Roots of Matrices

Nick Higham
School of Mathematics
The University of Manchester

higham@ma.man.ac.uk
http://www.ma.man.ac.uk/~higham/

Joint work with Lijing Lin

Brian Davies 65th Birthday Conference,
December 2009
Matrix $\rho$th Root

- $X$ is a $\rho$th root ($\rho \in \mathbb{Z}^+$) of $A \in \mathbb{C}^{n \times n}$ if $X^\rho = A$.
- Number of $\rho$th roots may be zero, finite or infinite.

**Definition**

For $A \in \mathbb{C}^{n \times n}$ with no eigenvalues on $\mathbb{R}^- = \{ x \in \mathbb{R} : x \leq 0 \}$, the **principal $\rho$th root**, $A^{1/\rho}$, is a unique $\rho$th root $X$ with spectrum in the wedge $|\arg(\lambda(X))| < \pi/\rho$. 
Matrix $p$th Root

- $X$ is a $p$th root ($p \in \mathbb{Z}^+$) of $A \in \mathbb{C}^{n \times n}$ $\iff$ $X^p = A$.
- Number of $p$th roots may be zero, finite or infinite.

**Definition**

For $A \in \mathbb{C}^{n \times n}$ with no eigenvalues on $\mathbb{R}^- = \{x \in \mathbb{R} : x \leq 0\}$ the principal $p$th root, $A^{1/p}$, is unique $p$th root $X$ with spectrum in the wedge $|\arg(\lambda(X))| < \pi/p$.

**Definition**

For $A \in \mathbb{C}^{n \times n}$ with no eigenvalues on $\mathbb{R}^-$ the principal logarithm, $\log(A)$, is unique solution of $e^X = A$ with $|\text{Im} \lambda(X)| < \pi$. 

MIMS

Nick Higham

Roots of Matrices 2 / 37
Definition

For $A \in \mathbb{C}^{n \times n}$ with no eigenvalues on $\mathbb{R}^-$ and $s \in [0, \infty)$, $A^s = e^{s \log A}$, where $\log A$ is the principal logarithm.

$$A^s = \frac{\sin(s\pi)}{s\pi} A \int_0^\infty (t^{1/s} I + A)^{-1} \, dt, \quad s \in (0, 1).$$
Arbitrary Power

**Definition**

For $A \in \mathbb{C}^{n \times n}$ with no eigenvalues on $\mathbb{R}^-$ and $s \in [0, \infty)$, $A^s = e^{s \log A}$, where $\log A$ is the principal logarithm.

$$A^s = \frac{\sin(s\pi)}{s\pi} A \int_0^\infty (t^{1/s} I + A)^{-1} \, dt, \quad s \in (0, 1).$$

**Applications:**

- Pricing American options (Berridge & Schumacher, 2004).
- Finite element discretizations of fractional Sobolev spaces (Arioli & Loghin, 2009).
If $A = XDX^{-1}$, $D = \text{diag}(d_i)$, then $f(A) = Xf(D)X^{-1}$. OK numerically if $X$ is well conditioned.

For any $A$, let $E = \epsilon \text{randn}(n)$, $A + E = XDX^{-1}$. Then (Davies, 2007)

$$f(A) \approx Xf(D)X^{-1}.$$ 

- Especially useful for $A^s$.
- A Test Problem for Computations of Fractional Powers of Matrices (Davies, 2008).
Turnbull (1927): $A_n^3 = I_n$, where

$$A_4 = \begin{bmatrix} -1 & 1 & -1 & 1 \\ -3 & 2 & -1 & 0 \\ -3 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{bmatrix}.$$

$B_n^2 = I_n$, where

$$B_4 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & -1 & -2 & -3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & -1 \end{bmatrix}.$$

Arises in BDF solvers for ODEs.
Bambaii & Chowla (1946): \( B_{n+1}^n = I_n \) where

\[
B_4 = \begin{bmatrix}
-1 & -1 & -1 & -1 \\
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0
\end{bmatrix}.
\]


Bauer (2002): “since then the value of mathematical methods in cryptology has been unchallenged.”

Real square roots of \(-I\):

\[
\begin{bmatrix}
a & 1 + a^2 \\
-1 & -a
\end{bmatrix}^2 = \begin{bmatrix}
-1 & 0 \\
0 & -1
\end{bmatrix}, \quad a \in \mathbb{C}.
\]
Markov Models

- Discrete-time Markov process with transition probability matrix $P$, time unit 1. Unit is 1 year in credit risk modelling.
- Transition matrix for fractional time unit $\alpha$ is $P^\alpha$.
- If $P$ is embeddable, $P = e^Q$ for generator $Q$ with $q_{ij} \geq 0$ ($i \neq j$), $\sum_{j=1}^{n} q_{ij} = 0$. Then $P^\alpha = e^{\alpha Q}$.

Problems:

- $P$ may not be embeddable.
- $P^{1/k}$ may not be a stochastic matrix.
- Is there a stochastic root?
The problem has arisen through proposed methodology on which the company will incur charges for use of an electricity network.

I have the use of a computer and Microsoft Excel.

I have an Excel spreadsheet containing the transition matrix of how a company’s [Standard & Poor’s] credit rating changes from one year to the next. I’d like to be working in eighths of a year, so the aim is to find the eighth root of the matrix.


M. Bladt & M. Sørensen. Efficient estimation of transition rates between credit ratings from observations at discrete time points. *Quantitative Finance, 2009.*
HIV to Aids Transition

- Estimated 6-month transition matrix.
- Four AIDS-free states and 1 AIDS state.
- 2077 observations (Charitos et al., 2008).

\[
P = \begin{bmatrix}
0.8149 & 0.0738 & 0.0586 & 0.0407 & 0.0120 \\
0.5622 & 0.1752 & 0.1314 & 0.1169 & 0.0143 \\
0.3606 & 0.1860 & 0.1521 & 0.2198 & 0.0815 \\
0.1676 & 0.0636 & 0.1444 & 0.4652 & 0.1592 \\
0 & 0 & 0 & 0 & 1
\end{bmatrix}.
\]

Want to estimate the 1-month transition matrix.

\[ \Lambda(P) = \{1, 0.9644, 0.4980, 0.1493, -0.0043\}. \]
Want techniques for evaluating interesting \( f \) at matrix arguments.

Example:

\[
\frac{d^2y}{dt^2} + Ay = 0, \quad y(0) = y_0, \quad y'(0) = y'_0
\]

\[\Rightarrow y(t) = \cos(\sqrt{A}t)y_0 + (\sqrt{A})^{-1}\sin(\sqrt{A}t)y'_0,\]

where \( \sqrt{A} \) is any square root of \( A \).

MATLAB has \texttt{expm}, \texttt{logm}, \texttt{sqrtm}, \texttt{funm} and \( \wedge \).
Visser Iteration for $A^{1/2}$

\[ X_{k+1} = X_k + \alpha (A - X_k^2), \quad X_0 = (2\alpha)^{-1}I. \]

- Used with $\alpha = 1/2$ by Visser (1932) to show positive operator on Hilbert space has a positive square root.
- Enables proof of existence of $A^{1/2}$ without using spectral theorem.
- Likewise in functional analysis texts, e.g. Riesz & Sz.-Nagy (1956).
- Elsner proves cgce for $A \in \mathbb{C}^{n\times n}$ with real, positive eigenvalues if $0 < \alpha \leq \rho(A)^{-1/2}$. 
Visser Convergence

\[ X_{k+1} = X_k + \alpha(A - X_k^2), \quad X_0 = (2\alpha)^{-1}I. \]

**Theorem (H, 2008)**

Let \( A \in \mathbb{C}^{n \times n} \) and \( \alpha > 0 \). If \( \Lambda(I - 4\alpha^2A) \) lies in the cardioid

\[ D = \{ 2z - z^2 : z \in \mathbb{C}, |z| < 1 \} \]

then \( A^{1/2} \) exists and \( X_k \to A^{1/2} \) linearly.
Rice (1982):

\[ X_{k+1} = X_k + \frac{1}{p}(A - X_k^p), \quad X_0 = 0. \]

For Hermitian pos def \( A \), \( 0 \leq X_k \leq X_{k+1} \) for all \( k \) and \( X_k \to A^{1/p} \).
Existence of \( p \)-th Roots

**Theorem (Psarrakos, 2002)**

A \( \in \mathbb{C}^{n \times n} \) has a \( p \)-th root iff for every integer \( \nu \geq 0 \) no more than one element of the **ascent sequence**” \( d_1, d_2, \ldots \) defined by

\[
d_i = \dim(\text{null}(A^i)) - \dim(\text{null}(A^{i-1}))
\]

lies strictly between \( p \nu \) and \( p(\nu + 1) \).

- For \( J = J(0) \in \mathbb{C}^{n \times n} \), \( \dim(\text{null}(J^k)) = k \), \( k = 0: n \), \( \{d_i\} = \{1, 1, \ldots, 1\} \); no \( p \)-th root for \( p \geq 2 \).
Theorem

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

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 pth 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 reduces $A$ to $\text{diag}(A_{11}, A_{22})$. 
Jordan canonical form $Z^{-1}AZ = J = \text{diag}(J_0, J_1)$.

All $p$th roots of $A$ are given by $A = Z\text{diag}(X_0, X_1)Z^{-1}$, where

- $X_1^p = J_1$ (have characterization),
- $X_0^p = J_0$ (no nice characterization).

**History:**

- Cayley (1858, 1872).
- Sylvester (1882, 1883).
A ∈ ℝⁿˣⁿ, A ≥ 0, Ae = e.

**Theorem**

Let $A ∈ ℝⁿˣⁿ$ be stochastic. Then

- $\rho(A) = 1$;
- 1 is a semisimple eigenvalue of $A$ with eigenvector $e$;
- if $A$ is irreducible, then 1 is a simple eigenvalue of $A$. 
Nonneg Root may not be Stochastic

\(X^p = A\) and \(X \geq 0\) imply that \(\rho(X) = \rho(A)^{1/p} = 1\) is an ei’val with ei’vec \(v \geq 0\) (Perron–Frobenius) but *not* that \(v = e\):

\[
A = \begin{bmatrix}
\frac{1}{2} & \frac{1}{2} & 0 \\
\frac{1}{2} & \frac{1}{2} & 0 \\
0 & 0 & 1
\end{bmatrix}, \quad \Lambda(A) = \{1, 1, 0\}.
\]

\(A = X^{2k}\) for

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

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

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

In fact . . .

Theorem

Let $C \geq 0$ be irreducible with $e'$vec $x > 0$ corr. to $\rho(C)$. 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.
Definition of Nonsingular $M$-matrix $A \in \mathbb{R}^{n \times n}$

$A = s\mathbf{I} - B$ with $B \geq 0$ and $s > \rho(B)$. 

### Definition of Nonsingular $M$-matrix $A \in \mathbb{R}^{n \times n}$

$$A = sI - B \text{ with } B \geq 0 \text{ and } s > \rho(B).$$

### Theorem

*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 "$M$", Re $\lambda_i(M) > 0$ so $M^{1/p}$ exists.
- $M^{1/p}$ is an $M$-matrix for all $p$ (Fiedler & Schneider, 1983)
- Thus $A^{1/p} = (M^{1/p})^{-1} \geq 0$ for all $p$, and $A^{1/p}e = e$ (shown via JCF arguments), so $A^{1/p}$ is stochastic.
Example 1

\[ A = \begin{bmatrix}
1 \\
\frac{1}{2} & \frac{1}{2} \\
\vdots & \vdots & \ddots \\
\frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n}
\end{bmatrix}. \]

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

\( A^{-1} \) is an \( M \)-matrix so \( A^{1/p} \) is stochastic for all \( p > 0 \).
Example 2

\[
Y^2 = \begin{bmatrix}
0 & 0 & 0 & 1 \\
0 & 0 & 1 & 1 \\
0 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 \\
\end{bmatrix}^2 = \begin{bmatrix}
1 & 1 & 1 & 1 \\
1 & 2 & 2 & 2 \\
1 & 2 & 3 & 3 \\
1 & 2 & 3 & 4 \\
\end{bmatrix} = M.
\]

\[
\lambda_k(M) = \frac{1}{4} \sec\left(\frac{k\pi}{2n+1}\right)^2, \ k = 1: n.
\]

Positive e’vec \( x \) for \( \rho(M) \).

- \( A = \rho(M)^{-1}D^{-1}MD \) is stochastic, where \( D = \text{diag}(x) \), has stochastic sq. root \( X = \rho(M)^{-1/2}D^{-1}YD \).

- Note: \( X \) is \textit{indefinite}.

- But \( A \) has another stochastic sq. root: \( A^{1/2} \), by previous theorem!
Example 2 cont.

For \( n = 4 \):

\[
\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.
\]
A stochastic matrix may have no pth root for any $p$. 
- A stochastic matrix may have no pth root for any p.
- A stochastic matrix may have pth roots but no stochastic pth root.
A stochastic matrix may have no pth root for any p.

A stochastic matrix may have pth roots but no stochastic pth root.

A stochastic matrix may have a stochastic principal pth root as well as a stochastic nonprimary pth root.
- A stochastic matrix may have no pth root for any $p$.
- A stochastic matrix may have pth roots but no stochastic pth root.
- A stochastic matrix may have a stochastic principal pth root as well as a stochastic nonprimary pth root.
- A stochastic matrix may have a stochastic principal pth root but no other stochastic pth root.
A stochastic matrix may have no pth root for any $p$.

A stochastic matrix may have pth roots but no stochastic pth root.

A stochastic matrix may have a stochastic principal pth root as well as a stochastic nonprimary pth root.

A stochastic matrix may have a stochastic principal pth root but no other stochastic pth root.

The principal pth root of a stochastic matrix with distinct, real, positive eigenvalues is not necessarily stochastic.
A (row) diagonally dominant stochastic matrix may not have a stochastic principal pth root.

\[
A = \begin{bmatrix}
9.9005 \times 10^{-1} & 9.9005 \times 10^{-7} & 9.9500 \times 10^{-3} \\
9.9005 \times 10^{-7} & 9.9005 \times 10^{-1} & 9.9500 \times 10^{-3} \\
4.9750 \times 10^{-3} & 4.9750 \times 10^{-3} & 9.9005 \times 10^{-1}
\end{bmatrix}.
\]

None of the 8 square roots of \( A \) is nonnegative.
A (row) diagonally dominant stochastic matrix may not have a stochastic principal pth root.

\[
A = \begin{bmatrix}
9.9005 \times 10^{-1} & 9.9005 \times 10^{-7} & 9.9500 \times 10^{-3} \\
9.9005 \times 10^{-7} & 9.9005 \times 10^{-1} & 9.9500 \times 10^{-3} \\
4.9750 \times 10^{-3} & 4.9750 \times 10^{-3} & 9.9005 \times 10^{-1}
\end{bmatrix}.
\]

None of the 8 square roots of \(A\) is nonnegative.

A stochastic matrix whose principal pth root is not stochastic may still have a primary stochastic pth root.

\[
A = \begin{bmatrix}
0 & 1 & 0 \\
0 & 0 & 1 \\
1 & 0 & 0
\end{bmatrix} = \begin{bmatrix}
0 & 0 & 1 \\
1 & 0 & 0 \\
0 & 1 & 0
\end{bmatrix}^2 = X^2.
\]

\[
\Lambda(A) = \Lambda(X) = \{e^{\pm 2\pi/3}, 1\}.
\]
When can nonsingular stochastic A be written $A = e^Q$ with $q_{ij} \geq 0$ for $i \neq j$ and $\sum_j q_{ij} = 0$, $i = 1 : n$?

Kingman (1962): holds iff for every positive integer $p$ there exists some stochastic $X$ such that $A = X^p$.

Conditions (e.g.)

- $\det(A) > 0$
- $\det(A) \leq \prod_i a_{ii}$

are necessary for embeddability of a stochastic $A$ but not necessary for existence of a stochastic $p$th root for a particular $p$.

New classes of embeddable matrices.
Karpelevič (1951) determined
\[ \Theta_n = \{ \lambda : \lambda \in \Lambda(A), \ A \in \mathbb{R}^{n \times n} \text{ stochastic} \} . \]

**Theorem**

\[ \Theta_n \subseteq \text{unit disk and intersects unit circle at } e^{2i\pi a/b}, \text{ all } a, b \text{ s.t. } 0 \leq a < b \leq n. \text{ Boundary of } \Theta_n \text{ is curvilinear arcs defined by} \]

\[ \lambda^q (\lambda^s - t)^r = (1 - t)^r, \]
\[ (\lambda^b - t)^d = (1 - t)^d \lambda^q, \]

*where* \( 0 \leq t \leq 1, \text{ and } b, d, q, s, r \in \mathbb{Z}^+ \text{ determined from certain specific rules.} \)
$n = 3, 4$
If $A$ and $X$ are stochastic and $X^p = A$ then it is necessary that

$$\lambda_i(A) \in \Theta_n^p := \{\lambda^p : \lambda \in \Theta_n\} \text{ for all } i.$$
Powers 2, 3, 4, 5 for $n = 3$
Powers 2, 3, 4, 5 for $n = 4$
A example

\[
A = \begin{bmatrix}
\frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\
\frac{1}{2} & 0 & \frac{1}{2} & 0 \\
\frac{10}{11} & 0 & 0 & \frac{1}{11} \\
\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4}
\end{bmatrix}.
\]

\(A\) cannot have a stochastic 12th root, but may have a stochastic 52nd root. None of the 52nd roots is stochastic; \(A^{1/12}\) and \(A^{1/52}\) both have negative elements.
Dependence on \( n \)

- \( \Theta_3 \subseteq \Theta_4 \subseteq \Theta_5 \subseteq \ldots \)

- The number of points at which \( \Theta_n \) intersects the unit circle increases rapidly with \( n \): 23 intersection points for \( \Theta_8 \) and 80 for \( \Theta_{16} \).

- As \( n \) increases, the region \( \Theta_n \) and its powers tend to fill the unit circle.
HIV-Aids matrix has spectrum

\[ \Lambda(P) = \{1, 0.9644, 0.4980, 0.1493, -0.0043\} \].

No real \( p \)th root for even \( p \).

Practitioners regularize the principal \( p \)th root—several approaches.

Practitioners probably unaware of existence of a non-principal stochastic root.
Conclusions

- Literature on roots of stochastic matrices emphasizes computational aspects over theory.
- Identified two classes of stochastic matrices for which $A^{1/p}$ is stochastic for all $p$.
- Wide variety of possibilities for existence and uniqueness, in particular re. primary versus nonprimary roots.
- Gave some necessary spectral conditions for existence.
- More work needed on theory and algorithms.

M. Arioli and D. Loghin.  
Discrete interpolation norms with applications.  

F. L. Bauer.  
*Decrypted Secrets: Methods and Maxims of Cryptology*.  
xii+474 pp.
References II

S. Berridge and J. M. Schumacher. 
An irregular grid method for high-dimensional free-boundary problems in finance. 

T. Charitos, P. R. de Waal, and L. C. van der Gaag. 
Computing short-interval transition matrices of a discrete-time Markov chain from partially observed data. 

E. B. Davies. 
Approximate diagonalization. 
M. Fiedler and H. Schneider. 
Analytic functions of $M$-matrices and generalizations. 

S. Fiori. 
Leap-frog-type learning algorithms over the Lie group of unitary matrices. 

N. J. Higham. 
The Matrix Function Toolbox. 

N. J. Higham and L. Lin.
On $p$th roots of stochastic matrices.
19 pp.

F. Karpelevič.
On the characteristic roots of matrices with nonnegative elements.
J. F. C. Kingman.
The imbedding problem for finite Markov chains.

P. J. Psarrakos.
On the $m$th roots of a complex matrix.

N. M. Rice.
On $n$th roots of positive operators.

H. W. Turnbull.
The matrix square and cube roots of unity.