contains one scaling vector x ; $0 < t_m \leq \min_{1 \leq j \leq n} x_j$; $0 < t_m \leq \frac{1}{\gamma} \min_{1 \leq j \leq n} (\delta - \frac{1}{x_j})$ (with δ any real number satisfying $0 < \min_{1 \leq j \leq n} (\delta - \frac{1}{x_j})$, that is to say $\max_{1 \leq j \leq n} \frac{1}{x_j} < \delta$); if $\beta < 0$ we also require $0 < t_m \leq \sqrt{\frac{\epsilon - n}{\beta}}$ (with ϵ any real number satisfying $0 < \epsilon < n$).

In the definition of S , note that ϵ , δ , and the vector x are fixed, v is any unit vector such that $(Ax)^T v = 0$, and the assumption on t , about S only containing the one scaling vector x , is allowed since A has only finitely many scalings.

S is a subset of the nonnegative orthant. This is true because $\|v\| = 1$ which implies $1 \geq v_i \geq -1$, then $x_i + tv_i \geq 0$ is clearly true for any $v_i \geq 0$. Whereas if $0 > v_i \geq -1$, from the way we chose t_m such that $0 \leq t_m \leq \min_{1 \leq j \leq n} x_j$ we again have $x_i + tv_i \geq 0$. A is positive definite on S , since when $y = \lambda(x + tv) \in S$ we have $y^T Ay = [\lambda(x + tv)]^T A[\lambda(x + tv)] = \lambda^2[x^T Ax + t^2 v^T Av] \geq \lambda^2[n + t^2 \beta]$. If $\beta \geq 0$ then $y^T Ay > 0$, unless $\lambda = 0$. If $\beta < 0$ then $y^T Ay \geq \lambda^2[n + (\frac{\epsilon - n}{\beta})\beta] = \lambda^2 \epsilon > 0$, unless $\lambda = 0$.

S is a cone, since for $a = \lambda_1(x + t_1 v_1)$ and $b = \lambda_2(x + t_2 v_2)$ in S (with $\lambda_1 + \lambda_2 \neq 0$) we have $a + b = (\lambda_1 + \lambda_2)[x + \frac{1}{\lambda_1 + \lambda_2} t_1 \lambda_1 v_1 + \frac{1}{\lambda_1 + \lambda_2} t_2 \lambda_2 v_2]$, and $x \geq -t_1 v_1, x \geq -t_2 v_2$ implies $\frac{\lambda_1}{\lambda_1 + \lambda_2} x \geq -\frac{1}{\lambda_1 + \lambda_2} t_1 \lambda_1 v_1$ and $\frac{\lambda_2}{\lambda_1 + \lambda_2} x \geq -\frac{1}{\lambda_1 + \lambda_2} t_2 \lambda_2 v_2$, which imply $x \geq -(\frac{1}{\lambda_1 + \lambda_2} t_1 \lambda_1 v_1 + \frac{1}{\lambda_1 + \lambda_2} t_2 \lambda_2 v_2)$, so $a + b \in S$, and evidently any $\mu a + \nu b \in S$, for any $\mu, \nu \geq 0$. Also, if we write $\frac{1}{\lambda_1 + \lambda_2} t_1 \lambda_1 v_1 + \frac{1}{\lambda_1 + \lambda_2} t_2 \lambda_2 v_2 = sv$, where $\|v\| = 1$, then $s^2 = \frac{1}{(\lambda_1 + \lambda_2)^2} [t_1^2 \lambda_1^2 \|v_1\|^2 + 2t_1 t_2 \lambda_1 \lambda_2 v_1^T v_2 + t_2^2 \lambda_2^2 \|v_2\|^2] = \frac{1}{(\lambda_1 + \lambda_2)^2} [t_1^2 \lambda_1^2 + 2t_1 t_2 \lambda_1 \lambda_2 v_1^T v_2 + t_2^2 \lambda_2^2] \leq \frac{1}{(\lambda_1 + \lambda_2)^2} (t_1 \lambda_1 + t_2 \lambda_2)^2 \leq t_m^2$, where we used the fact that $v_1^T v_2 \leq \|v_1\| \|v_2\| = 1$. S is an n -cone since there are $n - 1$ linearly independent choices for v , and we also have $x \in S$. Since S is closed we now have condition (i).

Next, $Ax > 0$ so A is semipositive. x is in the interior of $T = S \cap \{y \mid Ay \geq 0, y^T Ay = n\}$, since $Ax > 0, x^T Ax = n$, and for every $v \neq 0$, such that $(Ax)^T v = 0$, we have for all $t \in [0, t_m]$ that $\lambda(x + tv) \in S$ and $\lambda(x - tv) \in S$, for any λ , and since also for each $y = \lambda(x + tv)$, we have $y^T Ay = n$ for an appropriate choice of $\lambda > 0$. Finally, we know that $Ay > 0$, since $Ay = A\lambda(x + tv) = \lambda(\frac{1}{x_1} + t(Av)_1, \dots, \frac{1}{x_n} + t(Av)_n)^T$, and $\frac{1}{x_i} + t(Av)_i \geq \frac{1}{x_i} + t\gamma \geq \delta > 0$, for each $i, 1 \leq i \leq n$. Hence, x serves as the vector u in condition (ii).

Now consider any $y \in T$. From the Arithmetic-Geometric Mean inequality we know that $y_1(Ay)_1 \cdots y_n(Ay)_n \leq 1$. Also, we know that we have equality in the Arithmetic-Geometric Mean inequality if and only if $y_1(Ay)_1 = \cdots = y_n(Ay)_n$, in which case y scales A . Since T only contains the one scaling vector x , we must have $y_1 \cdots y_n(Ay)_1 \cdots (Ay)_n < 1 = x_1 \cdots x_n(Ax)_1 \cdots (Ax)_n$ for all $y \in T$ where $y \neq x$, giving conditions (iii) and (iv). ■

Let A be the matrix of the example at the end of Section 3. Condition (iii) fails for A as $f(x_1, x_2) = x_1(Ax)_1 x_2(Ax)_2$ can be seen to be constant on T . So A scalable

(without the finiteness condition) will not imply the four conditions. We do not have an example of a matrix with infinitely many scalings that satisfies conditions (i)-(iv), and are currently unable to characterize when a symmetric matrix $A \in \mathbf{R}^{n \times n}$ has only finitely many scalings. It seems likely to us that, under the additional condition that all diagonal entries are positive, there are only finitely many scalings.

We note that the strongest, previously known, sufficient condition for scalability now follows easily from our theorem.

Corollary 5 (*Marshall-Olkin*) *Let $A \in \mathbf{R}^{n \times n}$ be strictly copositive. Then A is scalable.*

Proof In the event that A is strictly copositive then S may be chosen to be the nonnegative orthant, which provides condition (i) of the theorem. It is possible to show that A is semipositive by arguing as follows. Define $\alpha = \min_{\|x\|=1, x \geq 0} x^T A x > 0$, then $x^T A x \geq \alpha x^T x$, for all $x \geq 0$. If the minimum is achieved at $x_0 \geq 0$, so $x_0^T (A - \alpha I)x_0 = 0$, since $A - \alpha I$ is copositive it follows that $(A - \alpha I)x_0 \geq 0$ (for instance by using Lemma 1 in [12]) ensuring that $Ax_0 \geq 0$, and then use induction on n to demonstrate that A is semipositive. If $Au = v > 0$ for $u > 0$, then taking $w = \sqrt{\frac{n}{u^T v}}u > 0$, we have $w^T Aw = n$ and $Aw = \sqrt{\frac{n}{u^T v}}v > 0$, which is condition (ii). The statement that x is on the boundary of $S \cap \{x | Ax \geq 0, x^T Ax = n\}$ means that $x_i = 0$ or $(Ax)_i = 0$ for some i , so that $w_1 \cdots w_n (Aw)_1 \cdots (Aw)_n > 0$, which is condition (iii). Condition (iv) follows from the Arithmetic-Geometric Mean inequality, since $x^T Ax = n$ on T , as in the remark in the proof of the theorem. ■

Of course, Marshall and Olkin's result (their different proof may be found in [27]) implies that if A has all positive entries or if A is positive definite then A is scalable. Alfred Horn constructed an $n \times n$ copositive matrix (see [11]), which is not the sum of a nonnegative matrix and a positive semidefinite matrix. It is perhaps worth mentioning that Corollary 5 does not address scalability when it is only known that a given matrix A is copositive (and not strictly copositive). Alfred Horn's example is

$$A = \begin{pmatrix} 1 & -1 & 1 & 1 & -1 \\ -1 & 1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & 1 & -1 \\ -1 & 1 & 1 & -1 & 1 \end{pmatrix}, \quad (19)$$

which is copositive, not strictly copositive, neither nonnegative nor positive semidefinite, and is scalable (in fact is already scaled). An example of a copositive matrix that is not scalable can be found in the paragraph just after Lemma 6 (below).

5 Scalability and positive semidefiniteness

Marcus and Newman claimed in [25] (see also [24], [26] and [28]) that if a matrix A is positive semidefinite, has nonnegative entries, and doesn't have a zero row, then A is scalable (their proof has not appeared in the literature). Our next theorem determines when a positive semidefinite matrix is scalable. We will use the fact that for A positive semidefinite, $x^T Ax = 0$ implies $Ax = 0$. This is true since if $A = B^T B$ for $B \in \mathbf{R}^{n \times n}$, then $0 = x^T Ax = x^T B^T Bx = (Bx)^T Bx$, in which case $Bx = 0$ and $B^T Bx = Ax = 0$. We begin with a lemma which is one variation of the Theorem of the Alternative [20].

Lemma 6 *Let $A \in \mathbf{R}^{n \times n}$ be symmetric. Then it is not possible to have both statements I and II below.*

I A is semipositive

II There exists $w \geq 0$ ($w \neq 0$) such that $Aw \leq 0$

Proof Statement I means that $Ax = y > 0$ for some $x > 0$. If we also have statement II then $0 \geq w^T Ax = w^T y > 0$, which is not possible. ■

Evidently, if A is scalable we cannot have statement II of the lemma. We will find this fact to be useful while considering the scalability of positive semidefinite matrices. For instance, $A = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$ is not scalable, using the lemma with $w = e$.

Theorem 7 *Let $A \in \mathbf{R}^{n \times n}$ be positive semidefinite. Then the following are equivalent.*

(i) A is semipositive

(ii) A is strictly copositive

(iii) A does not have a nonnegative vector in its null space

(iv) A is scalable

Proof Suppose (i) is true. We know that $x^T Ax \geq 0$ for all $x \in \mathbf{R}^n$, since A is positive semidefinite, so certainly $x^T Ax \geq 0$ for all $x \geq 0$. If $x^T Ax = 0$ for some $x \geq 0$ ($x \neq 0$), then $Ax = 0$ (since A is positive semidefinite) but then from the lemma, with $w = x$, we have contradicted (i), proving (ii). Suppose now that (ii) is true, then A is scalable from Corollary 5, so (ii) implies (iv), and clearly (iv) implies (i). Also, (i) implies (iii) or else the null vector would serve as a w in the lemma. We show now that (iii) implies (i). (ii) is equivalent to (i), and we shall show that if (ii) does not hold then (iii) does not hold. If A is not strictly copositive then since A is positive semidefinite there must exist some $x \geq 0$ such that $x^T Ax = 0$ ($x \neq 0$). But then $Ax = 0$, which was what we wanted to prove. ■

Marcus and Newman's result, mentioned in the paragraph before Lemma 6, follows immediately now, since if A is positive semidefinite and nonnegative and doesn't have a zero row then $Ae > 0$, so A is semipositive. Our next two corollaries follow from Corollary 5 and Theorem 7. Corollary 8 implies that if A is scalable then A^2 is scalable.

Corollary 8 *Let $A \in \mathbf{R}^{n \times n}$ be symmetric. If A is semipositive then A^2 is scalable.*

Proof Suppose A is semipositive. If A is nonsingular then A^2 is positive definite, then Corollary 5 implies that A^2 is scalable. If A is singular then A^2 is positive semidefinite. If A^2 is not scalable then from the theorem A^2 has a nonnegative null vector, so $A^2x = 0$ for some $x \geq 0$. But then $x^T A^2 x = (Ax)^T (Ax) = 0$, so $Ax = 0$, but then x would serve as w in the lemma. ■

It is not necessarily true that if A and B are symmetric and scalable then AB is scalable. For example when $A = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 3 & 2 \\ 2 & 1 \end{pmatrix}$ then $AB = \begin{pmatrix} 4 & 3 \\ -1 & -1 \end{pmatrix}$, then if $AB \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = \begin{pmatrix} 1/d_1 \\ 1/d_2 \end{pmatrix}$, this contradicts both d_i 's being positive. If we demand that AB is symmetric, then with $A = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}$ and $B = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$ we have $AB = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$. But AB is not scalable, since, with $D = \text{diag}(d_1, d_2)$, if $DABD$ were stochastic we would have $d_1 d_2 = 1$ and $d_1 d_2 + d_2^2 = 1$, which is not possible. We do however have the following result, which relies on AB being positive semidefinite (so AB must necessarily be symmetric), so that this corollary does not say much more than Theorem 7. This requires that A commutes with B , since AB is symmetric iff $(AB)^T = AB$ iff $B^T A^T = AB$ iff $BA = AB$.

Corollary 9 *Let both $A, B \in \mathbf{R}^{n \times n}$ be both symmetric and semipositive, with B having nonnegative entries, and AB positive semidefinite. Then AB is scalable.*

Proof If AB is nonsingular then AB is positive definite, so AB is scalable. Suppose AB is singular. If AB is not scalable then AB has a nonnegative null vector w , from the theorem, so $ABw = 0$. We can't have $Bw = 0$ ($w \neq 0$) from the lemma, since B is semipositive. But if A has the nonnegative null vector Bw , this contradicts A being semipositive, since from the lemma A cannot be both semipositive and have $A(Bw) = 0$. ■

6 Total scalability and strict copositivity

Equivalent conditions for a matrix to be copositive or strictly copositive have appeared already in the literature. See [10] for a survey. Cottle, Pang and Stone [6] proved Theorem 10.

Theorem 10 *Let $A \in \mathbf{R}^{n \times n}$ be symmetric. Then A is copositive if and only if every $r \times r$ principal submatrix B of A has a vector $x \in \mathbf{R}^r$ such that $x \geq 0$ and $Bx \geq 0$, for $1 \leq r \leq n$.*

Kaplan [13] proved Theorem 11, which we will use to prove Theorem 12.

Theorem 11 *Let $A \in \mathbf{R}^{n \times n}$ be symmetric. Then A is strictly copositive if and only if every $r \times r$ principal submatrix B of A does not have an eigenvector $v \in \mathbf{R}^r$ such that $v > 0$ with corresponding eigenvalue $\lambda \leq 0$, for $1 \leq r \leq n$.*

We say that a symmetric matrix $A \in \mathbf{R}^{n \times n}$ is totally scalable if every $r \times r$ principal submatrix B of A is scalable, for $1 \leq r \leq n$. A matrix being totally scalable extends the list of conditions equivalent to strict copositivity, as we now show.

Theorem 12 *Let $A \in \mathbf{R}^{n \times n}$ be symmetric. Then the following are equivalent.*

- (i) *A is totally scalable*
- (ii) *Every principal submatrix B of A is semipositive*
- (iii) *No principal submatrix B of A has an eigenvector $v \in \mathbf{R}^r$ such that $v > 0$ with corresponding eigenvalue $\lambda \leq 0$*
- (iv) *A is strictly copositive*

Proof Suppose that (i) holds. Every principal submatrix of A being scalable implies in particular that every principal submatrix B is semipositive, which is statement (ii). But then, no principal submatrix can have an eigenvector $v \in \mathbf{R}^r$ such that $v > 0$ with corresponding eigenvalue $\lambda \leq 0$, because of Lemma 6, which proves (iii). Theorem 11 then implies (iv). A being strictly copositive implies that every principal submatrix B is strictly copositive, so every B is scalable from Corollary 5, so (i) holds. ■

References

- [1] R. Bapat, D_1AD_2 theorems for multidimensional matrices, *Linear Algebra and its Applications*, **48** (1982) 437–442.
- [2] A. Borobia and R. Cantó, Matrix scaling: a geometric proof of Sinkhorn’s theorem, *Linear Algebra and its Applications*, **268** (1998) 1–8.
- [3] R. Brualdi, The DAD theorem for arbitrary row sums, *Proc. Amer. Math. Soc.* **45**(2) (1974) 189–194.
- [4] R. A. Brualdi, S. V. Parter and H. Schneider, The diagonal equivalence of a non-negative matrix to a stochastic matrix, *Journal of Math. Anal. and Appl.* **16** (1966) 31–50.
- [5] J. Csima and B. N. Datta, The DAD theorem for symmetric non-negative matrices, *Journal of Combinatorial Theory* **12** (1972) 147–152.
- [6] R. W. Cottle, Jong-Shi Pang and R. E. Stone, *The linear complementarity problem*, Academic Press, Boston (1992).

- [7] D. Ž. Djoković, Note on nonnegative matrices, *Proc. Amer. Math. Soc.* **25** (1970) 80–82.
- [8] C. Hobby and R. Pyke, Doubly stochastic operators from positive operators, *Pacific J. Math.* **15** (1965) 153–157.
- [9] R. A. Horn and C. R. Johnson, *Matrix analysis*, Cambridge University Press (1985).
- [10] Kh. D. Ikramov and N. V. Savel’eva, Conditionally definite matrices, *Journal of Mathematical Sciences* **98**(1) (2000) 1–50.
- [11] M. Hall Jr., *Combinatorial theory*, John Wiley and Sons, New York (1986).
- [12] C. R. Johnson and R. Reams, Spectral theory of copositive matrices, *Linear Algebra and its Applications* **395** (2005) 275–281.
- [13] W. Kaplan, A test for copositive matrices, *Linear Algebra and its Applications* **313** (2000) 203–206.
- [14] L. Khachiyan and B. Kalantari, Diagonal matrix scaling and linear programming, *SIAM Journal on Optimization* **2** (1992) 668–672.
- [15] L. Khachiyan and B. Kalantari, On the complexity of nonnegative-matrix scaling *Linear Algebra and its Applications* **240** (1996) 87–103.
- [16] P. Lancaster and M. Tismenetsky, *The Theory of Matrices, with applications* 2nd Ed. Academic Press, San Diego (1985).
- [17] G. Letac, A unified treatment of some theorems on positive matrices, *Proc. Amer. Math. Soc.* **43** (1974) 11–17.
- [18] D. London, On matrices with a doubly stochastic pattern, *Journal of Math. Anal. and Appl.* **34** (1971) 648–652.
- [19] N. Linial, A. Samorodnitsky and A. Wigderson, A deterministic strongly polynomial algorithm for matrix scaling and approximate permanents, *Combinatorica* **20** (2000) 545–568.
- [20] O. Mangasarian, *Nonlinear programming*, SIAM Classics in applied math, Vol. 10 (1994).
- [21] M. V. Menon, Reduction of a matrix with positive elements to a doubly stochastic matrix, *Proc. Amer. Math. Soc.* **18** (1967) 244–247.
- [22] H. Minc, *Nonnegative matrices*, John Wiley and Sons, New York (1988).

- [23] K. Murty and S. N. Kabadi, Some NP-complete problems in quadratic and nonlinear programming, *Mathematical Programming* **39** (1987) 117–129.
- [24] M. Marcus and M. Newman, The permanent of a symmetric matrix, Abstract 587–85, *Notices Amer. Math. Soc.* **8** (1961) 595.
- [25] M. Marcus and M. Newman, Generalized functions of symmetric matrices, *Proc. Amer. Math. Soc.* **16** (1965) 826–830.
- [26] M. Marcus and M. Newman, *A survey of matrix theory and matrix inequalities*, Dover (1992).
- [27] A. W. Marshall and I. Olkin, Scaling of matrices to achieve specified row and column sums, *Numerische Mathematik* **12** (1968) 83–90.
- [28] J. Maxfield and H. Minc, A doubly stochastic matrix equivalent to a given matrix, Abstract 592–81, *Notices Amer. Math. Soc.* **9** (1962) 309.
- [29] A. Nemirovski and U. Rothblum, On complexity of matrix scaling, *Linear Algebra and its Applications* **302/303** (1999) 435–460.
- [30] R. Pereira, Differentiators and the geometry of polynomials, *Journal of Math. Anal. and Appl.* **285**(1) (2003) 336–348.
- [31] U. Rothblum and H. Schneider, Scalings of matrices which have prespecified row sums and column sums via optimization, *Linear Algebra and its Applications* **114/115** (1989) 737–764.
- [32] R. Sinkhorn, A relationship between arbitrary positive matrices and doubly stochastic matrices, *Ann. Math. Statist.* **35** (1964) 876–879.
- [33] R. Sinkhorn, A relationship between arbitrary positive matrices and stochastic matrices, *Canadian J. Math.* **13** (1966) 303–306.
- [34] R. Sinkhorn and P. Knopp, Concerning nonnegative matrices and doubly stochastic matrices, *Pacific Journal of Mathematics* **21**(2) (1967) 343–348.
- [35] R. Sinkhorn, Diagonal equivalence to matrices with prescribed row and column sums, *American Mathematical Monthly* **74** (1967) 402–405.
- [36] R. Sinkhorn, Diagonal equivalence to matrices with prescribed row and column sums II, *Proc. Amer. Math. Soc.* **45**(2) (1974) 195–198.