Nonnegative Normal Matrices

S. K. Jain and L. E. Snyder
Department of Mathematics
Ohio University
Athens, Ohio 45701

Submitted by Richard A. Brualdi

ABSTRACT

A structural characterization is given for the class of those nonnegative matrices for which the transpose is a polynomial in the matrix with the polynomial having nonnegative coefficients and no constant term.

1. INTRODUCTION

A large number of characterizations of the important class of normal matrices are given by Grone et al. in [2]. One of the characterizations for a real normal matrix is that its transpose is a polynomial in the matrix. A characterization for the class of those stochastic matrices for which the transpose is equal to some power of the matrix is given by Sinkhorn in [4]. Our result extends Sinkhorn's result to any nonnegative A for which $A^T = \alpha_1 A^{m_1} + \dots + \alpha_k A^{m_k}$, where each $\alpha_i > 0$ and $0 < m_1 < m_2 < \cdots < m_k$.

2. DEFINITIONS AND NOTATION

A matrix $A = (a_{ij})$ is said to be *nonnegative* if $a_{ij} \geqslant 0$ for all i, j, and A is said to be *positive* if $a_{ij} > 0$ for all i, j. A real matrix A is said to be *normal* if A commutes with A^T, the transpose of A. X is said to be a 1-*inverse* of A if $AXA = A$, and is said to be a $\{1,2\}$-*inverse* of A if in addition $XAX = X$. A is said to be *stochastic* if A is nonnegative and the

655 Avenue of the Americas, New York, NY 10010 0024-3795/93/$6.00

sum of each row is one. $A = (a_{ij})$ is called a *Z-matrix* if $a_{ij} \leqslant 0$ for $i \neq j$; A is an *M-matrix* if A can be expressed in the form $A = sI - B$, where $s > 0$, $B \geqslant 0$, and $s > \rho(B)$, the spectral radius of B.

3. MAIN RESULTS

THEOREM 1. *If A is a nonnegative, nonzero matrix such that*

$$A^T = p(A),$$

where p is a polynomial,

$$p(t) = \alpha_1 t^{m_1} + \alpha_2 t^{m_2} + \cdots + \alpha_k t^{m_k},$$

with all $\alpha_i > 0$, and $0 < m_1 < m_2 < \cdots < m_k$, then either A is symmetric or there is a permutation matrix P such that

$$PAP^T = \begin{pmatrix} J & 0 \\ 0 & 0 \end{pmatrix},$$

where J is a direct sum of matrices of the following types:

(I) $\beta x x^T$, *where x is a positive column vector with $x^T x = 1$,*
(II)

$$\beta \begin{pmatrix} 0 & x_1 x_2^T & 0 & 0 & \cdots & 0 \\ 0 & 0 & x_2 x_3^T & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & & \vdots \\ 0 & 0 & 0 & \cdots & 0 & x_{d-1} x_d^T \\ x_d x_1^T & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}$$

where the x_i's are positive unit column vectors, possibly of different orders, and β is the unique positive fixed point of the polynomial p. For each summand of type (II) *which is a $d \times d$ block matrix, d must divide $m_i + 1$ for each of the exponents m_i in the polynomial p.*

Proof. By hypothesis

$$A^T = p(A),$$

where p is a polynomial,

$$p(t) = \alpha_1 t^{m_1} + \alpha_2 t^{m_2} + \cdots + \alpha_k t^{m_k},$$

such that $\alpha_i > 0$ and $m_i > 0$ for each i. Then $A = p(A^T) = p(p(A))$, i.e.,

$$A = \alpha_1[p(A)]^{m_1} + \dots + \alpha_k[p(A)]^{m_k}. \tag{1}$$

First we determine the structure of matrices satisfying the polynomial equation (1) with positive coefficients.

Suppose $m_1 = 1$. Then from (1) we obtain

$$(1 - \alpha_1^2)A = \alpha_1(\alpha_2 A^{m_2} + \dots) + \alpha_2[p(A)]^{m_2} + \cdots + \alpha_k[p(A)]^{m_k}. \tag{2}$$

If $\alpha_1 = 1$, then $\alpha_2 = \alpha_3 = \cdots = \alpha_k = 0$, since the left side of (2) is the zero matrix and so $k = 1$. Hence A is symmetric. If $\alpha_1 \neq 1$, then $\alpha_1 < 1$ since the right side of Equation (2) is nonnegative. Equation (2) thus allows us to express A as

$$A = A^2 X, \tag{3}$$

where X is a nonnegative matrix which commutes with A. Consequently, A has a nonnegative 1-inverse.

In case $m_1 \geqslant 2$, it follows easily from Equation (1) that A can be expressed as in (3), so again A has a nonnegative 1-inverse.

If X is a nonnegative 1-inverse of A, then $Y = XAX$ is a nonnegative $\{1,2\}$-inverse of A. Consequently, by Theorem 1 of [3] there is a permutation matrix P such that

$$PAP^T = M = \begin{pmatrix} J & JD & 0 & 0 \\ 0 & 0 & 0 & 0 \\ CJ & CJD & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

where C and D are some nonnegative matrices of appropriate sizes and J is a direct sum of matrices of the following types:

(I) βxy^T, where x and y are positive vectors with $y^Tx = 1$,
(II)

$$\begin{pmatrix} 0 & \beta_{12}x_1y_2^T & 0 & 0 & \cdots & 0 \\ 0 & 0 & \beta_{23}x_2y_3^T & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & & \vdots \\ 0 & 0 & 0 & \cdots & 0 & \beta_{d-1d}x_{d-1}y_d^T \\ \beta_{d1}x_dy_1^T & 0 & 0 & \cdots & 0 & 0 \end{pmatrix},$$

where x_i and y_i are positive vectors of the same order with $y_i^Tx_i = 1$ and the β_{ij}'s are positive. x_i and x_j are not necessarily of the same order if $i \neq j$.

Since $A^T = p(A)$, it is also true that $M^T = PA^TP^T = Pp(A)P^T = p(PAP^T) = p(M)$. From this it follows that the (2, 1) block and the (1, 3) block of M^T must be zero matrices, or equivalently, $JD = 0$ and $CJ = 0$. That is,

$$M = PAP^T = \begin{pmatrix} J & 0 \\ 0 & 0 \end{pmatrix}.$$

The matrix J also will satisfy the condition that $J^T = p(J)$, and J being a direct sum of matrices implies that each of the summands S will likewise satisfy the condition $S^T = p(S)$ for the same polynomial p.

Let us consider a summand S of type I, i.e., $S = \beta xy^T$, with $y^Tx = 1$ and x, y being positive vectors. Note that $(xy^T)^j = xy^T$ for all $j \geqslant 1$. Hence

$$\beta yx^T = p(\beta xy^T) = (\alpha_1\beta^{m_1} + \alpha_2\beta^{m_2} + \cdots + \alpha_k\beta^{m_k})xy^T. \qquad (4)$$

Postmultiply both sides of (4) by x, then

$$(\beta x^Tx)y = (\alpha_1\beta^{m_1} + \cdots + \alpha_k\beta^{m_k})x. \qquad (5)$$

It follows that y equals a positive scalar times the positive vector x. Consequently there is no loss in generality in assuming that S has the form $S = \beta xx^T$, with x being a positive unit vector. Then (4) becomes

$$\beta xx^T = (\alpha_1\beta^{m_1} + \cdots + \alpha_k\beta^{m_k})xx^T$$

which implies that $\beta = p(\beta)$ so that β is the unique positive fixed point of the polynomial p.

Next let us consider a summand S of the type II and the powers of S. For example, when $d = 4$, then

$$S = \begin{pmatrix} 0 & \beta_{12}x_1y_2^T & 0 & 0 \\ 0 & 0 & \beta_{23}x_2y_3^T & 0 \\ 0 & 0 & 0 & \beta_{34}x_3y_4^T \\ \beta_{41}x_4y_1^T & 0 & 0 & 0 \end{pmatrix},$$

$$S^2 = \begin{pmatrix} 0 & 0 & \beta_{12}\beta_{23}x_1y_3^T & 0 \\ 0 & 0 & 0 & \beta_{23}\beta_{34}x_2y_4^T \\ \beta_{34}\beta_{41}x_3y_1^T & 0 & 0 & 0 \\ 0 & \beta_{41}\beta_{12}x_4y_2^T & 0 & 0 \end{pmatrix},$$

$$S^3 = \begin{pmatrix} 0 & 0 & 0 & \beta_{12}\beta_{23}\beta_{34}x_1y_4^T \\ \beta_{23}\beta_{34}\beta_{41}x_2y_1^T & 0 & 0 & 0 \\ 0 & \beta_{34}\beta_{41}\beta_{12}x_3y_2^T & 0 & 0 \\ 0 & 0 & \beta_{41}\beta_{12}\beta_{23}x_4y_3^T & 0 \end{pmatrix},$$

and S^4 is a block diagonal matrix with the ith block having the form $\mu x_i y_i^T$, where $\mu = \beta_{12}\beta_{23}\beta_{34}\beta_{41}$. In general S^d is a block diagonal matrix and the ith block is $\mu x_i y_i^T$, where $\mu = \beta_{12}\beta_{23} \cdots \beta_{d1}$. Likewise S^{qd} is a block diagonal matrix for any q, and the ith block is $\mu^q x_i y_i^T$.

From this it follows that for $0 < r < d$, $S^{dq+r} = \mu^q S^r$.

Each of the exponents m_i in the polynomial p can be expressed as $m_i = dq_i + r_i$, where $0 \leqslant r_i < d$. Then

$$S^T = p(S) = \sum \alpha_i S^{dq_i + r_i} = \sum \alpha_i \mu^{q_i} S^{r_i}. \tag{6}$$

Since each $\alpha_i > 0$ and only the nonzero blocks of S^{d-1} can match the corresponding nonzero blocks of S^T, it follows that each $r_i = d - 1$.

Therefore, for any $d \geqslant 2$, the exponents m_i of the polynomial p must be of the form $m_i = dq_i + d - 1$. The $(1, d)$ blocks in both sides of Equation (6) are equal to

$$\beta_{d1} y_1 x_d^T = \sum \alpha_i \mu^{q_i} \beta_{12}\beta_{23} \cdots \beta_{d-1,d} x_1 y_d^T.$$

Multiplying this equation on the right by x_d leads to the result that y_1 is a positive scalar multiple of x_1.

Similarly it can be shown that each y_i is a positive scalar multiple of x_i, and so without loss of generality it may be assumed that each x_i is a unit vector and that $y_i = x_i$ for each i.

Now let us equate the corresponding blocks of the diagonal in the equation $SS^T = S^TS$. For the (1, 1) blocks we have $\beta_{12}^2 x_1 x_1^T = \beta_{d1}^2 x_1 x_1^T$, and for the other blocks we have $\beta_{23}^2 x_2 x_2^T = \beta_{12}^2 x_2 x_2^T$, $\beta_{34}^2 x_3 x_3^T = \beta_{23}^2 x_3 x_3^T$, etc. Hence $\beta_{12} = \beta_{23} = \beta_{34} = \cdots = \beta_{d1}$.

Thus the matrix S must have the form

$$S = \beta \begin{pmatrix} 0 & x_1 x_2^T & 0 & 0 & \dots & 0 \\ 0 & 0 & x_2 x_3^T & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & & \vdots \\ 0 & 0 & 0 & \dots & 0 & x_{d-1} x_d^T \\ x_d x_1^T & 0 & 0 & \dots & 0 & 0 \end{pmatrix}.$$

Let $C = (1/\beta)S$. Since $C^{m_i} = C^T$ whenever $m_i + 1$ is a multiple of d, it follows from $S^T = p(S)$ that $\beta = p(\beta)$. ■

REMARK 1. If xy^T is stochastic, then all components of x are equal. Furthermore, if all components of y are also equal though not necessarily the same as for x, then $xy^T = J_k$, where k is the size of x and J_k is the square matrix of order k with all entries equal to $1/k$.

A normal stochastic matrix A is doubly stochastic (see [4, p. 225]), i.e., A^T is stochastic. Sinkhorn's structural characterization [4] of those stochastic matrices A for which $A^T = A^k$, follows from Theorem 1 and Remark 1.

The conditions in Theorem 1 are also sufficient.

THEOREM 2. *If there is a permutation matrix P such that PAP^T is a direct sum of the nonnegative matrices of the types* (I) *and* (II), *then there is a monomial p with $p(\beta) = \beta$ and $A^T = p(A)$.*

Proof. Note that if S is a summand of type (II), then $S = \beta C$, where C is a periodic matrix, i.e., $C^{d+1} = C$. Let $m = \operatorname{lcm}(d_1, d_2, \dots, d_k) - 1$, where each summand of type (II) is a $d_i \times d_i$ block matrix. [If only type (I) summands occur, then A will be symmetric and the result is obvious.] Then

$S^m = \beta^m C^m = \beta^m C^T = \beta^{m-1} S^T$. Hence, if we let $p(t) = (1/\beta^{m-1})t^m$, then $p(S) = S^T$ for each summand S; therefore $p(A) = A^T$, ∎

REMARK 2. Note that a positive normal matrix A which is not symmetric cannot satisfy the relation $p(A) = A^T$ for any polynomial p with nonnegative coefficients and no constant term. For example, the circulant matrix

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

is such a matrix.

A natural question arises as to what the structure is for matrices A for which $A^T = p(A)$ where $p(t)$ is a polynomial with nonnegative coefficients and $p(0) \neq 0$. We have not been able to settle this question, but we give below some observations, contained in Propositions 1 and 2, which may be of interest.

PROPOSITION 1. *If A is a nonnegative matrix such that*

$$A^T = p(A), \qquad \textit{where} \quad p(A) = \alpha_o I + \alpha_1 A^{m_1} + \dots \alpha_k A^{m_k},$$

with $\alpha_0, \alpha_1, \dots, \alpha_k$ positive, then

(i) *A has at most two positive eigenvalues,*
(ii) *A^{-1} exists and is an M-matrix.*

Proof. (i): First we note that if λ_1 is a positive eigenvalue of A, then $p(\lambda_1)$ is also a positive eigenvalue of A because A and A^T have the same eigenvalues. Since the function p is strictly increasing on the interval $[0,\infty)$, it follows that any positive eigenvalue of A must be a fixed point of p. Also, the graph of p is concave upward on the interval $[0,\infty)$, and so there can be at most two such fixed points.

(ii): From $A^T = p(A)$, we have

$$A = p(A^T) = p(p(A)) = \alpha_0 I + \alpha_1 [p(A)]^{m_1} + \cdots + \alpha_k [p(A)]^{m_k},$$

so $A = p(\alpha_o)I + Aq(A)$ for some polynomial q with nonnegative coefficients. Note that $p(\alpha_0) > 0$, so

$$I = \frac{1}{p(\alpha_0)}[A - Aq(A)] = \frac{1}{p(\alpha_0)}[I - q(A)]A.$$

This shows that A^{-1} exists and that A^{-1} is a Z-matrix [since $q(A) \geqslant 0$]. By condition N_{38} of [1, p. 137], A^{-1} is an M-matrix. ■

PROPOSITION 2. *If A is a nonsymmetric matrix with $A \geqslant 0$ and if $A^T = p(A)$, where $p(t)$ has nonnegative coefficients and $p(0) \neq 0$, then* deg $p(t) > 2$.

Proof. Assume deg $p(t) = 2$. From $A = p(p(A))$ we have

$$0 = p(p(A)) - A = \beta_0 I - \beta_1 A + \beta_2 A^2 + \beta_3 A^3 + \beta_4 A^4, \qquad (7)$$

where β_0, β_1, and β_4 are positive and the other β's are nonnegative. From Proposition 1 we know that A^{-1} is an M-matrix, so the eigenvalues of A^{-1} and A must have positive real parts (condition F_{12} of [1, p. 150] and that A has at most two real eigenvalues. Hence the minimum polynomial of A must contain a factor $t - \lambda_1$ and a factor of the form $t^2 - at + b$, where λ_1, a, and b are positive. The product of these two factors, say $(t - \lambda_1)(t^2 - at + b) = t^3 - \gamma_2 t^2 + \gamma_1 t - \gamma_0$, with each $\gamma_i \geqslant 0$ and $\gamma_0 > 0$, must divide the right side of (7). The quotient must be of the form $\beta_4(t - \lambda_2)$ with $\lambda_2 > 0$, since $\beta_0 > 0$. However, this would then contradict the fact that $\beta_3 \geqslant 0$. Hence we must have deg $p(t) > 2$, completing the proof. ■

EXAMPLE. The following matrix A satisfies the equation $A^T = \frac{1}{8}I + \frac{1}{2}A + \frac{1}{8}A^2$ and has the eigenvalues $\lambda_1 = 2 - \sqrt{3}$ and $\lambda_2 = 2 + \sqrt{3}$:

$$A = \begin{pmatrix} 2 & \sqrt{3} \\ \sqrt{3} & 2 \end{pmatrix}.$$

Although it is possible to find larger matrices which satisfy the same polynomial equation in the example above, Proposition 2 shows that any such matrix must be symmetric. The authors have been unable to find a nonsymmetric matrix A and a polynomial $p(t)$ with nonnegative coefficients and $p(0) \neq 0$ for which $A^T = p(A)$.

The authors wish to thank the referee for his helpful suggestions and comments.

REFERENCES

1 A. Berman and R. J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*, Academic, New York, 1979.
2 R. Grone, C. R. Johnson, E. M. Sa, and H. Wolkowicz, Normal matrices, *Linear Algebra Appl.* 87:213–225 (1987).
3 S. K. Jain, E. K. Kwak, and V. K. Goel, Decomposition of nonnegative group-monotone matrices, *Trans. Amer. Math. Soc.* 257:371–385 (1980).
4 R. Sinkhorn, Power symmetric stochastic matrices, *Linear Algebra Appl.* 40:225–228 (1981).

Received 24 July 1991; final manuscript accepted 26 January 1992