



ELSEVIER

Contents lists available at ScienceDirect

Linear Algebra and its Applications

journal homepage: www.elsevier.com/locate/laa



On p th roots of stochastic matrices

Nicholas J. Higham ^{*,1}, Lijing Lin ^{**}

School of Mathematics, The University of Manchester, Manchester M13 9PL, UK

ARTICLE INFO

Article history:

Available online 5 May 2010

Submitted by V. Mehrmann

Dedicated to Pete Stewart on the occasion of his 70th birthday.

AMS classification:

15A51

65F15

Keywords:

Stochastic matrix
Nonnegative matrix
Matrix p th root
Primary matrix function
Nonprimary matrix function
Perron–Frobenius theorem
Markov chain
Transition matrix
Embeddability problem
 M -matrix
Inverse eigenvalue problem

ABSTRACT

In Markov chain models in finance and healthcare a transition matrix over a certain time interval is needed but only a transition matrix over a longer time interval may be available. The problem arises of determining a stochastic p th root of a stochastic matrix (the given transition matrix). By exploiting the theory of functions of matrices, we develop results on the existence and characterization of matrix p th roots, and in particular on the existence of stochastic p th roots of stochastic matrices. Our contributions include characterization of when a real matrix has a real p th root, a classification of p th roots of a possibly singular matrix, a sufficient condition for a p th root of a stochastic matrix to have unit row sums, and the identification of two classes of stochastic matrices that have stochastic p th roots for all p . We also delineate a wide variety of possible configurations as regards existence, nature (primary or nonprimary), and number of stochastic roots, and develop a necessary condition for existence of a stochastic root in terms of the spectrum of the given matrix.

© 2010 Elsevier Inc. All rights reserved.

* Corresponding author.

** Corresponding author.

E-mail addresses: higham@ma.man.ac.uk (N.J. Higham), Lijing.Lin@postgrad.manchester.ac.uk (L. Lin).

URL: <http://www.ma.man.ac.uk/~higham> (N.J. Higham), <http://www.maths.manchester.ac.uk/~lijing> (L. Lin).

¹ The work of this author was supported by a Royal Society-Wolfson Research Merit Award and by Engineering and Physical Sciences Research Council Grant EP/D079403.

1. Introduction

Discrete-time Markov chains are in widespread use for modelling processes that evolve with time. Such processes include the variations of credit risk in the finance industry and the progress of a chronic disease in healthcare, and in both cases the particular problem considered here arises.

In credit risk, a transition matrix records the probabilities of a firm's transition from one credit rating to another over a given time interval [22]. The shortest period over which a transition matrix can be estimated is typically one year, and annual transition matrices can be obtained from rating agencies such as Moody's Investors Service and Standard & Poor's. However, for valuation purposes, a transition matrix for a period shorter than one year is usually needed. A short term transition matrix can be obtained by computing a root of an annual transition matrix. A six-month transition matrix, for example, is a square root of the annual transition matrix. This property has led to interest in the finance literature in the computation or approximation of roots of transition matrices [21,25]. Exactly the same mathematical problem arises in Markov models of chronic diseases, where the transition matrix is built from observations of the progression in patients of a disease through different severity states. Again, the observations are at an interval longer than the short time intervals required for study and the need for a matrix root arises [4]. An early discussion of this problem, which identifies the need for roots of transition matrices in models of business and trade, is that of Waugh and Abel [37].

A transition matrix is a stochastic matrix: a square matrix with nonnegative entries and row sums equal to 1. The applications we have described require a stochastic root of a given stochastic matrix A , that is, a stochastic matrix X such that $X^p = A$, where p is typically a positive integer. Mathematically, there are three main questions.

1. Under what conditions does a given stochastic matrix A have a stochastic p th root, and how many roots are there?
2. If a stochastic root exists, how can it be computed?
3. If a stochastic root does not exist, what is an appropriate approximate stochastic root to use in its place?

The focus of this work is on the first question, which has not previously been investigated in any depth. Regarding the second and third questions, various methods are available for computing matrix p th roots [2,11,13,14], [16, Chapter 7], [19,20,34], but there are currently no methods tailored to finding a stochastic root. Current approaches are based on computing *some* p th root and perturbing it to be stochastic [4,21,25].

In Section 2 we recall known results on the existence of matrix p th roots and derive a new characterization of when a real matrix has a real p th root. With the aid of a lemma describing the p th roots of block triangular matrices whose diagonal blocks have distinct spectra, we obtain a classification of p th roots of possibly singular matrices. In Section 3 we derive a sufficient condition for a p th root of a stochastic matrix A to have unit row sums; we show that this condition is necessary for primary roots and that a nonnegative p th root always has unit row sums when A is irreducible. We use the latter result to connect the stochastic root problem with the problem of finding nonnegative roots of nonnegative matrices. Two classes of stochastic matrices are identified that have stochastic principal p th roots for all p : one is the inverse M -matrices and the other is a class of symmetric positive semidefinite matrices explicitly obtained from a construction of Soules. In Section 4 we demonstrate a wide variety of possible scenarios for the existence and uniqueness of stochastic roots of a stochastic matrix—in particular, with respect to whether a stochastic root is principal, primary, or nonprimary. In Section 5 we exploit a result on the inverse eigenvalue problem for stochastic matrices in order to obtain necessary conditions that the spectrum of a stochastic matrix must satisfy in order for the matrix to have a stochastic p th root. Finally, some conclusions are given in Section 6.

2. Theory of matrix p th roots

We are interested in the nonlinear equation $X^p = A$, where p is assumed to be a positive integer. In practice, p might be rational—for example if a transition matrix is observed for a five year time interval

but the interval of interest is two years. If $p = r/s$ for positive integer r and s then the problem is to solve the equation $X^r = A^s$, and this reduces to the original problem with $p \leftarrow r$ and $A \leftarrow A^s$, since any positive integer power of a stochastic matrix is stochastic.

We can understand the nonlinear equation $X^p = A$ through the theory of functions of matrices. The following theorem classifies all p th roots of a nonsingular matrix [16, Theorem 7.1], [34] and will be exploited below.

Theorem 2.1 (classification of p th roots of nonsingular matrices). *Let the nonsingular matrix $A \in \mathbb{C}^{n \times n}$ have the Jordan canonical form $Z^{-1}AZ = J = \text{diag}(J_1, J_2, \dots, J_m)$, with Jordan blocks $J_k = J_k(\lambda_k) \in \mathbb{C}^{m_k \times m_k}$, and let $s \leq m$ be the number of distinct eigenvalues of A . Let $L_k^{(j_k)} = L_k^{(j_k)}(\lambda_k)$, $k = 1 : m$, denote the p th roots of J_k given by*

$$L_k^{(j_k)}(\lambda_k) := \begin{bmatrix} f_{j_k}(\lambda_k) & f'_{j_k}(\lambda_k) & \cdots & \frac{f_{j_k}^{(m_k-1)}(\lambda_k)}{(m_k-1)!} \\ & f_{j_k}(\lambda_k) & \ddots & \vdots \\ & & \ddots & f'_{j_k}(\lambda_k) \\ & & & f_{j_k}(\lambda_k) \end{bmatrix}, \tag{2.1}$$

where $j_k \in \{1, 2, \dots, p\}$ denotes the branch of the p th root function $f(z) = \sqrt[p]{z}$. Then A has precisely p^s p th roots that are expressible as polynomials in A , given by

$$X_j = Z \text{diag}(L_1^{(j_1)}, L_2^{(j_2)}, \dots, L_m^{(j_m)})Z^{-1}, \quad j = 1 : p^s, \tag{2.2}$$

corresponding to all possible choices of j_1, \dots, j_m , subject to the constraint that $j_i = j_k$ whenever $\lambda_i = \lambda_k$. If $s < m$ then A has additional p th roots that form parametrized families

$$X_j(U) = ZU \text{diag}(L_1^{(j_1)}, L_2^{(j_2)}, \dots, L_m^{(j_m)})U^{-1}Z^{-1}, \quad j = p^s + 1 : p^m, \tag{2.3}$$

where $j_k \in \{1, 2, \dots, p\}$, U is an arbitrary nonsingular matrix that commutes with J , and for each j there exist i and k , depending on j , such that $\lambda_i = \lambda_k$ while $j_i \neq j_k$.

In the theory of matrix functions the roots (2.2) are called primary functions of A , and the roots in (2.3), which exist only if A is derogatory (that is, if some eigenvalue appears in more than one Jordan block), are called nonprimary functions [16, Chapter 1]. A distinguishing feature of the primary roots (2.2) is that they are expressible as polynomials in A , whereas the nonprimary roots are not. To give some insight into the theorem and the nature of nonprimary roots, we consider

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

which is already in Jordan form, and for which $m = 2, s = 1$. All square roots are given by

$$\pm \begin{bmatrix} 1 & \frac{1}{2} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \pm U \begin{bmatrix} 1 & \frac{1}{2} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} U^{-1},$$

where from the standard characterization of commuting matrices [16, Theorem 1.25] we find that U is an arbitrary nonsingular matrix of the form

$$U = \begin{bmatrix} a & b & d \\ 0 & a & 0 \\ 0 & e & c \end{bmatrix}.$$

While a nonsingular matrix always has a p th root, the situation is more complicated for singular matrices, as the following result of Psarrakos [33] shows.

Theorem 2.2 (existence of p th root). $A \in \mathbb{C}^{n \times n}$ has a p th root if and only if the “ascent sequence” of integers d_1, d_2, \dots defined by

$$d_i = \dim(\text{null}(A^i)) - \dim(\text{null}(A^{i-1})) \tag{2.4}$$

has the property that for every integer $v \geq 0$ no more than one element of the sequence lies strictly between pv and $p(v + 1)$.

For real A , the above theorems do not distinguish between real and complex roots. The next theorem provides a necessary and sufficient condition for the existence of a real p th root of a real A ; it generalizes [18, Theorem 6.4.14], which covers the case $p = 2$, and [36, Corollary to Theorem 1], which applies to nonsingular A .

Theorem 2.3 (existence of real p th root). $A \in \mathbb{R}^{n \times n}$ has a real p th root if and only if it satisfies the ascent sequence condition (2.4) and, if p is even, A has an even number of Jordan blocks of each size for every negative eigenvalue.

Proof. First, we note that a given Jordan canonical form J is that of some real matrix if and only if for every nonreal eigenvalue λ occurring in r Jordan blocks of size q there are also r Jordan blocks of size q corresponding to $\bar{\lambda}$; in other words, the Jordan blocks of each size for nonreal eigenvalues come in complex conjugate pairs. This property is a consequence of the real Jordan form and its relation to the complex Jordan form [18, Section 3.4], [26, Section 6.7].

(\Rightarrow) If A has a real p th root then by Theorem 2.2 it must satisfy (2.4). Suppose that p is even, that A has an odd number, $2k + 1$, of Jordan blocks of size m for some m and some eigenvalue $\lambda < 0$, and that there exists a real X with $X^p = A$. Since a nonsingular Jordan block does not split into smaller Jordan blocks when raised to a positive integer power [16, Theorem 1.36], the Jordan form of X must contain exactly $2k + 1$ Jordan blocks of size m corresponding to eigenvalues μ_j with $\mu_j^p = \lambda$, which implies that each μ_j is nonreal since $\lambda < 0$ and p is even. In order for X to be real these Jordan blocks must occur in complex conjugate pairs, but this is impossible since there is an odd number of them. Hence we have a contradiction, so A must have an even number of Jordan blocks of size m for λ .

(\Leftarrow) A has a Jordan canonical form $Z^{-1}AZ = J = \text{diag}(J_0, J_1)$, where J_0 collects together all the Jordan blocks corresponding to the eigenvalue 0 and J_1 contains the remaining Jordan blocks. Since (2.4) holds for A it also holds for J_0 , so J_0 has a p th root W_0 , and W_0 can be taken real in view of the construction given in [33, Section 3]. Form a p th root W_1 of J_1 by taking a p th root of each constituent Jordan block in such a way that every nonreal root has a matching complex conjugate—something that is possible because the Jordan blocks for nonreal eigenvalues occur in complex conjugate pairs since A is real, while if p is even the Jordan blocks of A for negative eigenvalues occur in pairs, by assumption. Then, with $W = \text{diag}(W_0, W_1)$, we have $W^p = J$. Since the Jordan blocks of W occur in complex conjugate pairs it is similar to a real matrix, Y . With \sim denoting similarity, we have $Y^p \sim W^p = J \sim A$. Since Y^p and A are real and similar, they are similar via a real similarity [18, Section 3.4]. Thus $A = GY^pG^{-1}$ for some real, nonsingular G , which can be rewritten as $A = (GYG^{-1})^p = X^p$, where X is real. \square

The next theorem identifies the number of real primary p th roots of a real matrix.

Theorem 2.4. Let the nonsingular matrix $A \in \mathbb{R}^{n \times n}$ have r_1 distinct positive real eigenvalues, r_2 distinct negative real eigenvalues, and c distinct complex conjugate pairs of eigenvalues. If p is even there are (a) $2^{r_1}p^c$ real primary p th roots if $r_2 = 0$ and (b) no real primary p th roots if $r_2 > 0$. If p is odd there are p^c real primary p th roots.

Proof. By transforming A to real Schur form² R our task reduces to counting the number of real p th roots of the diagonal blocks, since a primary p th root of R has the same quasitriangular structure as

² Here, R is block upper triangular with diagonal blocks either 1×1 or 2×2 , and any 2×2 diagonal blocks have complex conjugate eigenvalues.

R and its off-diagonal blocks are uniquely determined by the diagonal blocks [16, Section 7.2], [34]. Consider a 2×2 diagonal block C , which contains a complex conjugate pair of eigenvalues. Let

$$Z^{-1}CZ = \text{diag}(\lambda, \bar{\lambda}) = \theta I + i\mu K, \quad K = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Then $C = \theta I + \mu W$, where $W = iZKZ^{-1}$, and since $\theta, \mu \in \mathbb{R}$ it follows that $W \in \mathbb{R}^{2 \times 2}$. The real primary p th roots of C are $X = ZDZ^{-1} = Z\text{diag}(\alpha + i\beta, \alpha - i\beta)Z^{-1} = \alpha I + \beta W$, where $(\alpha + i\beta)^p = \theta + i\mu$, since the eigenvalues must occur in complex conjugate pairs. There are p such choices, giving p^c choices in total.

Every real eigenvalue must be mapped to a real p th root, and the count depends on the parity of p . There is obviously no real primary p th root if $r_2 > 0$ and p is even, while for odd p any negative eigenvalue $-\lambda$ must be mapped to $-\lambda^{1/p}$, which gives no freedom. Each positive eigenvalue λ yields two choices $\pm\lambda^{1/p}$ for even p , but only one choice $\lambda^{1/p}$ for odd p . This completes the proof. \square

The next lemma enables us to extend the characterization of p th roots in Theorem 2.1 to singular A . We denote by $\Lambda(A)$ the spectrum of A .

Lemma 2.5. *Let*

$$A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} \in \mathbb{C}^{n \times n},$$

where $\Lambda(A_{11}) \cap \Lambda(A_{22}) = \emptyset$. Then any p th root of A has the form

$$X = \begin{bmatrix} X_{11} & X_{12} \\ 0 & X_{22} \end{bmatrix},$$

where $X_{ii}^p = A_{ii}$, $i = 1, 2$ and X_{12} is the unique solution of the Sylvester equation $A_{11}X_{12} - X_{12}A_{22} = X_{11}A_{12} - A_{12}X_{22}$.

Proof. It is well known (see, e.g., [16, Problem 4.3]) that if W satisfies the Sylvester equation $A_{11}W - WA_{22} = A_{12}$ then

$$D = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix} = \begin{bmatrix} I & -W \\ 0 & I \end{bmatrix}^{-1} \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} \begin{bmatrix} I & -W \\ 0 & I \end{bmatrix} \equiv R^{-1}AR.$$

The Sylvester equation has a unique solution since A_{11} and A_{22} have no eigenvalue in common. It is easy to see that any p th root of $A = RDR^{-1}$ has the form $X = RYR^{-1}$, where $Y^p = D$. To characterize all such Y we partition Y conformably with D and equate the off-diagonal blocks in $YD = DY$ to obtain the nonsingular Sylvester equations $Y_{12}A_{22} - A_{11}Y_{12} = 0$ and $Y_{21}A_{11} - A_{22}Y_{21} = 0$, which yield $Y_{12} = 0$ and $Y_{21} = 0$, from which $Y_{ii}^p = A_{ii}$, $i = 1, 2$, follows. Therefore

$$X = RYR^{-1} = \begin{bmatrix} I & -W \\ 0 & I \end{bmatrix} \text{diag}(Y_{11}, Y_{22}) \begin{bmatrix} I & -W \\ 0 & I \end{bmatrix}^{-1} = \begin{bmatrix} Y_{11} & Y_{11}W - WY_{22} \\ 0 & Y_{22} \end{bmatrix}.$$

The Sylvester equation for X_{12} follows by equating the off-diagonal blocks in $XA = AX$, and again this equation is nonsingular. \square

We can now extend Theorem 2.1 to possibly singular matrices.

Theorem 2.6 (classification of p th roots). *Let $A \in \mathbb{C}^{n \times n}$ have the Jordan canonical form $Z^{-1}AZ = J = \text{diag}(J_0, J_1)$, where J_0 collects together all the Jordan blocks corresponding to the eigenvalue 0 and J_1 contains the remaining Jordan blocks. Assume that A satisfies the condition of Theorem 2.2. All p th roots of A are given by $A = Z \text{diag}(X_0, X_1)Z^{-1}$, where X_1 is any p th root of J_1 , characterized by Theorem 2.1, and X_0 is any p th root of J_0 .*

Proof. Since A satisfies the condition of Theorem 2.2, J_0 does as well. It suffices to note that by Lemma 2.5 any p th root of J has the form $\text{diag}(X_0, X_1)$, where $X_0^p = J_0$ and $X_1^p = J_1$. \square

Among all p th roots the principal p th root is the most used in theory and in practice. For $A \in \mathbb{C}^{n \times n}$ with no eigenvalues on \mathbb{R}^- , the closed negative real axis, the principal p th root, written $A^{1/p}$, is the unique p th root of A all of whose eigenvalues lie in the segment $\{z : -\pi/p < \arg(z) < \pi/p\}$ [16, Theorem 7.2]. It is a primary matrix function and it is real when A is real.

3. p th roots of stochastic matrices

We now focus on p th roots of stochastic matrices, and in particular the question of the existence of stochastic roots.

In the next two theorems we recall some key facts from Perron–Frobenius theory that will be needed below [1, Chapter 2], [18, Chapter 8], [26, Chapter 15]. Recall that $A \in \mathbb{R}^{n \times n}$ is reducible if there is a permutation matrix P such that

$$P^T A P = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}, \tag{3.1}$$

where A_{11} and A_{22} are square, nonempty submatrices. A is irreducible if it is not reducible. We write $X \geq 0$ ($X > 0$) to denote that the elements of X are all nonnegative (positive), and denote by $\rho(A)$ the spectral radius of A , $e = [1, 1, \dots, 1]^T$ the vector of 1s, and e_k the unit vector with 1 in the k th position and zeros elsewhere.

Theorem 3.1 (Perron–Frobenius). *If $A \in \mathbb{R}^{n \times n}$ is nonnegative then $\rho(A)$ is an eigenvalue of A with a corresponding nonnegative eigenvector. If, in addition, A is irreducible then*

- (a) $\rho(A) > 0$;
- (b) there is an $x > 0$ such that $Ax = \rho(A)x$;
- (c) $\rho(A)$ is a simple eigenvalue of A (that is, it has algebraic multiplicity 1).

Theorem 3.2. *Let $A \in \mathbb{R}^{n \times n}$ be stochastic. Then*

- (a) $\rho(A) = 1$;
- (b) 1 is a semisimple eigenvalue of A (that is, it appears only in 1×1 Jordan blocks in the Jordan canonical form of A) and has a corresponding eigenvector e ;
- (c) if A is irreducible, then 1 is a simple eigenvalue of A .

Proof. The first part is straightforward. The semisimplicity of the eigenvalue 1 is proved by Minc [30, Chapter 6, Theorem 1.3], while the last part follows from Theorem 3.1. \square

For a p th root X of a stochastic A to be stochastic there are two requirements: that X is nonnegative and that $Xe = e$. While $X^p = A$ and $X \geq 0$ together imply that $\rho(X) = 1$ is an eigenvalue of X with a corresponding nonnegative eigenvector v (by Theorem 3.1), it does not follow that $v = e$. The matrices A and X in Fact 4.11 below provide an example, with $v = [1, 1, 2^{1/2}]^T$. The next result shows that a sufficient condition for a p th root of a stochastic matrix to have unit row sums is that every copy of the eigenvalue 1 of A is mapped to an eigenvalue 1 of X .

Lemma 3.3. *Let $A \in \mathbb{R}^{n \times n}$ be stochastic and let $X^p = A$, where for any eigenvalue μ of X with $\mu^p = 1$ it holds that $\mu = 1$. Then $Xe = e$.*

Proof. Since A is stochastic and so has 1 as a semisimple eigenvalue with corresponding eigenvector e , it has the Jordan canonical form $A = ZJZ^{-1}$ with $J = \text{diag}(I, J_2, J_0)$, where $1 \notin \Lambda(J_2)$, $J_0 \in \mathbb{C}^{k \times k}$ contains all the Jordan blocks corresponding to zero eigenvalues, and $Ze_1 = e$. By Theorem 2.6 any p th root X of A satisfying the assumption of the lemma has the form $X = ZULU^{-1}Z^{-1}$, where $L = \text{diag}(I, L_2, Y_0)$ with $Y_0^p = J_0$, and where $U = \text{diag}(\tilde{U}, I_k)$ with \tilde{U} an arbitrary nonsingular matrix that commutes with $\text{diag}(I, J_2)$ and hence is of the form $\tilde{U} = \text{diag}(\tilde{U}_1, \tilde{U}_2)$. Then

$$Xe = ZULU^{-1}Z^{-1}e = ZULU^{-1}e_1 = Z\text{diag}(I, \tilde{U}_2 L_2 \tilde{U}_2^{-1}, Y_0)e_1 = Ze_1 = e,$$

as required. \square

The sufficient condition of the lemma for X to have unit row sums is not necessary, as the example $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, p = 2$, shows. However, for primary roots the condition is necessary, since every copy of the eigenvalue 1 is mapped to the same root ξ , and $Xe = \xi e$ (which can be proved using the property $f(ZJZ^{-1}) = Zf(J)Z^{-1}$ of primary matrix functions f), so we need $\xi = 1$. The condition is also necessary when A is irreducible, as the next corollary shows.

Corollary 3.4. *Let $A \in \mathbb{R}^{n \times n}$ be an irreducible stochastic matrix. Then for any nonnegative X with $X^p = A, Xe = e$.*

Proof. Since A is stochastic and irreducible, 1 is a simple eigenvalue of A , by Theorem 3.2. As noted just before Lemma 3.3, $X^p = A$ and $X \geq 0$ imply that $\rho(X) = 1$ is an eigenvalue of X , and this is the only eigenvalue μ of X with $\mu^p = 1$, since 1 is a simple eigenvalue of A . Therefore the condition of Lemma 3.3 is satisfied. \square

The next result shows an important connection between stochastic roots of stochastic matrices and nonnegative roots of irreducible nonnegative matrices.

Theorem 3.5. *Suppose C is an irreducible nonnegative matrix with positive eigenvector x corresponding to the eigenvalue $\rho(C)$. Then $A = \rho(C)^{-1}D^{-1}CD$ is stochastic, where $D = \text{diag}(x)$. Moreover, if $C = Y^p$ with Y nonnegative then $A = X^p$, where $X = \rho(C)^{-1/p}D^{-1}YD$ is stochastic.*

Proof. The eigenvector x necessarily has positive elements in view of the fact that C is irreducible and nonnegative, by Theorem 3.1. The stochasticity of A is standard (see [30, Chapter 6, Theorem 1.2], for example), and can be seen from the observation that, since $De = x, Ae = \rho(C)^{-1}D^{-1}Cx = \rho(C)^{-1}D^{-1}\rho(C)x = e$. We have $X^p = \rho(C)^{-1}D^{-1}Y^pD = \rho(C)^{-1}D^{-1}CD = A$. Finally, the irreducibility of C implies that of A , and hence the nonnegative matrix X has unit row sums, by Corollary 3.4. \square

We can identify an interesting class of stochastic matrices for which a stochastic p th root exists for all p . Recall that $A \in \mathbb{R}^{n \times n}$ is a nonsingular M -matrix if $A = sI - B$ with $B \geq 0$ and $s > \rho(B)$. It is a standard property that the inverse of a nonsingular M -matrix is nonnegative [1, Chapter 6].

Theorem 3.6. *If the stochastic matrix $A \in \mathbb{R}^{n \times n}$ is the inverse of an M -matrix then $A^{1/p}$ exists and is stochastic for all p .*

Proof. Since $M = A^{-1}$ is an M -matrix, the eigenvalues of M all have positive real part and hence $M^{1/p}$ exists. Furthermore, $M^{1/p}$ is also an M -matrix for all p , by a result of Fiedler and Schneider [8]. Thus $A^{1/p} = (M^{1/p})^{-1} \geq 0$ for all p , and $A^{1/p}e = e$ follows from the comments following Lemma 3.3, so $A^{1/p}$ is stochastic. \square

If $A \geq 0$ and we can compute $B = A^{-1}$ then it is straightforward to check whether B is an M -matrix: we just have to check whether $b_{ij} \leq 0$ for all $i \neq j$ [1, Chapter 6]. An example of a stochastic inverse M -matrix is given in Fact 4.8 below. Another example is the lower triangular matrix

$$A = \begin{bmatrix} 1 & & & \\ \frac{1}{2} & \frac{1}{2} & & \\ \vdots & \vdots & \ddots & \\ \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \end{bmatrix}, \tag{3.2}$$

for which

$$A^{-1} = \begin{bmatrix} 1 & & & & & \\ -1 & 2 & & & & \\ 0 & -2 & 3 & & & \\ \vdots & \vdots & \ddots & \ddots & & \\ 0 & 0 & \cdots & -(n-1) & n & \end{bmatrix}.$$

Clearly, A^{-1} is an M -matrix and hence from Theorem 3.6, $A^{1/p}$ is stochastic for any positive integer p .

A particular class of inverse M -matrices is the strictly ultrametric matrices, which are the symmetric positive semidefinite matrices for which $a_{ij} \geq \min(a_{ik}, a_{kj})$ for all i, j, k and $a_{ii} > \min\{a_{ik} : k \neq i\}$ (or, if $n = 1, a_{11} > 0$). The inverse of such a matrix is a strictly diagonally dominant M -matrix [29,31].

Using a construction of Soules [35] (also given in a different form by Perfect and Mirsky [32, Theorem 8]), a class of symmetric positive semidefinite stochastic matrices with stochastic roots can be built explicitly.

Theorem 3.7. Let $Q \in \mathbb{R}^{n \times n}$ be an orthogonal matrix with first column $n^{-1/2}e$, $q_{ij} > 0$ for $i + j < n + 2$, $q_{ij} < 0$ for $i + j = n + 2$, and $q_{ij} = 0$ for $i + j > n + 2$. If $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n, \lambda_1 > 0$, and

$$\frac{1}{n}\lambda_1 + \frac{1}{n(n-1)}\lambda_2 + \frac{1}{(n-1)(n-2)}\lambda_3 + \dots + \frac{1}{1 \cdot 2}\lambda_n \geq 0 \tag{3.3}$$

then

- (a) $A = \lambda_1^{-1}Q \text{diag}(\lambda_1, \dots, \lambda_n)Q^T$ is a symmetric stochastic matrix;
- (b) if $\lambda_1 > \lambda_2$ then $A > 0$;
- (c) if $\lambda_n \geq 0$ then $A^{1/p}$ is stochastic for all p .

Proof. (a) is proved by Soules [35, Corollary 2.4]. (b) is shown by Elsner et al. [7, p. 327]. To show (c), if $\lambda_n \geq 0$ then $\lambda_1^{1/p} \geq \lambda_2^{1/p} \geq \dots \geq \lambda_n^{1/p}$ holds and (3.3) trivially remains true with λ_i replaced by $\lambda_i^{1/p}$ for all i and so $A^{1/p}$ is stochastic by (a). \square

A family of matrices Q of the form specified in the theorem can be constructed as a product of Givens rotations G_{ij} , where G_{ij} is a rotation in the (i, j) plane designed to zero the j th element of the vector it premultiplies and produce a nonnegative i th element. Choose rotations G_{ij} so that

$$Ge := G_{12}G_{23} \cdots G_{n-1,n}e = n^{1/2}e_1.$$

Then G has positive elements on and above the diagonal, negative elements on the first subdiagonal, and zeros everywhere else. We have $G^T e_1 = n^{-1/2}e$, and defining Q as G^T with the order of its rows reversed yields a Q of the desired form. For example, for $n = 4$,

$$Q = \begin{bmatrix} 0.5000 & 0.2887 & 0.4082 & 0.7071 \\ 0.5000 & 0.2887 & 0.4082 & -0.7071 \\ 0.5000 & 0.2887 & -0.8165 & 0 \\ 0.5000 & -0.8660 & 0 & 0 \end{bmatrix}.$$

There is a close relation between Theorems 3.6 and 3.7. If $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n > 0$ in Theorem 3.7 then A in Theorem 3.7 has the property that A^{-1} is an M -matrix and, moreover, A^{-k} is an M -matrix for all positive integers k [7, Corollary 2.4].

It is possible to generalize Theorem 3.7 to nonsymmetric stochastic matrices with positive real eigenvalues (using [5, Section 3], for example) but we will not pursue this here.

Finally, we note some more specific results. He and Gunn [15] explicitly find all p th roots of 2×2 stochastic matrices and show how to construct all primary p th roots of 3×3 stochastic matrices.

Marcus and Minc [28] give a sufficient condition for the principal square root of a symmetric positive semidefinite matrix to be stochastic.

Theorem 3.8. *Let $A \in \mathbb{R}^{n \times n}$ be a symmetric positive semidefinite stochastic matrix with $a_{ii} \leq 1/(n - 1), i = 1: n$. Then $A^{1/2}$ is stochastic.*

Proof. See [28, Theorem 2] or [30, Chapter 5, Theorem 4.2]. \square

4. Scenarios for existence and uniqueness of stochastic roots

Existence and uniqueness of p th roots under the requirement of preserving stochastic structure is not a straightforward matter. We present a sequence of facts that demonstrate the wide variety of possible scenarios. In particular, we show that if the principal p th root is not stochastic there may still be a primary stochastic p th root, and if there is no primary stochastic p th root there may still be a nonprimary stochastic p th root.

Fact 4.1. *A stochastic matrix may have no p th root for any p . Consider the stochastic matrix $A = J_n(0) + e_n e_n^T \in \mathbb{R}^{n \times n}$, where $J_n(0), n > 2$, is an $n \times n$ Jordan block with eigenvalue 0. The ascent sequence (2.4) is easily seen to be $n - 1$ 1s followed by zeros. Hence by Theorem 2.2, A has no p th root for any $p > 1$.*

Fact 4.2. *A stochastic matrix may have p th roots but no stochastic p th root. This is true for even p because if A is nonsingular and has some negative eigenvalues then it has p th roots but may have no real p th roots, by Theorem 2.3. An example illustrating this fact is the stochastic matrix*

$$A = \begin{bmatrix} 0.5000 & 0.3750 & 0.1250 \\ 0.7500 & 0.1250 & 0.1250 \\ 0.0833 & 0.0417 & 0.8750 \end{bmatrix}, \quad \Lambda(A) = \{1, 3/4, -1/4\},$$

which has p th roots for all p but no real p th roots for any even p .

Fact 4.3. *A stochastic matrix may have a stochastic principal p th root as well as a stochastic nonprimary p th root. Consider the family of 3×3 stochastic matrices [25]*

$$X(p, x) = \begin{bmatrix} 0 & p & 1 - p \\ x & 0 & 1 - x \\ 0 & 0 & 1 \end{bmatrix},$$

where $0 < p < 1$ and $0 < x < 1$, and let $a = px$. The eigenvalues of $X(p, x)$ are $1, a^{1/2}$, and $-a^{1/2}$. The matrix

$$A = X(p, x)^2 = \begin{bmatrix} a & 0 & 1 - a \\ 0 & a & 1 - a \\ 0 & 0 & 1 \end{bmatrix}$$

is stochastic. But there is another stochastic matrix \tilde{X} that is also a square root of A :

$$\tilde{X} = \begin{bmatrix} a^{1/2} & 0 & 1 - a^{1/2} \\ 0 & a^{1/2} & 1 - a^{1/2} \\ 0 & 0 & 1 \end{bmatrix}.$$

Note that \tilde{X} is the principal square root of A (and hence a primary square root) while all members of the family $X(p, x)$ are nonprimary, since A is upper triangular but the $X(p, x)$ are not.

Fact 4.4. *A stochastic matrix may have a stochastic principal p th root but no other stochastic p th root.*

The matrix (3.2) provides an example.

Fact 4.5. *The principal pth root of a stochastic matrix with distinct, real, positive eigenvalues is not necessarily stochastic.*

This fact is easily verified experimentally. For a parametrized example, let

$$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & \beta \end{bmatrix}, \quad P = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 0 \end{bmatrix}, \quad 0 < \alpha, \beta < 1.$$

Then the matrix

$$X = PDP^{-1} = \frac{1}{4} \begin{bmatrix} 1 + \alpha + 2\beta & 1 + \alpha - 2\beta & 2 - 2\alpha \\ 1 + \alpha - 2\beta & 1 + \alpha + 2\beta & 2 - 2\alpha \\ 1 - \alpha & 1 - \alpha & 2 + 2\alpha \end{bmatrix} \tag{4.1}$$

has unit row sums, and $A = PD^2P^{-1}$ can be obtained by replacing α, β in (4.1) with α^2, β^2 , respectively. Clearly, X is nonnegative if and only if $\beta \leq (1 + \alpha)/2$ while A is nonnegative if and only if $\beta \leq ((1 + \alpha^2)/2)^{1/2}$. If we let $(1 + \alpha)/2 < \beta \leq ((1 + \alpha^2)/2)^{1/2}$ then A is stochastic and its principal square root $X = A^{1/2}$ is not nonnegative; moreover, for $\alpha = 0.5, \beta = 0.751$ (for example) it can be verified that none of the eight square roots of A is stochastic.

Fact 4.6. *A (row) diagonally dominant stochastic matrix (one for which $a_{ii} \geq \sum_{j \neq i} a_{ij}$ for all i) may not have a stochastic principal pth root.*

The matrix A of the previous example serves to illustrate this fact. For $\alpha = 0.99, \beta = 0.9501$,

$$A = \begin{bmatrix} 9.9005 \times 10^{-1} & 9.9005 \times 10^{-7} & 9.9500 \times 10^{-3} \\ 9.9005 \times 10^{-7} & 9.9005 \times 10^{-1} & 9.9500 \times 10^{-3} \\ 4.9750 \times 10^{-3} & 4.9750 \times 10^{-3} & 9.9005 \times 10^{-1} \end{bmatrix}, \tag{4.2}$$

which has strongly dominant diagonal. Yet none of the eight square roots of A is nonnegative.

Fact 4.7. *A stochastic matrix whose principal pth root is not stochastic may still have a primary stochastic pth root.* This fact can be seen from the permutation matrices

$$X = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} = X^2. \tag{4.3}$$

The eigenvalues of A are distinct (they are $-\frac{1}{2} \pm \frac{\sqrt{3}}{2}i$ and 1), so all roots are primary. The matrix X , which is not the principal square root (X has the same eigenvalues as A), is easily checked to be the only stochastic square root of A .

Fact 4.8. *A stochastic matrix with distinct eigenvalues may have a stochastic principal pth root and a different stochastic primary pth root.* As noted in [16, Problem 1.31], the symmetric positive definite matrix M with $m_{ij} = \min(i, j)$ has a square root Y with

$$y_{ij} = \begin{cases} 0, & i + j \leq n, \\ 1, & i + j > n. \end{cases}$$

For example,

$$\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}^2 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 \\ 1 & 2 & 3 & 3 \\ 1 & 2 & 3 & 4 \end{bmatrix}.$$

It is also known that the eigenvalues of M are $\lambda_k = (1/4)\sec(k\pi/(2n + 1))^2, k = 1: n$, so $\rho(M) = (1/4)\sec(n\pi/(2n + 1))^2 =: r_n$ [9]. Since M has all positive elements it has a positive eigenvector x corresponding to $\rho(M)$ (the Perron vector), and so we can apply Theorem 3.5 to deduce that the stochastic matrix $A = r_n^{-1}D^{-1}MD$, where $D = \text{diag}(x)$, has stochastic square root $X = r_n^{-1/2}D^{-1}YD$, and X obviously has the same anti-triangular structure as Y . Since X is clearly indefinite, it is not the principal square root. However, since the eigenvalues of M , and hence A , are distinct, all the square roots of A are primary square roots. The stochastic square root X has $\lceil n/2 \rceil$ positive eigenvalues and $\lfloor n/2 \rfloor$ negative eigenvalues, which follows from the inertia properties of a 2×2 block symmetric matrix—see, for example, Higham and Cheng [17, Theorem 2.1]. However, X is not the only stochastic square root of A , as we now show.

Lemma 4.9. *The principal p th root of $A = r_n^{-1}D^{-1}MD$ is stochastic for all p .*

Proof. Because the row sums are preserved by the principal p th root, we just have to show that $A^{1/p}$ is nonnegative, or equivalently that $M^{1/p}$ is nonnegative. It is known that M^{-1} is the tridiagonal second difference matrix with typical row $[-1 \quad 2 \quad -1]$, except that the (n, n) element is 1. Since M^{-1} has nonpositive off-diagonal elements and M is nonnegative, M^{-1} is an M -matrix and it follows from Theorem 3.6 that $M^{1/p}$ is stochastic for all p . \square

For $n = 4$, A and its two stochastic square roots are

$$\begin{aligned} \begin{bmatrix} 0.1206 & 0.2267 & 0.3054 & 0.3473 \\ 0.0642 & 0.2412 & 0.3250 & 0.3696 \\ 0.0476 & 0.1790 & 0.3618 & 0.4115 \\ 0.0419 & 0.1575 & 0.3182 & 0.4825 \end{bmatrix} &= \begin{bmatrix} 0 & 0 & 0 & 1.0000 \\ 0 & 0 & 0.4679 & 0.5321 \\ 0 & 0.2578 & 0.3473 & 0.3949 \\ 0.1206 & 0.2267 & 0.3054 & 0.3473 \end{bmatrix}^2 \\ &= \begin{bmatrix} 0.2994 & 0.2397 & 0.2315 & 0.2294 \\ 0.0679 & 0.3908 & 0.2792 & 0.2621 \\ 0.0361 & 0.1538 & 0.4705 & 0.3396 \\ 0.0277 & 0.1117 & 0.2626 & 0.5980 \end{bmatrix}^2. \end{aligned}$$

Fact 4.10. *A stochastic matrix without primary stochastic p th roots may have nonprimary stochastic p th roots. Consider the circulant stochastic matrix*

$$A = \frac{1}{3} \begin{bmatrix} 1 - 2a & 1 + a & 1 + a \\ 1 + a & 1 - 2a & 1 + a \\ 1 + a & 1 + a & 1 - 2a \end{bmatrix}, \quad 0 < a \leq \frac{1}{3}.$$

The eigenvalues of A are $1, -a, -a$. The four primary square roots X of A are all nonreal, because in each case the two negative eigenvalues $-a$ and $-a$ are mapped to the same square root, which means that X cannot have complex conjugate eigenvalues. With $\omega = e^{-2\pi i/3}$, we have

$$A = Q^{-1}DQ, \quad Q = \begin{bmatrix} 1 & 1 & 1 \\ 1 & \omega & \omega^2 \\ 1 & \omega^2 & \omega \end{bmatrix}, \quad D = \text{diag}(1, -a, -a).$$

Let $X = Q^{-1} \text{diag}(1, ia^{1/2}, -ia^{1/2})Q$. Then

$$X = \frac{1}{3} \begin{bmatrix} 1 & 1 + (3a)^{1/2} & 1 - (3a)^{1/2} \\ 1 - (3a)^{1/2} & 1 & 1 + (3a)^{1/2} \\ 1 + (3a)^{1/2} & 1 - (3a)^{1/2} & 1 \end{bmatrix},$$

which is a stochastic, nonprimary square root of A .

Fact 4.11. *A nonnegative p th root of a stochastic matrix is not necessarily stochastic. Consider the nonnegative but non-stochastic matrix [27]*

$$X = \begin{bmatrix} 0 & 0 & 2^{-1/2} \\ 0 & 0 & 2^{-1/2} \\ 2^{-1/2} & 2^{-1/2} & 0 \end{bmatrix}, \quad \Lambda(X) = \{1, 0, -1\},$$

for which

$$A = X^{2k} \equiv \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \Lambda(A) = \{1, 1, 0\}$$

is stochastic. Note that A is its own stochastic p th root for any integer p .

Fact 4.12. *A stochastic matrix may have a stochastic p th root for some, but not all, p .*

Consider again the matrix

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

appearing in Fact 4.7. We have $A^3 = I$, which implies $A^{3k+1} = A$ and $(A^2)^{3k+2} = A^4 = A$ for all non-negative integers k . Hence A is its own stochastic p th root for $p = 3k + 1$ and A^2 is a stochastic p th root of A for $p = 3k + 2$. However, $\Lambda(A) = \{1, \omega, \bar{\omega}\}$ with $\omega = e^{-2\pi i/3}$, and the arguments in Section 5 show that A has no stochastic cube root (since ω lies outside the region Θ_3^3 in Fig. 5.2). Hence A does not have a stochastic root for $p = 3k$.

The stochastic root problem is intimately related to the embeddability problem in discrete-time Markov chains, which asks when a nonsingular stochastic matrix A can be written $A = e^Q$ for some Q with $q_{ij} \geq 0$ for $i \neq j$ and $\sum_j q_{ij} = 0, i = 1: n$. (For background on the embeddability problem see Davies [6] or Higham [16, Section 2.3] and the references therein.) Kingman [24, Proposition 7] shows that this condition holds if and only if for every positive integer p there exists some stochastic X_p such that $A = X_p^p$. (Thus the matrices identified in Theorems 3.6 and 3.7 form two classes of embeddable matrices.) The condition that A is embeddable is therefore much stronger than the condition that A has a stochastic p th root for a particular p . This is emphasized by the following facts, which show that certain necessary conditions derived in the literature for A to be embeddable are not necessary for A to have a stochastic p th root for certain p . Moreover, A may of course be singular in the stochastic root problem, in which case it cannot be the exponential of any matrix.

Fact 4.13. *$\det(A) > 0$ is (obviously) necessary for the embeddability of a stochastic matrix A ; it is also necessary for the existence of a stochastic p th root when p is even, but it is not necessary when p is odd.*

The matrix

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

has $\det(A) = -1$, but A is its own stochastic p th root for any odd p .

Fact 4.14. *$\det(A) \leq \prod_i a_{ii}$ is necessary for the embeddability of A [10, Theorem 6.1], but it is not necessary for the existence of a stochastic p th root. For example, let A be the matrix X in (4.3). Then $A^3 = I$, so $(A^2)^2 = A$ and A has a stochastic square root, but $\det(A) = 1 > 0 = a_{11}a_{22}a_{33}$.*

Fact 4.15. *If there is a sequence $k_0 = i, k_1, k_2, \dots, k_m = j$ such that $a_{k_\ell k_{\ell+1}} > 0$ for each ℓ but $a_{ij} = 0$ then A is not embeddable [12, Section 6.10], but it is still possible for A to have a stochastic p th root for some p . See the matrix A in (4.3), which has a stochastic root but for which $a_{12} > 0$ and $a_{23} > 0$, while $a_{13} = 0$.*

In both the embeddability problem and the stochastic root problem it is difficult to identify conditions that guarantee the existence of a logarithm or root of the required form. For some further insight, consider a nonsingular upper triangular stochastic matrix T , taking $n = 3$ for simplicity. The equation $U^2 = T$ can be solved for U (assumed upper triangular) a diagonal at a time by a recurrence of Björck and Hammarling [3], [16, Section 6.2]. This gives $u_{ii} = t_{ii}^{1/2}$, $i = 1:3$ (since we require U nonnegative), $u_{i,i+1} = t_{i,i+1}/(u_{ii} + u_{i+1,i+1})$, $i = 1:2$, and $u_{13} = (t_{13} - u_{12}u_{23})/(t_{11}^{1/2} + t_{33}^{1/2})$. Hence $u_{13} \geq 0$ when

$$t_{13} - \frac{t_{12}t_{23}}{(t_{11}^{1/2} + t_{22}^{1/2})(t_{22}^{1/2} + t_{33}^{1/2})} \geq 0.$$

If we assume that T is diagonally dominant, which implies $t_{ii} \geq 1/2$, $i = 1, 2$, and note that $t_{33} = 1$, we obtain the sufficient condition for nonnegativity that $(1 + 2^{1/2})t_{13} \geq t_{12}t_{23}$. But diagonal dominance alone is not sufficient to ensure nonnegativity. Thus even for diagonally dominant triangular matrices the stochasticity of the principal square root depends in a complicated way on the relationships between the matrix entries.

We can also consider general strictly diagonally dominant stochastic matrices, for which $a_{ii} > 1/2$ for all i . Let $m = \min_i a_{ii}$ and write $A = mI + E$. Then $E \geq 0$ and $Ee = (A - mI)e = (1 - m)e$, so $\|E\|_\infty = 1 - m$. Hence we can write $A = m(I + F)$, where $\|F\|_\infty = \|E\|_\infty/m = (1 - m)/m < 1$. Then the principal p th root can be expressed as

$$A^{1/p} = m^{1/p}(I + F)^{1/p} = m^{1/p} \left(I + \frac{1}{p}F + \frac{1}{2!} \frac{1}{p} \left(\frac{1}{p} - 1 \right) F^2 + \dots \right).$$

Unfortunately, it is difficult to obtain from this expansion useful sufficient conditions for $A^{1/p} \geq 0$. Nonnegativity is guaranteed if all the off-diagonal elements of F are positive and $\|F\|_\infty$ is sufficiently small, but as the matrix (4.2) shows, “small” here may have to be very small.

5. Necessary conditions based on inverse eigenvalue problem

Karpelevič [23] has determined the set Θ_n of all eigenvalues of all stochastic $n \times n$ matrices. This set provides the solution to the inverse eigenvalue problem for stochastic matrices, which asks when a given complex scalar is the eigenvalue of some $n \times n$ stochastic matrix. (Note the distinction with the problem of determining conditions under which a set of n complex numbers comprises the eigenvalues of some $n \times n$ stochastic matrix, which is called the inverse spectrum problem by Minc [30].)

The following theorem gives the main points of Karpelevič’s theorem on the characterization of Θ_n ; full details on the “specific rules” mentioned therein can be found in [23] and [30, Chapter 7, Theorem 1.8].

Theorem 5.1. *The set Θ_n is contained in the unit disk and is symmetric with respect to the real axis. It intersects the unit circle at points $e^{2i\pi a/b}$ where a and b range over all integers such that $0 \leq a < b \leq n$. For $n > 3$, the boundary of Θ_n consists of curvilinear arcs connecting these points in circular order. Any point λ on these arcs must satisfy one of the parametric equations*

$$\lambda^q(\lambda^s - t)^r = (1 - t)^r, \tag{5.1}$$

$$(\lambda^b - t)^d = (1 - t)^d \lambda^q, \tag{5.2}$$

where $0 \leq t \leq 1$, and b, d, q, s, r are positive integers determined from certain specific rules.

The set Θ_3 of eigenvalues of 3×3 stochastic matrices consists of points in the interior and on the boundary of an equilateral triangle of maximal size inscribed in the unit circle with one of its vertices at the point $(1,0)$, as well as all points on the segment $[-1, 1]$; see Fig. 5.1. The boundary of Θ_4 consists of curvilinear arcs determined by the parametric equations $\lambda^3 + \lambda^2 + \lambda + t = 0$ and $\lambda^3 + \lambda^2 - (2t - 1)\lambda - t^2 = 0$, $0 \leq t \leq 1$, together with line segments linking $(1,0)$ with $(0,1)$, and $(1,0)$ with $(0, -1)$, respectively, as can also be seen in Fig. 5.1.

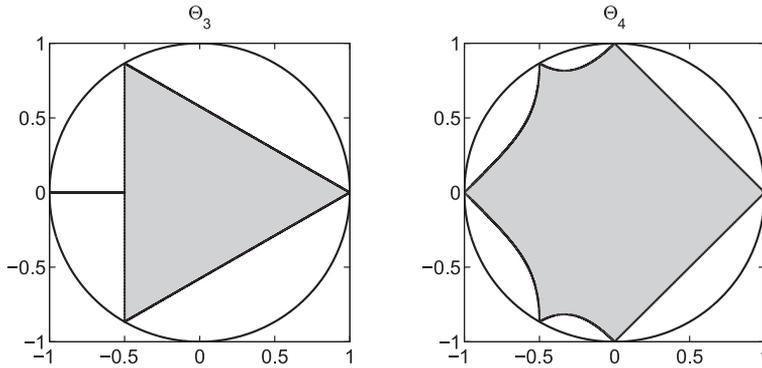


Fig. 5.1. The sets Θ_3 and Θ_4 of all eigenvalues of 3×3 and 4×4 stochastic matrices, respectively.

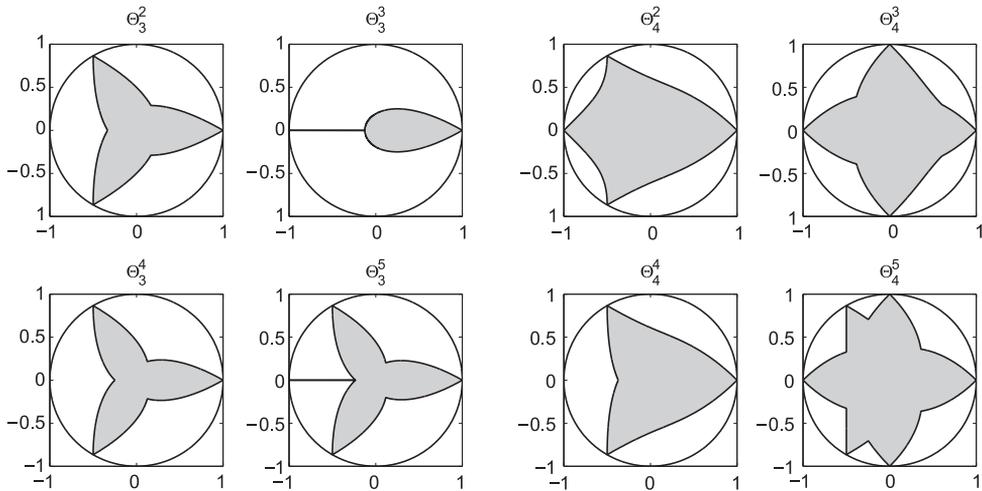


Fig. 5.2. Regions obtained by raising the points in Θ_3 (left) and Θ_4 (right) to the powers 2, 3, 4, and 5.

Denote by Θ_n^p the set of p th powers of points in Θ_n , i.e., $\Theta_n^p = \{\lambda^p : \lambda \in \Theta_n\}$. If A and X are stochastic $n \times n$ matrices such that $X^p = A$ then for any eigenvalue λ of X , λ^p is an eigenvalue of A . Hence, a necessary condition for A to have a stochastic p th root is that all the eigenvalues of A are in the set Θ_n^p . It can be shown that Θ_n^p is a closed set within the unit disk with boundary $\partial\Theta_n^p \subseteq \{\lambda^p : \lambda \in \partial\Theta_n\}$, where $\partial\Theta_n$ is the boundary of Θ_n , the points on which satisfy the parametric equation (5.1) or (5.2). Fig. 5.2 shows the second to fifth powers of Θ_3 and Θ_4 .

This approach provides necessary conditions for A to have a stochastic p th root. The conditions are not sufficient, because we are checking whether each eigenvalue of A is the eigenvalue of some p th power of a stochastic matrix, and not that every eigenvalue of A is an eigenvalue of the p th power of the same stochastic matrix.

To illustrate, consider the stochastic matrix

$$A = \begin{bmatrix} 1/3 & 1/3 & 0 & 1/3 \\ 1/2 & 0 & 1/2 & 0 \\ 10/11 & 0 & 0 & 1/11 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{bmatrix}. \tag{5.3}$$

From Fig. 5.3 we see that A cannot have a stochastic 12th root, but may have a stochastic 52nd root. In fact, both $A^{1/12}$ and $A^{1/52}$ have negative elements and none of the 52nd roots is stochastic.

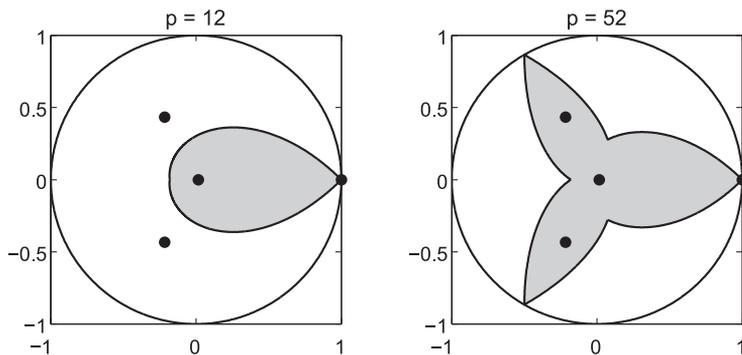


Fig. 5.3. Θ_4^p for $p = 12$ and $p = 52$ and the spectrum (shown as dots) of A in (5.3).

If $A \in \mathbb{R}^{n \times n}$ is stochastic then so is the matrix $\text{diag}(A, 1)$ of order $n + 1$, and it follows that $\Theta_3 \subseteq \Theta_4 \subseteq \Theta_5 \subseteq \dots$. Moreover, the number of points at which the region Θ_n intersects the unit circle increases rapidly with n ; for example, there are 23 intersection points for Θ_8 and 80 for Θ_{16} . As n increases the region Θ_n and its powers tend to fill the unit circle, so the necessary conditions given in this section are most useful for small dimensions. We emphasize, however, that small matrices do arise in practice; for example, in the model in [4] describing the progression to AIDS in an HIV-infected population the transition matrix³ is of dimension 5.

6. Conclusions

The existing literature on roots of stochastic matrices emphasizes computational aspects at the expense of a careful treatment of the underlying theory. We have used the theory of matrix functions to develop tools for analyzing the existence of stochastic roots of stochastic matrices. We have identified two classes of stochastic matrices for which the principal p th root is stochastic for all p . However, such matrices seem rare, and we have demonstrated a wide variety of possibilities for existence and uniqueness, in particular regarding primary versus nonprimary roots. We have also given some necessary spectral conditions for existence. We hope that as well as providing insight into what makes this interesting and practically important problem so difficult our work will prove useful for further development of theory and algorithms.

References

- [1] A. Berman, R.J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1994. ISBN 0 89871 321 8, xx + 340 pp. (Corrected republication, with supplement, of work first published in 1979 by Academic Press).
- [2] D.A. Bini, N.J. Higham, B. Meini, Algorithms for the matrix p th root, *Numer. Algorithms* 39 (4) (2005) 349–378.
- [3] Å. Björck, S. Hammarling, A Schur method for the square root of a matrix, *Linear Algebra Appl.* 52/53 (1983) 127–140.
- [4] T. Charitos, P.R. de Waal, L.C. van der Gaag, Computing short-interval transition matrices of a discrete-time Markov chain from partially observed data, *Stat. Med.* 27 (2008) 905–921.
- [5] M.Q. Chen, L. Han, M. Neumann, On single and double Soules matrices, *Linear Algebra Appl.* 416 (2006) 88–110.
- [6] E.B. Davies, *Embeddable Markov matrices*, 2010. Available from: <http://arxiv.org/abs/1001.1693>.
- [7] L. Elsner, R. Nabben, M. Neumann, Orthogonal bases that lead to symmetric nonnegative matrices, *Linear Algebra Appl.* 271 (1998) 323–343.
- [8] M. Fiedler, H. Schneider, Analytic functions of M -matrices and generalizations, *Linear and Multilinear Algebra* 13 (1983) 185–201.
- [9] J. Fortiana, C.M. Cuadras, A family of matrices, the discretized Brownian bridge, and distance-based regression, *Linear Algebra Appl.* 264 (1997) 173–188.
- [10] G.S. Goodman, An intrinsic time for non-stationary finite Markov chains, *Z. Wahrscheinlichkeitstheorie* 16 (1970) 165–180.
- [11] F. Greco, B. Iannazzo, A binary powering algorithm for computing primary matrix roots, *Numer. Algorithms*, in press.

³ This matrix has one negative eigenvalue and a square root is required; that no exact stochastic square root exists follows from Theorem 2.3.

- [12] G.R. Grimmett, D.R. Stirzaker, *Probability and Random Processes*. Oxford University Press, 2001. ISBN 0 19 857223 9, xii + 596 pp.
- [13] C.-H. Guo, On Newton's method and Halley's method for the principal p th root of a matrix, *Linear Algebra Appl.* 432 (2010) 1905–1922.
- [14] C.-H. Guo, N.J. Higham, A Schur–Newton method for the matrix p th root and its inverse, *SIAM J. Matrix Anal. Appl.* 28 (3) (2006) 788–804.
- [15] Q.-M. He, E. Gunn, A note on the stochastic roots of stochastic matrices, *J. Systems Sci. Systems Eng.* 12 (2) (2003) 210–223.
- [16] N.J. Higham, *Functions of Matrices: Theory and Computation*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2008. ISBN 978 0 898716 46 7, xx + 425 pp.
- [17] N.J. Higham, S.H. Cheng, Modifying the inertia of matrices arising in optimization, *Linear Algebra Appl.* 275–276 (1998) 261–279.
- [18] R.A. Horn, C.R. Johnson, *Matrix Analysis*. Cambridge University Press, Cambridge, UK, 1985. ISBN 0 521 30586 1, xiii + 561 pp.
- [19] B. Iannazzo, On the Newton method for the matrix p th root, *SIAM J. Matrix Anal. Appl.* 28 (2) (2006) 503–523.
- [20] B. Iannazzo, A family of rational iterations and its application to the computation of the matrix p th root, *SIAM J. Matrix Anal. Appl.* 30 (4) (2009) 1445–1462.
- [21] R.B. Israel, J.S. Rosenthal, J.Z. Wei, Finding generators for Markov chains via empirical transition matrices, with applications to credit ratings, *Math. Finance* 11 (2) (2001) 245–265.
- [22] R.A. Jarrow, D. Lando, S.M. Turnbull, A Markov model for the term structure of credit risk spreads, *Rev. Financial Stud.* 10 (2) (1997) 481–523.
- [23] F. Karpelevič, On the characteristic roots of matrices with nonnegative elements. *Izvestia Akademii Nauk SSSR, Mathematical Series*, vol. 15, 1951, pp. 361–383 (in Russian). English Translation appears in *Amer. Math. Soc. Trans., Series 2*, 140 (1988) 79–100.
- [24] J.F.C. Kingman, The imbedding problem for finite Markov chains, *Z. Wahrscheinlichkeitstheorie* 1 (1962) 14–24.
- [25] A. Kreinin, M. Sidelnikova, Regularization algorithms for transition matrices, *Algo Res. Quart.* 4 (1/2) (2001) 23–40.
- [26] P. Lancaster, M. Tismenetsky, *The Theory of Matrices*, second ed., Academic Press, London, 1985. ISBN 0 12 435560 9, xv + 570 pp.
- [27] D. London, Nonnegative matrices with stochastic powers, *Israel J. Math.* 2 (4) (1964) 237–244.
- [28] M. Marcus, H. Minc, Some results on doubly stochastic matrices, *Proc. Amer. Math. Soc.* 13 (4) (1962) 571–579.
- [29] S. Martínez, G. Michon, J.S. Martín, Inverse of strictly ultrametric matrices are of Stieltjes type, *SIAM J. Matrix Anal. Appl.* 15 (1) (1994) 98–106.
- [30] H. Minc, *Nonnegative Matrices*, Wiley, New York, 1988. ISBN 0 471 83966 3, xiii + 206 pp.
- [31] R. Nabben, R.S. Varga, A linear algebra proof that the inverse of a strictly ultrametric matrix is a strictly diagonally dominant Stieltjes matrix, *SIAM J. Matrix Anal. Appl.* 15 (1) (1994) 107–113.
- [32] H. Perfect, L. Mirsky, Spectral properties of doubly-stochastic matrices, *Monatsh. Math.* 69 (1) (1965) 35–57.
- [33] P.J. Psarrakos, On the m th roots of a complex matrix, *Electron. J. Linear Algebra* 9 (2002) 32–41.
- [34] M.I. Smith, A Schur algorithm for computing matrix p th roots, *SIAM J. Matrix Anal. Appl.* 24 (4) (2003) 971–989.
- [35] G.W. Soules, Constructing symmetric nonnegative matrices, *Linear and Multilinear Algebra* 13 (1983) 241–251.
- [36] G. ten Have, Structure of the n th roots of a matrix, *Linear Algebra Appl.* 187 (1993) 59–66.
- [37] F.V. Waugh, M.E. Abel, On fractional powers of a matrix, *J. Amer. Statist. Assoc.* 62 (1967) 1018–1021.