Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Linear Algebra and its Applications 392 (2004) 1–9
LINEAR ALGEBRA AND ITS APPLICATIONS
www.elsevier.com/locate/laa

Sums of totally positive matrices

C.R. Johnson [a], D.D. Olesky [b,*,1]

[a] *Department of Mathematics, College of William and Mary, P.O. Box 8795, Williamsburg, VA 23187-8795, USA*
[b] *Department of Computer Science, University of Victoria, P.O. Box 3055, Victoria, B.C., Canada V8W 3P6*

Received 13 October 2003; accepted 19 May 2004

Submitted by W. Barrett

Abstract

It is shown that an arbitrary $m \times n$ positive matrix can be written as a sum of at most $\min\{m, n\}$ totally positive matrices, and that this is in general the best possible value for the number of summands. Sufficient conditions are given under which fewer than $\min\{m, n\}$ totally positive summands are needed.

Keywords: Matrix sum; Positive matrix; Totally positive matrix

1. Introduction

Matrix factorizations are an important part of linear algebra. There are also well-known results regarding decomposition of a matrix into a sum of matrices from a specified class. For example, any symmetric matrix is the difference of two positive definite matrices, and any positive matrix is the sum of two P-matrices (i.e., matrices with all principal minors positive).

We show that an arbitrary $m \times n$ positive matrix can be written as a sum of at most $\min\{m, n\}$ totally positive matrices, and an example is given to show that this

* Corresponding author. Tel.: +1 250 721 7223; fax: +1 250 721 7292.
E-mail addresses: crjohnso@math.wm.edu (C.R. Johnson), dolesky@cs.uvic.ca (D.D. Olesky).
1 Research supported by an NSERC Discovery grant.

0024-3795/$ - see front matter
doi:10.1016/j.laa.2004.05.008

is the best possible value for the number of summands. We further give sufficient conditions under which fewer (than $\min\{m, n\}$) totally positive summands are needed.

2. Totally nonnegative and totally positive matrices

Properties and applications of the following classes of matrices can be found, for example, in [1,2,4].

Definition 1. An $m \times n$ matrix A is *totally nonnegative* (*totally positive*) if **all** of its minors are $\geqslant 0$ (>0).

We denote these two matrix classes by TN and TP, respectively.

Clearly any entrywise nonnegative matrix (written $A \geqslant 0$) with all of its nonzero entries in one row (or one column) is a TN matrix. Thus, by partitioning any $m \times n$ $A \geqslant 0$ into its row (or column) vectors, A can be written as a sum of m (or n) TN matrices with each summand containing one row (or column) of A, and all other entries equal to 0.

The following example illustrates that for every $n \geqslant 2$, there exists an $n \times n$ nonnegative matrix that cannot be written as a sum of fewer than n TN matrices.

Example 1. Let $n \geqslant 2$ and consider the $n \times n$ "backward identity" matrix

$$K = \begin{bmatrix} & & & & 1 \\ & 0 & & 1 & \\ & & \iddots & & \\ & 1 & & 0 & \\ 1 & & & & \end{bmatrix}.$$

Suppose that $K = \sum_{k=1}^{n-1} B_k$, where $B_k = \left[b_{ij}^{(k)}\right] \geqslant 0$. Since there are n positive entries in K, there must be two positive entries in at least one matrix B_k, say B_p. Since in every position in which K is 0, each B_k must have a 0, it follows that the 2×2 minor containing these two positive entries is negative and B_p is not a TN matrix. That is, K cannot be written as a sum of $n - 1$ TN matrices. (Note that this example can be extended to show that there exist $m \times n$ nonnegative matrices that cannot be written as a sum of $\min\{m, n\}$ TN matrices by bordering K with either rows or columns of 0's.)

Since every entry of a TP matrix is positive, any sum of TP matrices is a positive matrix. In the remainder of this paper, we address the question of whether or not an arbitrary positive matrix can be written as a sum of TP matrices, and if so, the number of such matrices that is required.

3. A positive matrix as a sum of TP matrices

Every row or column vector of an $m \times n$ entrywise positive matrix A (written $A > 0$) is a $1 \times n$ or $m \times 1$ TP submatrix, respectively. In Lemma 3 below, more general rectangular TP submatrices of an $m \times n$ matrix are considered, and it is shown that any such submatrix can be perturbed (leaving all of the entries of the given TP submatrix unchanged) so as to obtain an $m \times n$ TP matrix.

We first prove a lemma that shows that a TP matrix can be bordered by a row or column vector so that it remains a TP matrix. The second lemma below gives a known result, namely that a row (or column) vector can be inserted between any two rows (or columns) of a TP matrix so that it remains a TP matrix.

Lemma 1. *If $A > 0$ is an $m \times n$ TP matrix, then there exist positive row vectors w and x with n entries and positive column vectors y and z with m entries so that the bordered $(m + 1) \times n$ matrices*

$$\left[\frac{A}{w}\right] \quad \text{and} \quad \left[\frac{x}{A}\right]$$

and the bordered $m \times (n + 1)$ matrices

$$[A|y] \quad \text{and} \quad [z|A]$$

are TP *matrices.*

Proof. We show the existence of w, the other cases being similar. The first entry w_1 of w can be chosen to be an arbitrary positive number. Then, in turn, positive values for each of $w_2, w_3, \ldots, w_n$ can be determined so that every minor that is completely specified when these values are determined is positive. Note that each of these values enters into all such determinant computations positively, so that all of the relevant minors will be positive if each value (in turn) is chosen to be sufficiently large.

The vectors x, y and z can be determined similarly, with the values specified in the order x_n (arbitrary), $x_{n-1}, \ldots, x_1$; y_1 (arbitrary), $y_2, \ldots, y_m$; and z_m (arbitrary), $z_{m-1}, \ldots, z_1$. □

Lemma 2 [3, Theorem 2.3]. *Let $A > 0$ be an $m \times n$ TP matrix. Then there exists a positive row vector w with n entries (or a positive column vector y with m entries) so that the $(m + 1) \times n$ matrix (or the $m \times (n + 1)$ matrix) obtained by inserting w (or y) between any pair of specified rows (or columns) of A is a* TP *matrix.*

If α, β are nonempty subsets of distinct positive integers, then $A[\alpha, \beta]$ denotes the submatrix of A with entries from rows specified by α and columns specified by β. For a given $m \times n$ matrix A, let the support of A, denoted $\mathrm{supp}(A)$, be $\{(i, j) : a_{ij} \neq 0\}$. In the following, $\|\cdot\|$ denotes an arbitrary matrix norm.

Lemma 3. *Given* $\alpha \subseteq \{1, 2, \ldots, m\}$ *and* $\beta \subseteq \{1, 2, \ldots, n\}$, *suppose that* $A \geqslant 0$ *is an* $m \times n$ *matrix such that* $A[\alpha, \beta]$ *is a* TP *matrix, and all other* $a_{ij} = 0$. *Then for all* $\varepsilon > 0$, *there exists an* $m \times n$ TP *matrix* $\hat{A}$ *with* $\hat{A}[\alpha, \beta] = A[\alpha, \beta]$ *and* $\|A - \hat{A}\| < \varepsilon$.

Proof. Suppose that $\alpha = \{\alpha_1, \alpha_2, \ldots, \alpha_p\}$ and $\beta = \{\beta_1, \beta_2, \ldots, \beta_q\}$ in which $1 \leqslant p \leqslant m$, $1 \leqslant q \leqslant n$, $\alpha_i < \alpha_{i+1}$ and $\beta_j < \beta_{j+1}$ for all $i \leqslant p - 1$ and $j \leqslant q - 1$. If at least one of the index sets α, β is not a set of consecutive integers, then Lemma 2 can be used (perhaps repeatedly) to obtain a TP matrix B of order $(\alpha_p - \alpha_1 + 1) \times (\beta_q - \beta_1 + 1)$. Now, with $\tilde{\alpha} = \{\alpha_1, \alpha_1 + 1, \alpha_1 + 2, \ldots, \alpha_p\}$ and $\tilde{\beta} = \{\beta_1, \beta_1 + 1, \beta_1 + 2, \ldots, \beta_q\}$, let $\tilde{A} \geqslant 0$ be the $m \times n$ matrix with $\tilde{A}[\tilde{\alpha}, \tilde{\beta}] = B$, and all other $\tilde{a}_{ij} = 0$. Then $\tilde{A}$ is a TN matrix in which $\tilde{A}[\alpha, \beta] = A[\alpha, \beta]$ and $\tilde{A}[\tilde{\alpha}, \tilde{\beta}]$ are TP submatrices. (If both α and β are sets of consecutive integers, then $\tilde{\alpha} = \alpha$, $\tilde{\beta} = \beta$, $\tilde{A} = A$ and Lemma 2 is not required.)

Next, Lemma 1 can be used to border the rows and/or columns of $\tilde{A}[\tilde{\alpha}, \tilde{\beta}]$ so as to obtain an $m \times n$ TP matrix, say $\bar{A}$, such that $\bar{A}[\tilde{\alpha}, \tilde{\beta}] = \tilde{A}[\tilde{\alpha}, \tilde{\beta}]$ (and thus $\bar{A}[\alpha, \beta] = A[\alpha, \beta]$).

The matrix $\hat{A}$ is obtained from $\bar{A}$ by scaling the rows and columns outside of the index sets α and β, respectively, by positive diagonal matrices. Let $\hat{A} = D_1 \bar{A} D_2$, where $D_1 \equiv \left[d_{ij}^{(1)}\right]$ and $D_2 \equiv \left[d_{ij}^{(2)}\right]$ are diagonal matrices in which, for $1 \leqslant i \leqslant m$ and $1 \leqslant j \leqslant n$,

$$d_{ii}^{(1)} = \begin{cases} 1, & \text{if } i \in \alpha, \\ \hat{\varepsilon}, & \text{otherwise} \end{cases}$$

and

$$d_{jj}^{(2)} = \begin{cases} 1, & \text{if } j \in \beta, \\ \hat{\varepsilon}, & \text{otherwise.} \end{cases}$$

Then, for any given ε and sufficiently small $\hat{\varepsilon}$, $\|A - \hat{A}\| < \varepsilon$. □

In order to write a positive matrix A as a sum of TP matrices, we require a partitioning of the entries of A into disjoint rectangular TP submatrices. As stated at the beginning of this section, the row (or column) vectors of A always give such a partition. The following definition generalizes this concept.

Definition 2. A partition of an $m \times n$ matrix $A > 0$ into $k \geqslant 1$ rectangular TP submatrices $A[\alpha_r, \beta_r]$, for $1 \leqslant r \leqslant k$, such that every position (i, j) for $1 \leqslant i \leqslant m$ and $1 \leqslant j \leqslant n$ lies in one and only one block of the partition, is called a TP *k-partition* of A.

We now state our main result.

Theorem 1. *If an* $m \times n$ *matrix* $A > 0$ *has a* TP *k-partition*, *then* A *can be written as a sum of* k TP *matrices*.

Proof. Write A as a sum of k $m \times n$ TN matrices

$$A = A_1 + A_2 + \cdots + A_k,$$

where, for $1 \leqslant p \leqslant k$, $A_p \equiv \left[a_{ij}^{(p)}\right]$ and

$$a_{ij}^{(p)} = \begin{cases} a_{ij}, & \text{if } i \in \alpha_p \text{ and } j \in \beta_p, \\ 0, & \text{otherwise.} \end{cases}$$

By Lemma 3, there exists an $m \times n$ matrix $E_1 \geqslant 0$ with $\|E_1\|$ arbitrarily small so that $A_1 + E_1 \equiv A_1^{(1)}$ is a TP matrix. Since any sufficiently small perturbation of a TP matrix is a TP matrix, E_1 can be chosen sufficiently small so that if

$$A = A_1^{(1)} + A_2^{(1)} + A_3^{(1)} + \cdots + A_k^{(1)}$$

in which, for $2 \leqslant p \leqslant k$, $\operatorname{supp}\left(A_p^{(1)}\right) = \operatorname{supp}(A_p)$, then $A_p^{(1)}[\alpha_p, \beta_p]$ is still a TP matrix.

Now similarly, by Lemma 3, there exists $E_2 \geqslant 0$ with $\|E_2\|$ arbitrarily small so that $A_2^{(1)} + E_2 \equiv A_2^{(2)}$ is a TP matrix. Moreover, E_2 can be chosen sufficiently small so that if

$$A = A_1^{(2)} + A_2^{(2)} + A_3^{(2)} + \cdots + A_k^{(2)}$$

in which $\operatorname{supp}\left(A_p^{(2)}\right) = \operatorname{supp}\left(A_p^{(1)}\right)$ for $p \neq 2$, then $A_1^{(2)}$ is still a TP matrix and $A_p^{(2)}[\alpha_p, \beta_p]$ is still a TP matrix for all $p \geqslant 3$.

Continuing in this manner, k such steps complete the proof, by, in turn, making each of the k summands a (full) TP matrix. □

Since both the set of row vectors and the set of column vectors of a positive matrix are a TP partition, we have the following.

Corollary 1. *Any $m \times n$ matrix $A > 0$ is a sum of at most* $\min\{m, n\}$ TP *matrices.*

The next result follows since any real matrix can be written as the difference of two positive matrices.

Corollary 2. *An arbitrary $m \times n$ real matrix is a sum and difference of at most* 2 $\min\{m, n\}$ TP *matrices.*

The following example illustrates Theorem 1.

Example 2. Let

$$A = \begin{bmatrix} 1 & 3 & 2 \\ 2 & 1 & 1 \\ 4 & 2 & 1 \end{bmatrix}$$

and consider the TP 3-partition of A into its row vectors. Then, for example,

$$A_1 + E_1 = \begin{bmatrix} 1 & 3 & 2 \\ 0.001 & 0.004 & 0.006 \\ 0.001 & 0.005 & 0.02 \end{bmatrix} \equiv A_1^{(1)}$$

is a TP matrix, and $E_1 \geqslant 0$ is sufficiently small so that $A = A_1^{(1)} + A_2^{(1)} + A_3^{(1)}$ with

$$A_2^{(1)} = \begin{bmatrix} 0 & 0 & 0 \\ 1.999 & 0.996 & 0.994 \\ 0 & 0 & 0 \end{bmatrix}, \quad A_3^{(1)} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 3.999 & 1.995 & 0.98 \end{bmatrix},$$

and $A_2^{(1)}[\{2\}, \{1, 2, 3\}]$ and $A_3^{(1)}[\{3\}, \{1, 2, 3\}]$ are TP matrices.

Next, for example,

$$A_2^{(1)} + E_2 = \begin{bmatrix} 0.005 & 0.002 & 0.001 \\ 1.999 & 0.996 & 0.994 \\ 0.001 & 0.006 & 0.02 \end{bmatrix} \equiv A_2^{(2)}$$

is a TP matrix, and $E_2 \geqslant 0$ is sufficiently small so that $A = A_1^{(2)} + A_2^{(2)} + A_3^{(2)}$ with

$$A_1^{(2)} = \begin{bmatrix} 0.995 & 2.998 & 1.999 \\ 0.001 & 0.004 & 0.006 \\ 0.001 & 0.005 & 0.02 \end{bmatrix},$$

$$A_3^{(2)} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 3.998 & 1.989 & 0.96 \end{bmatrix},$$

and $A_1^{(2)}$ and $A_3^{(2)}[\{3\}, \{1, 2, 3\}]$ are TP matrices.

After one more such step, we obtain, for example,

$$A = \begin{bmatrix} 0.993 & 2.9976 & 1.9989 \\ 0.001 & 0.004 & 0.006 \\ 0.001 & 0.005 & 0.02 \end{bmatrix} + \begin{bmatrix} 0.005 & 0.002 & 0.001 \\ 1.992 & 0.993 & 0.993 \\ 0.001 & 0.006 & 0.02 \end{bmatrix} + \begin{bmatrix} 0.002 & 0.0004 & 0.0001 \\ 0.007 & 0.003 & 0.001 \\ 3.998 & 1.989 & 0.96 \end{bmatrix},$$

which is the sum of three TP matrices.

4. Sharpness of our main result

We first give an example to show that the converse of Theorem 1 is not, in general, true.

Example 3. The matrix

$$A = \begin{bmatrix} 2 & 4.1 & 5.3 \\ 2.7321 & 5.5651 & 6.9848 \\ 4 & 8.1 & 10.01 \end{bmatrix}$$

is the sum of the $k = 2$ TP matrices

$$\begin{bmatrix} 1 & 1 & 1 \\ 1.7321 & 2 & 2.025 \\ 3 & 4 & 4.01 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 3.1 & 4.3 \\ 1 & 3.5651 & 4.9598 \\ 1 & 4.1 & 6 \end{bmatrix},$$

but A does not have a TP 2-partition since all minors of A of orders 2 and 3 are negative.

The following example shows that Corollary 1 is sharp in that for all $m, n \geqslant 2$, there exists an $m \times n$ matrix $A > 0$ that cannot be written as a sum of fewer than $\min\{m, n\}$ TP matrices.

This example, in which the matrix K of Example 1 is modified to be positive, is stated for the case $m = n$, but is easily extended to the rectangular case.

Example 4. Let $n \geqslant 2$, $t > 1$ and let $A > 0$ be an $n \times n$ matrix with

$$a_{ij} = \begin{cases} tn, & \text{if } j = n - i + 1, \\ \varepsilon, & \text{otherwise.} \end{cases}$$

If $A = \sum_{k=1}^{n-1} B_k$, where $B_k = \left[b_{ij}^{(k)}\right] > 0$, then since there are n entries in A equal to tn and only $n - 1$ summands B_k, in at least one of these matrices, say B_p, there are two distinct entries $b_{i,n-i+1}^{(p)}$ whose product is greater than 1 (for some fixed t sufficiently large). Thus, for any $\varepsilon < 1$, the 2×2 minor containing these two entries $b_{i,n-i+1}^{(p)}$ is negative, and B_p is not a TP matrix.

5. Concluding remarks

Example 3 shows that the sum of two 3×3 TP matrices need not have a positive minor of order $\geqslant 2$. However, Example 4 shows that something more than just positivity is required for an $m \times n$ matrix $A > 0$ to be a sum of fewer than $\min\{m, n\}$ TP matrices.

Question 1. For any fixed k $(2 \leqslant k < \min\{m, n\})$, what characterizes the sum of k TP matrices?

The case $k = 2$ seems to be especially interesting, as the following examples illustrate.

Example 5. Let

$$A = \begin{bmatrix} 1 & 10 & 1000 \\ 10 & 1 & 10 \\ 1000 & 10 & 1 \end{bmatrix}.$$

If $A = B + C$ is the sum of two TP matrices, with $B \equiv [b_{ij}]$ and $C \equiv [c_{ij}]$, then

$$\begin{pmatrix} 1 & 10 \\ 10 & 1 \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} + \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} = \begin{pmatrix} b_{22} & b_{23} \\ b_{32} & b_{33} \end{pmatrix} + \begin{pmatrix} c_{22} & c_{23} \\ c_{32} & c_{33} \end{pmatrix}$$

are also sums of two TP matrices. Without loss of generality, suppose that $b_{12} \geqslant c_{12}$. Thus $b_{12} \geqslant 5$ and since $b_{12}b_{21}$ must be <1, we obtain $b_{21} < \frac{1}{b_{12}} \leqslant \frac{1}{5}$, and thus $c_{21} = 10 - b_{21} \geqslant 10 - \frac{1}{5} = 9.8$. Similarly, at least one of $\{b_{23}, b_{32}, c_{23}, c_{32}\}$ is $\geqslant 9.8$. A similar analysis of the submatrix $\begin{pmatrix} 1 & 1000 \\ 1000 & 1 \end{pmatrix}$ of A shows that at least one of $\{b_{13}, b_{31}, c_{13}, c_{31}\}$ is $\geqslant 999.998$. An examination of all possible cases now shows that A cannot be written as a sum of two TP matrices.

Example 6. Let J_n denote the $n \times n$ matrix with each entry equal to 1. Then

$$J_3 = \begin{bmatrix} \frac{1}{2} & \frac{1}{3} & \frac{1}{5} \\ \frac{2}{3} & \frac{1}{2} & \frac{1}{3} \\ \frac{4}{5} & \frac{2}{3} & \frac{1}{2} \end{bmatrix} + \begin{bmatrix} \frac{1}{2} & \frac{2}{3} & \frac{4}{5} \\ \frac{1}{3} & \frac{1}{2} & \frac{2}{3} \\ \frac{1}{5} & \frac{1}{3} & \frac{1}{2} \end{bmatrix}$$

is a sum of two TP matrices.

Question 2. Can the rank 1 matrix J_n (for each $n \geqslant 4$) be written as a sum of two TP matrices? (Note that if this is possible, then any rank 1 matrix can be so written because of diagonal scaling.) It is likely that this is true for all n, and we have verified this for $n \leqslant 7$.

In view of Examples 5 and 6, we pose the following.

Question 3. If $A > 0$ is an $m \times n$ matrix, does there exist a value c (possibly depending on $\min\{m, n\}$) such that if

$$c < \frac{\min\limits_{i,j} a_{ij}}{\max\limits_{i,j} a_{ij}},$$

then A can be written as the sum of two TP matrices?

References

[1] T. Ando, Totally positive matrices, Linear Algebra Appl. 90 (1987) 165–219.
[2] F.R. Gantmacher, M.G. Krein, Oszillationskerne und Kleine Schwingungen Mechanischer Systeme, Akademie-Verlag, Berlin, 1960.
[3] C.R. Johnson, R.L. Smith, Line insertions in totally positive matrices, J. Approx. Theory 105 (2000) 305–312.
[4] S. Karlin, Total Positivity, vol. I, Stanford University Press, Stanford, 1968.