

ON NONNEGATIVE SIGN EQUIVALENT AND SIGN SIMILAR FACTORIZATIONS OF MATRICES*

DANIEL HERSHKOWITZ[†] AND ALLAN PINKUS[‡]

Dedicated to Hans Schneider on the occasion of his eightieth birthday

Abstract. It is shown that every real $n \times n$ matrix is a product of at most two nonnegative sign equivalent matrices, and every real $n \times n$ matrix, $n \geq 2$, is a product of at most three nonnegative sign similar matrices. Finally, it is proved that every real $n \times n$ matrix is a product of totally positive sign equivalent matrices. However, the question of the minimal number of such factors is left open.

Key words. Sign equivalent matrices, Sign similar matrices, Totally positive matrices, Matrix factorizations.

AMS subject classifications. 15A18, 15A29.

1. Nonnegative Sign Equivalent Factorization.

NOTATION 1.1. Let n be a positive integer, let A be an $n \times n$ matrix and let α and β be nonempty subsets of $\{1, \dots, n\}$. We denote by $A[\alpha|\beta]$ the submatrix of A whose rows and columns are indexed by α and β , respectively, in natural lexicographic order.

DEFINITION 1.2.

- i) A matrix A is said to be *totally positive* if all minors A are nonnegative.
- ii) A matrix A is said to be *strictly totally positive* if all minors A are strictly positive.
- iii) An upper triangular matrix A is said to be *triangular strictly totally positive* if all minors that can possibly be nonzero are strictly positive. That is, the determinant of $A[\alpha|\beta]$ is strictly positive whenever $\alpha = \{i_1, \dots, i_k\}$, $i_1 < \dots < i_k$, $\beta = \{j_1, \dots, j_k\}$, $j_1 < \dots < j_k$, and $i_m \leq j_m$, $m = 1, \dots, k$, for all possible α , β and k .

DEFINITION 1.3. A matrix is said to be *nonnegative sign equivalent* if it can be written in the form $D_1 Q D_2$ with Q (entrywise) nonnegative and D_1 and D_2 diagonal matrices with diagonal elements equal to ± 1 .

Clearly, not every matrix is nonnegative sign equivalent. Fully supported matrices that are nonnegative sign equivalent were characterized in [2, Theorem 4.12], while the general case is covered in [5], [6], [8], [9] and [10]. It is interesting to ask whether every real matrix is a product of nonnegative sign equivalent matrices. Also, if a matrix is a product of such matrices, what is the minimal number of nonnegative sign equivalent matrices in such a factorization?

In this section we show that *every* real matrix is a product of *at most two* nonnegative sign equivalent matrices.

*Received by the editors 21 March 2007. Accepted for publication on 12 July 2007. Handling Editor: Ludwig Elsner.

[†]Department of Mathematics, Technion - Israel Institute of Technology, Haifa 32000, Israel (her-shkow@technion.ac.il).

[‡]Department of Mathematics, Technion - Israel Institute of Technology, Haifa 32000, Israel (pinkus@technion.ac.il).

Our main result is the following factorization theorem.

THEOREM 1.4. *Let A be a real $n \times n$ matrix. Then we can always factor A in the form $A = DQB$, where D is a diagonal matrix with diagonal elements equal to ± 1 , Q is a nonnegative matrix, and B is the inverse of an upper triangular strictly totally positive matrix with diagonal elements equal to 1.*

Proof. We first construct the matrix $D = \text{diag}(d_{11}, \dots, d_{nn})$ as follows. Let $i \in \{1, \dots, n\}$. If the i th row of A contains a nonzero element then we set $d_{ii} = \text{sgn}(a_{ij})$, where $j = \min\{k : a_{ik} \neq 0\}$. If the i th row of A is identically zero then we choose d_{ii} as 1 or -1 , arbitrarily. Let P be the permutation matrix rearranging the rows of $PDA = C = \{c_{ij}\}_{i,j=1}^n$ to have the form

$$\begin{bmatrix} + & & & & & & \\ \vdots & & & & & & \\ + & & & & & & \\ 0 & + & & & & & \\ \vdots & \vdots & & & & & \\ 0 & + & & & & & \\ 0 & 0 & + & & & & \\ \vdots & & & & & & \\ 0 & \dots & \dots & \dots & 0 & + & \dots \end{bmatrix}. \quad (1.5)$$

That is, if $c_{ij} = 0$, $j = 1, \dots, m$, then $c_{lj} = 0$, $j = 1, \dots, m$, $l = i, \dots, n$. In other words, there exist n integers $l_0 = 0 \leq l_1 \leq \dots \leq l_n \leq n$ such that for every $k \in \{1, \dots, n\}$ we have

$$c_{ij} = 0, \quad i = l_{k-1} + 1, \dots, l_k, \quad 1 \leq j < k$$

$$c_{ik} > 0, \quad i = l_{k-1} + 1, \dots, l_k$$

$$c_{ij} = 0, \quad i > l_k, \quad 1 \leq j \leq k.$$

(If $l_{k-1} = l_k$ then the first two conditions are empty.) Note that if $l_n < n$, then the rows $l_n + 1, \dots, n$ are identically zero.

We have $P^{-1} = P^T$ and $D^{-1} = D$. Thus, $A = DP^TC$. We now construct an $n \times n$ matrix M which is a unit diagonal upper triangular strictly totally positive matrix and such that CM is a nonnegative matrix. We define $M = \{m_{ij}\}_{i,j=1}^n$ using the following algorithm.

Initialization: Let

$$\begin{cases} m_{ii} = 1 \\ m_{ij} = 0, \quad i > j \end{cases}, \quad i, j = 1, \dots, n. \quad (1.6)$$

Step p , $p = 1, \dots$: Let i be the largest index of a row containing elements that have not yet been fixed, and let j be the largest index such that m_{ij} has not yet been fixed. Note that by (1.6) we have $i < j$. We choose $m_{ij} > 0$ sufficiently large so that

$$\det M[\{i, i+1, \dots, i+k\}|\{j, j+1, \dots, j+k\}] > 0, \quad k = 0, \dots, n-j. \quad (1.7)$$

Observe that this is possible since we have already fixed all other entries of the rows i, \dots, n and columns j, \dots, n , and because

$$\det M[\{i+1, \dots, i+k\}|\{j+1, \dots, j+k\}] > 0, \quad k = 1, \dots, n-j.$$

Also, if $l_{i-1} < l_i$ then $m_{ij} > 0$ is chosen sufficiently large so that

$$(CM)_{pj} = c_{pi}m_{ij} + \sum_{t=i+1}^j c_{pt}m_{tj} > 0, \quad p = l_{i-1}, \dots, l_i.$$

Note that this is doable since m_{tj} have already been specified for $i < t \leq j$.

Finalization: The algorithm terminates whenever the matrix M is fully fixed.

It follows from the form (1.5) of C and the definition of the matrix M that $Q_1 = CM$ is a nonnegative matrix. Also, since M satisfies (1.6) as well as (1.7) for all i and j , it follows by [1] that M is a unit diagonal upper triangular strictly totally positive matrix. Therefore, we have $C = Q_1B$, where Q_1 is a nonnegative matrix and B is the inverse of a unit diagonal upper triangular strictly totally positive matrix. As $A = DP^TC = DP^TQ_1B$, our claim follows. \square

COROLLARY 1.8. *Every real matrix is a product of at most two nonnegative sign equivalent matrices.*

Proof. Let $A = DQB$ be a factorization of a real $n \times n$ matrix A as proven in Theorem 1.4. Being the inverse of a unit diagonal upper triangular strictly totally positive matrix, the matrix B has the form $B = D^*\hat{R}D^*$ where \hat{R} is a unit diagonal upper triangular strictly totally positive matrix, and D^* is the $n \times n$ diagonal matrix with diagonal elements $d_{ii}^* = (-1)^{i+1}$, $i = 1, \dots, n$. Thus, A can be factored in the form $A = DQD^*\hat{R}D^*$. \square

2. Nonnegative Sign Similar Factorization.

DEFINITION 2.1. A matrix is said to be *nonnegative sign similar* if it can be written in the form DQD with Q nonnegative and D a diagonal matrix with diagonal elements equal to ± 1 .

Clearly, not every matrix is nonnegative sign similar. It is known that an irreducible real matrix A is nonnegative sign similar if and only if all cyclic products of A are nonnegative. This is an easy consequence of [2, Theorem 4.1] or of [3, Theorem 4.1]. The treatment of reducible matrices can be found in [8], [9] and [10], in [5] and in [6]. In view of Corollary 1.8 it is also natural to ask whether every real matrix is a product of nonnegative sign similar matrices. In this section we show that with the obvious exception of negative 1×1 matrices, every real matrix is a product of at most three nonnegative sign similar matrices.

THEOREM 2.2. *Every real $n \times n$ matrix, $n \geq 2$, is a product of at most three nonnegative sign similar matrices.*

We divide the proof of this theorem into three parts because of the different methods of proof in each.

PROPOSITION 2.3. *Every real matrix which is not a diagonal matrix with all diagonal elements negative, is a product of at most three nonnegative sign similar matrices.*

Proof. Let $A = \{a_{ij}\}_{i,j=1}^n$ be an $n \times n$ matrix which is not a diagonal matrix with all diagonal elements negative. It follows that for some $k \in \{1, \dots, n\}$ either $a_{kk} \geq 0$ or the k th row of the matrix A contains a nonzero off-diagonal element. We define an $n \times n$ matrix $M = \{m_{ij}\}_{i,j=1}^n$ as follows. We set $m_{ii} = 1$, $i = 1, \dots, n$, and $m_{ij} = 0$ whenever $i \neq j$ and $j \neq k$. Observe that the k th column of AM is obtained by adding columns $1, \dots, k-1, k+1, \dots, n$ of A , multiplied by $m_{1k}, \dots, m_{k-1,k}, m_{k+1,k}, \dots, m_{nk}$, respectively, to the k th column. Therefore, we can assign values $m_{1k}, \dots, m_{k-1,k}, m_{k+1,k}, \dots, m_{nk}$ such that $(AM)_{ik} \neq 0$ whenever the i th row of A contains a nonzero element (including the case $i = k$) and such that $(AM)_{kk} \geq 0$. Let D_1 be the diagonal sign matrix defined by

$$(D_1)_{ii} = \begin{cases} 1, & (AM)_{ik} \geq 0 \\ -1, & (AM)_{ik} < 0. \end{cases}$$

Observe that the k th column of D_1AM is nonnegative. Since $(D_1)_{kk} = 1$, it follows that the k th column of $C = D_1AMD_1$ too is nonnegative. We now define a nonnegative $n \times n$ matrix $Q = \{q_{ij}\}_{i,j=1}^n$ as follows. We set $q_{ii} = 1$, $i = 1, \dots, n$, and $q_{ij} = 0$ whenever $i \neq j$ and $i \neq k$. Observe that $B = CQ$ is the matrix obtained by adding the k th column of C multiplied by $q_{k1}, \dots, q_{k,k-1}, q_{k,k+1}, \dots, q_{kn}$ to columns $1, \dots, k-1, k+1, \dots, n$, respectively. Therefore, we can choose $q_{k1}, \dots, q_{k,k-1}, q_{k,k+1}, \dots, q_{kn}$ to be positive numbers such that B is a nonnegative matrix. Since $D_1^{-1} = D_1$, we have

$$A = D_1BQ^{-1}D_1M^{-1}.$$

It is easy to verify that

$$(M^{-1})_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j, j \neq k \\ -m_{ik}, & j = k, i \neq k \end{cases}$$

and

$$(Q^{-1})_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j, i \neq k \\ -q_{kj}, & i = k, j \neq k. \end{cases}$$

Let D_2 and D_3 be the diagonal matrices defined by

$$(D_2)_{ii} = \begin{cases} 1, & i = k \\ 1, & i \neq k, m_{ik} < 0 \\ -1, & i \neq k, m_{ik} \geq 0 \end{cases}$$

and

$$(D_3)_{ii} = \begin{cases} 1, & i \neq k \\ -1, & i = k. \end{cases}$$

It follows that $M_1 = D_2 M^{-1} D_2$ and $Q_1 = D_3 Q^{-1} D_3 = Q$ are nonnegative matrices. Hence, we have

$$A = D_1 B D_3 Q_1 D_3 D_1 D_2 M_1 D_2 = (D_1 B D_1)(D_1 D_3 Q_1 D_3 D_1)(D_2 M_1 D_2),$$

proving our claim. \square

PROPOSITION 2.4. *Let n be an even positive integer, and let A be a diagonal $n \times n$ matrix with all diagonal elements negative. Then A is a product of two nonnegative sign similar matrices.*

Proof. Let $P = P^T$ be the permutation matrix corresponding to the permutation $(1, 2)(3, 4) \dots (n-1, n)$, and let D be the $n \times n$ diagonal matrix diagonal elements $d_{ii} = (-1)^{i+1}$, $i = 1, \dots, n$. Observe that the matrix $Q = DAPD$ is nonnegative. Therefore we have $A = DQDP$, and our claim follows. \square

PROPOSITION 2.5. *Let n be an odd positive integer, $n \geq 3$, and let A be a diagonal $n \times n$ matrix with all diagonal elements negative. Then A is a product of at most two nonnegative sign similar matrices.*

Proof. Let $Q = \{q_{ij}\}_{i,j=1}^n$ be any strictly totally positive matrix. Since n is odd, we can increase $q_{2,n-1}$ so that the determinant of $Q[\{2, \dots, n\}|\{1, \dots, n-1\}]$ becomes negative. Now, we increase $q_{1,n-1}$ so that the determinants of all $(n-1) \times (n-1)$ submatrices whose upper-rightmost element is $q_{1,n-1}$ become negative, and we increase q_{2n} so that the determinants of all $(n-1) \times (n-1)$ submatrices whose upper-rightmost element is q_{2n} become negative. Finally, we increase q_{1n} so that the determinants of all $(n-1) \times (n-1)$ submatrices whose upper-rightmost element is q_{1n} become negative and the determinant of the whole resulting matrix \bar{Q} is positive. The matrix \bar{Q} is entrywise positive, with a positive determinant, and all minors of order $n-1$ negative. Thus, the sign of the $(\bar{Q}^{-1})_{ij}$ is $(-1)^{i+j+1}$. Let $B = -\bar{Q}^{-1}$ and let D be the $n \times n$ diagonal matrix diagonal elements $d_{ii} = (-1)^{i+1}$, $i = 1, \dots, n$. Note that DBD is a nonnegative matrix and that $\bar{Q}B = -I$. Let D_1 be the diagonal matrix whose diagonal elements are the absolute values of the elements of A . A required factorization is now $A = (D_1 \bar{Q})B$. \square

REMARK 2.6. Proposition 2.4 also follows easily by the method of proof in Proposition 2.5. The desired matrix \bar{Q} is obtained by taking any strictly totally positive matrix and simply reversing the order of the rows (or columns).

REMARK 2.7. Clearly, a 1×1 matrix with a negative element cannot be written as a product of 1×1 nonnegative sign similar matrices.

In view of Theorem 2.2 the minimal number of nonnegative sign similar matrices in a factorization of a real matrix does not exceed three. The following example shows that the minimal number is three.

EXAMPLE 2.8. Let

$$A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Since $a_{22} < 0$, the matrix A is not nonnegative sign similar. We now show that A is also not a product of two nonnegative sign similar matrices. Assume to the contrary that $A = B_1 B_2$, where B_1 and B_2 are nonnegative sign similar. It is easy to verify that a 2×2 matrix is nonnegative sign similar if and only if it is either of type

$$\begin{bmatrix} + & + \\ + & + \end{bmatrix} \quad (2.9)$$

or of type

$$\begin{bmatrix} + & - \\ - & + \end{bmatrix}, \quad (2.10)$$

where “+” denotes a nonnegative element and “−” denotes a nonpositive element. Since the product of two matrices of type (2.9) is a matrix of type (2.9) and a product of two matrices of type (2.10) is a matrix of type (2.10), the matrices B_1 and B_2 are of different types. Since $A = A^T$, without loss of generality we may assume that B_1 is of type (2.9) and B_2 is of type (2.10). Let $D = \text{diag}(1, -1)$. Since $B_1 B_2 = A$, we have $B_1(B_2 D) = I$, and so $B_1^{-1} = B_2 D$. Note that $B_2 D$ is of type

$$\begin{bmatrix} + & + \\ - & - \end{bmatrix}. \quad (2.11)$$

Observe that if $\det B_1 > 0$ then B_1^{-1} is of type (2.10). For B_1^{-1} to be of both types (2.10) and (2.11) it requires that the second column of B_1^{-1} is a zero column, which is impossible. Similarly, if $\det B_1 < 0$ then B_1^{-1} is of type

$$\begin{bmatrix} - & + \\ + & - \end{bmatrix}. \quad (2.12)$$

For B_1^{-1} to be of both types (2.11) and (2.12) it requires that the first column of B_1^{-1} is a zero column, which is impossible. Therefore, our assumption that A is a product of two nonnegative sign similar matrices is false.

There are numerous products of three nonnegative sign similar matrices that give A . For example,

$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}.$$

3. Totally Positive Sign Equivalent Factorization.

DEFINITION 3.1. A matrix is said to be *totally positive sign equivalent* if it can be written in the form $D_1 Q D_2$ with Q totally positive and D_1 and D_2 diagonal matrices with diagonal elements equal to ± 1 .

It is interesting to ask whether every real matrix is a product of totally positive sign equivalent matrices. Also, if a matrix is a product of such matrices, what is the minimal number of totally positive sign equivalent matrices in such a factorization? We will prove that every matrix is a product of totally positive sign equivalent matrices. However we do not know the minimal number of totally positive sign equivalent matrices in such a factorization. The number is at least three, as we shall show. But this question remains open.

PROPOSITION 3.2. *Every square real matrix is a product of totally positive sign equivalent matrices.*

Proof. The proof is a simple consequence of the fact that every matrix $A = \{a_{ij}\}_{i,j=1}^n$, where $a_{ii} \geq 0$, $i = 1, \dots, n$, $a_{j,j+1} \geq 0$ (or $a_{j+1,j} \geq 0$), $j = 1, \dots, n-1$, and all other entries identically zero, is totally positive. As such the class of totally positive sign equivalent matrices includes all *bidiagonal* matrices, that is matrices of the form $A = \{a_{ij}\}_{i,j=1}^n$, where the only nonzero entries are possibly a_{ii} , $i = 1, \dots, n$, and $a_{j,j+1}$ (or $a_{j+1,j}$), $j = 1, \dots, n-1$. It is well-known that every matrix can be factored as a product of bidiagonal matrices. The existence of such a factorization, the so-called Loewner-Neville factorization, was proven in [4] for square matrices satisfying certain invertibility conditions. It was proven for general matrices in [7], where there is also a discussion on the minimal number of required factors. \square

EXAMPLE 3.3. Let us show that the matrix

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

is the product of exactly three totally positive sign equivalent matrices, and no less. To see that three suffices consider, for example, the factorization:

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

It is easily verified that each factor on the right-hand-side of this factorization is totally positive sign equivalent.

It remains to show that A cannot be factored as a product of two totally positive sign equivalent matrices. Assume to the contrary that

$$A = D_1 B D_2 C D_3,$$

where D_1 , D_2 and D_3 are diagonal matrices with diagonal elements equal to ± 1 , while B and C are totally positive. Thus

$$D_1 A D_3 = B D_2 C$$

and as is easily checked,

$$D_1 A D_3 = E = \begin{bmatrix} 0 & \sigma_1 \\ \sigma_2 & 0 \end{bmatrix},$$

where $\sigma_1, \sigma_2 \in \{-1, 1\}$. As E is nonsingular, so are B and C , and therefore

$$B^{-1} = D_2 C E^{-1}.$$

Let

$$B^{-1} = \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix}.$$

As B is totally positive we have that $\det B^{-1} > 0$, $g_{11}, g_{22} > 0$ and $g_{12}, g_{21} \leq 0$.

Let

$$D_2 = \begin{bmatrix} d_1 & 0 \\ 0 & d_2 \end{bmatrix}, \quad C = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}.$$

As

$$E^{-1} = \begin{bmatrix} 0 & \sigma_2 \\ \sigma_1 & 0 \end{bmatrix}$$

we have

$$D_2 C E^{-1} = \begin{bmatrix} d_1 \sigma_1 c_{12} & d_1 \sigma_2 c_{11} \\ d_2 \sigma_1 c_{22} & d_2 \sigma_2 c_{21} \end{bmatrix}.$$

We now compare B^{-1} and $D_2 C E^{-1}$, and arrive at a contradiction.

As C is totally positive and nonsingular we have $c_{11}, c_{22} > 0$, $c_{12}, c_{21} \geq 0$, and

$$\det C = c_{11}c_{22} - c_{12}c_{21} > 0.$$

Thus, from the form of B^{-1} , $d_1 \sigma_1 = d_2 \sigma_2 = 1$ while $d_1 \sigma_2 = d_2 \sigma_1 = -1$. Now from the properties of B^{-1} and C ,

$$0 < \det B^{-1} = \det D_2 C E^{-1} = c_{12}c_{21} - c_{11}c_{22} < 0,$$

which is a contradiction.

REFERENCES

- [1] C. W. Cryer. The LU -factorization of totally positive matrices. *Linear Algebra and its Applications*, 7:83–92, 1973.
- [2] G. M. Engel and H. Schneider. Cyclic and diagonal products on a matrix. *Linear Algebra and its Applications*, 7:301–335, 1973.
- [3] G. M. Engel and H. Schneider. Diagonal similarity and diagonal equivalence for matrices over groups with 0. *Czechoslovak Mathematical Journal*, 25(100):387–403, 1975.
- [4] M. Fiedler and T. L. Markham. Consecutive-column and -row Properties of matrices and the Loewner-Neville factorization. *Linear Algebra and its Applications*, 266:243–259, 1977.
- [5] M. v. Golitschek, U. G. Rothblum, and H. Schneider. A conforming decomposition theorem, a piecewise nonlinear theorem of the alternative and scalings of matrices satisfying lower and upper bounds. *Mathematical Programming*, 27:291–306, 1983.

- [6] D. Hershkowitz and H. Schneider. One-sided simultaneous inequalities and sandwich theorems for diagonal similarity and diagonal equivalence of nonnegative matrices. *Electronic Journal of Linear Algebra*, 10:81–101, 2003.
- [7] C. R. Johnson, D. D. Olesky, and P. van den Driessche. Elementary bidiagonal factorizations. *Linear Algebra and its Applications*, 292:233–244, 1999.
- [8] B. D. Saunders and H. Schneider. Flows of graphs applied to diagonal similarity and equivalence of matrices. *Discrete Mathematics*, 24:139–162, 1978.
- [9] B. D. Saunders and H. Schneider. Applications of the Gordan-Stiemke theorem in combinatorial matrix theory. *SIAM Review*, 21:528–541, 1979.
- [10] B. D. Saunders and H. Schneider. Cones, graphs and optimal scalings of matrices. *Linear and Multilinear Algebra*, 8:121–135, 1979.