Some properties of kernel matrices

Hoekstra, Æ.H.

Published: 01/01/1981

Document Version
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the author’s version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher’s website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 12. Dec. 2018
Some properties of kernel matrices

by

A.H. Hoekstra

Eindhoven, June 1981

The Netherlands
Some properties of kernel matrices

1. Definition and elementary properties

We consider discrete-time Markov processes of the type discussed in [4]: the processes have stationary transition probabilities and the transition distribution function \( H(y|x) = P(X_{n+1} = y|X_n = x) \) is given by

\[
(1.1) \quad H(y|x) = \sum_{j=1}^{r} a_j(x)B_j(y) \quad (x,y \in \mathbb{R})
\]

where the \( a_j \) and \( B_j \) are real valued functions, the \( a_j \) are measurable and the \( B_j \) are of bounded variation and continuous from the right.

Definition 1.1: The \( r \times r \) matrix \( C \) with entries

\[
(1.2) \quad c_{ij} = \int a_j(x)dB_i(x) \quad (i,j = 1, \ldots, r)
\]

is called the kernel matrix corresponding to \( (1.1) \).

We shall denote a (column) vector with \( j \)-th component \( v_j \) by \( v \) and its transpose by \( ^tv \).

Proposition 1.2: The \( n \)-step transition distribution function \( H^{(n)}(y|x) \) is given by

\[
(1.3) \quad H^{(n)}(y|x) = ^tv(x)C^{n-1}B(y) \quad (x,y \in \mathbb{R}, \; n \geq 1)
\]

with \( C^0 = I \), the unit matrix.

Proof: see [4].

Proposition 1.3: The representation \( (1) \) of the transition function \( H \) is minimal, i.e. \( H \) cannot be expressed as a sum with less than \( r \) terms, if and only if both the \( a_j(x) \) and \( B_j(x) \) are linearly independent.
Proposition 1.4: Two minimal representations of a transition function $H$ can only differ by a nonsingular linear transformation $T$, i.e. if $H(y|x) = \sum_{j=1}^{r} a_j(x)B_j(y) = \sum_{j=1}^{r} a^*_j(x)B^*_j(y)$ and $r$ is minimal, then there is a nonsingular $r \times r$ matrix $T$, such that $\tau a_j(x) = \tau a_j(x)T$ and $B^*_j(y) = T^{-1}B_j(y)$.

Proof: for the case of a finite Markov chain a proof, which can easily be generalized, will be given in section 2.

Proposition 1.5: Among the minimal representation of a transition function $H$ there is at least one in which the $B_j$ are distribution functions. The corresponding kernel matrix $C$ has all its row sums equal to 1. There is also one in which the $a_j$ are bounded between 0 and 1.

Proof: see [4].

Corollary 1.6: In every fixed minimal representation of $H$ the functions $a_j$ are bounded.

From here on we assume that $r$ in (1.1) is minimal, and we then say that $H$ is of rank $r$.

Let $\mathcal{C}_r(H)$ be the set of all $r \times r$ kernel matrices that correspond to a fixed $H$ of rank $r$. If $C \in \mathcal{C}_r(H)$ then also $T^{-1}CT \in \mathcal{C}_r(H)$ for every nonsingular $r \times r$ matrix $T$, since $T^{-1}CT$ corresponds to the representation $H(y|x) = (\tau a(x)T)(T^{-1}B(y))$ if $C$ corresponds to $H(y|x) = \tau a(x)B(y)$. Combining this with Proposition 1.4 we find that $\mathcal{C}_r(H)$ is a complete class of similar $r \times r$ matrices. Let $\mathcal{C}_r$ be the set of all $r \times r$ kernel matrices.

Proposition 1.7: If $C \in \mathcal{C}_r$ then $C^n \in \mathcal{C}_r$ for $n \geq 2$.

Proof: $C$ is kernel matrix for a transition function $H(y|x) = \tau a(x)B(y)$. Take $H^{(n)}(y|x)$ as a new one-step transition function $\tilde{H}(y|x)$ with

\[(1.4) \quad \tilde{H}(y|x) = \tau a(x)C^{n-1}B(y) = \tau a(x)B(y).\]
We obtain the kernel matrix \( \widetilde{C} = (\int a_j(x) dB_i(x))_{i,j} = C^{-1}(\int a_j(x) dB_i(x))_{i,j} = C^n. \)

**Proposition 1.8:** If \( C \in \mathcal{C}_r \) then every convex combination of \( C, C^2, \ldots, C^n \) is an element of \( \mathcal{C}_r \).

**Proof:** the matrix \( \alpha_1 C + \alpha_2 C^2 + \ldots + \alpha_n C^n \) \( (\alpha_1 \geq 0; \alpha_1 + \ldots + \alpha_n = 1) \) is a kernel matrix for the transition function \( a_1 H(y|x) + a_2 H(2)(y|x) + \ldots + a_n H(n)(y,x) \), as can easily be checked.

As we shall see in section 3 convex combinations of the \( C^k \) including \( C^0 = I \) do not necessarily belong to \( \mathcal{C}_r \). In this respect kernel matrices differ from Markov transition matrices.

**Proposition 1.9:** If \( C \in \mathcal{C}_r \), then \( C \) has an eigenvalue 1 and \( |\lambda| \leq 1 \) for all eigenvalues \( \lambda \) of \( C \).

**Proof:** see [4].

**Proposition 1.10:** If there exists a matrix \( C \in \mathcal{C}_r \) with eigenvalue \( \lambda_0 \), then for each \( \rho \) with \( 0 \leq \rho \leq 1 \) there exists a matrix \( C_\rho \in \mathcal{C}_r \) with eigenvalue \( \rho \lambda_0 \).

**Proof:** Let \( C \), corresponding to \( H = \sum_{j=1}^r a_j(x)B_j(y) \) with the \( B_j \) distribution functions, have an eigenvalue \( \lambda_0 \). For \( 0 \leq \rho \leq 1 \) define

\[
(1.5) \quad H_\rho(y|x) = \sum_{j=1}^r \rho a_j(x)B_j(y) + (1-\rho)B_\rho(y) = \\
= \sum_{j=1}^{r-1} \rho a_j(x)B_j(y) + [\rho a_\rho(x) + 1-\rho]B_\rho(y).
\]

This is a transition distribution function with kernel matrix

\[
(1.6) \quad C_\rho = \rho C + (1-\rho)
\]
We have

\[
(1.7) \quad \det(C_\rho - \lambda I) = \begin{vmatrix}
\rho c_{11} - \lambda & \rho c_{12} & \cdots & \rho c_{1r} + 1 - \rho \\
\rho c_{21} & \rho c_{22} - \lambda & \cdots & \rho c_{2r} + 1 - \rho \\
\vdots & \vdots & \ddots & \vdots \\
\rho c_{r1} & \rho c_{r2} & \cdots & \rho c_{rr} + 1 - \rho - \lambda 
\end{vmatrix}
\]

\[
= \begin{vmatrix}
\rho c_{11} - \lambda & \rho c_{12} & \cdots & \rho c_{1r-1} \\
\rho c_{21} & \rho c_{22} - \lambda & \cdots & \rho c_{2r-1} \\
\vdots & \vdots & \ddots & \vdots \\
\rho c_{r1} & \rho c_{r2} & \cdots & \rho c_{r,r-1} 
\end{vmatrix}
(1 - \lambda) =
\]

\[
= \rho^{n-1}\det(C - \frac{\lambda}{\rho} I),
\]

from which it follows that \( C_\rho \) has an eigenvalue \( \rho \lambda_0 \).

2. The finite case

A finite Markov chain with \( s \times s \) transition matrix \( M = (m_{ij}) \) is a special case of the general Markov process considered in section 1. Let the state space be \( S = \{x_1, \ldots, x_s\} \) and take

\[
(2.1) \quad a_j(x) = \begin{cases}
  m_{ij} & \text{for } x = x_i, \\
  0 & \text{for } x \notin S \text{ and } j = 1, \ldots, s - 1, \\
  1 & \text{for } x \notin S \text{ and } j = s.
\end{cases}
\]

\[
(2.2) \quad B_j(y) = \begin{cases}
  1 & \text{for } y \geq x_j, \\
  0 & \text{for } y < x_j.
\end{cases}
\]

These functions satisfy (1.1) with \( r = s \), the \( B_j \) being distribution functions, but the representation not necessarily being minimal. To get a minimal one suppose that \( M \) has rank \( r \). Then \( M \) can be written as the product of an \( s \times r \) matrix \( A \) and an \( r \times s \) matrix \( B \), where \( A \) and \( B \) both have rank \( r \). This is
trivial if $r = s$ (take $A = I_r$ and $B = M$, which is actually (2.1)); if $r < s$ we can choose $r$ independent rows of $M$, say (after re-ordering) the first $r$ rows. Since every row of $M$ is now a linear combination of these $r$ rows, we have

$$m_{ij} = \sum_{k=1}^{r} a_{ik} m_{kj}$$

with $a_{ik} = \delta(i,k)$ for $i = 1, \ldots, r$ and $\sum_{k=1}^{r} a_{ik} = 1$ for $i = 1, \ldots, s$. Now let $A = (a_{ik})_{i=1, \ldots, s}$ and $B = (m_{kj})_{k=1, \ldots, r}$. We evidently have

$$M = AB.$$ 

The rows of $B$ are probability distributions. The upper $r \times r$ part of $A$ is the unit matrix $I_r$; in the lower $(s - r) \times r$ part some of the entries may be negative or larger than 1.

The kernel matrix $C$ now takes the simple form

$$C = BA.$$ 

We shall now give a proof a Proposition 1.4 in this special case. We write $\mathcal{C}_r(M)$ for the set of all kernel matrices corresponding to a given transition matrix $M$ of rank $r$.

"Proof" of Proposition 1.4: Suppose $M = AB = A^* B^*$; $C = BA \in \mathcal{C}_r(M)$ and $C^* = B^* A^* \in \mathcal{C}_r(M)$.

$A$ has rank $r$, so we can choose $r$ independent rows of $A$, say $\tilde{a}_1, \ldots, \tilde{a}_r$, that together form a nonsingular $r \times r$ matrix $\tilde{A}$. Let $\tilde{A}^*$ be the $r \times r$ matrix consisting of the corresponding rows of $A^*$, and define

$$T = \tilde{A}^{-1} \tilde{A}^*.$$ 

From $AB = A^* B^*$ it follows that $\tilde{A}B = \tilde{A}^* B^*$ and hence $B = TB^*$. We now have $ATB^* = A^* B^*$ and hence $AT = A^*$ (consider only $r$ independent columns of $B^*$).
Since $B$ and $B^*$ both have rank $r$, $T$ is nonsingular, and we may conclude that

\[ (2.7) \quad C^* = T^{-1}CT. \]

For a fixed $M$ all matrices $C \in \mathcal{C}_r(M)$ have the same characteristic polynomial $P_C(\lambda) = \det(C - \lambda I_r)$, as they are similar. If $P_M(\lambda)$ denotes the characteristic polynomial of $M$, we have

**Proposition 2.1:**

\[ (2.8) \quad P_M(\lambda) = \lambda^{s-r}P_C(\lambda), \]

where $s$ is the order of $M$.

**Proof:** Let $A$ and $B$ be the matrices that produce $C$ and define matrices $A_0$ and $B_0$ by

\[ (2.9) \quad A_0 = [A | R], \quad B_0 = [B | \theta], \]

where $R$ is any $s \times (s - r)$ matrix that gives $A_0$ rank $s$ and $\theta$ has all entries equal to 0. From $M = AB = A_0B_0$ we find

\[ (2.10) \quad P_M(\lambda) = \det(M - \lambda I_s) = \det(A_0B_0 - \lambda I_s) \]

\[ = \det A_0(B_0A_0 - \lambda I_s)A_0^{-1} = \det(B_0A_0 - \lambda I_s) \]

\[ = \det(BA - \lambda I_r)\det(-\lambda I_{s-r}) = \lambda^{s-r}P_C(\lambda). \]

**Corollary 2.2:** All nonzero eigenvalues of $M$ are also eigenvalues of $C \in \mathcal{C}_r(M)$, with the same multiplicity.

**Corollary 2.3:** $C \in \mathcal{C}_r(M)$ has an eigenvalue 1 and $|\lambda| < 1$ for all eigenvalues $\lambda$ of $C$ (since this is true for the transition matrix $M$; Cf. Proposition 1.9).

**Corollary 2.4:** The trace $\sigma(C)$ of $C \in \mathcal{C}_r(M)$, i.e. the sum of the eigenvalues of $C$, is nonnegative.
It does not follow that all eigenvalues of a matrix $C \in \mathcal{C}_r(M)$ are nonzero.
In fact, $C$ can even have an eigenvalue 0 of multiplicity $r - 1$. Take e.g.

$$M = \begin{pmatrix}
0 & 1 & 0 & \cdots & 0 \\
0 & 0 & 1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & \cdots & \cdots & 0 & 1
\end{pmatrix} = \begin{pmatrix}
1 & 0 & \cdots & 0 \\
0 & 1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & \cdots & \cdots & 1
\end{pmatrix} \begin{pmatrix}
0 & 1 & 0 & \cdots & 0 \\
0 & 0 & 1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & \cdots & \cdots & 0 & 1
\end{pmatrix}$$

(2.11)

$$C = \begin{pmatrix}
0 & 1 & 0 & \cdots & 0 \\
0 & 0 & 1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & \cdots & \cdots & 0 & 1
\end{pmatrix}$$

(2.12)

\[M\text{ is a } (r + 1) \times (r + 1) \text{ transition matrix of rank } r, \text{ and } C \text{ is a } r \times r \text{ matrix of rank } r - 1, \text{ and } \lambda^r = \lambda^{r-1}(1 - \lambda).\]

If, as in the example above, $C$ has an eigenvalue 0, the rank of $C$ is less than $r$. We can then apply the factorisation procedure to $C$ instead of $M$, and find $C = A_1 B_1$ with $C_1 = B_1 A_1$ and $M^n = A C^{n-1} B = A A_{r} C_{r}^{n-2} B_1 B$. If we go on we eventually obtain a matrix $C_{r_0}$ with only nonzero eigenvalues and

$$M^n = \tilde{A} C_{r_0}^{n-1-r_0} \tilde{B} \quad (n \geq r_0 + 1)$$

(2.13)

where $\tilde{A} = A_{r_0} \ldots A_{r_0}$ and $\tilde{B} = B_{r_0} \ldots B_1 B$. In the above example $r_0 = r - 1$ and $C_{r_0} = (1)$. Indeed, $M^n$ is constant for $n \geq r$ in this case.

**Proposition 2.5:** If $C \in \mathcal{C}_r(M)$ then $\Gamma := \lim_{n \to \infty} C^n$ exists if and only if $M^{\infty} := \lim_{n \to \infty} M^n$ exists.

**Proof:** the statement follows immediately from $M^n = AC^{n-1}B$ and $C^n = BM^{n-1}A$. $\square$
Proposition 2.6: Let $C = BA \in C^r(M)$ and let the rows of $B$ be probability distributions. If $C$ has a single eigenvalue $\lambda$ and no other eigenvalues with $|\lambda| = 1$, then $\Gamma$ exists, has identical rows with sums 1, and

\[ (2.14) \quad \Gamma = MA. \]

Proof: under the conditions of the proposition $M^\infty$ exists and has identical rows with sums 1 (probability distributions). Hence $\Gamma$ exists with $\Gamma = BM^\infty = M^\infty A$, and $\Gamma$ also has identical rows with sums 1.

Proposition 2.7: Every kernel matrix $C$ is the limit of a sequence of kernel matrices corresponding to finite transition matrices.

Proof: Let $C$ correspond to $H(y|x) = \sum_{j=1}^r a_j(x) B_j(y)$. It is no restriction to assume that the $B_j$ are distribution functions. By Corollary 1.6 the $a_j$ are bounded, say $|a_j(x)| \leq L$ for all $x$ and all $j \in \{1, \ldots, r\}$. For each fixed $k \in \mathbb{N}$ and $j \in \{1, \ldots, r\}$ the sets

\[ (2.15) \quad A_{j,k} = \{x \mid \frac{L}{k} \leq a_j(x) < \frac{L + 1}{k}\}, \quad L = -Lk, -Lk+1, \ldots, Lk-1, \]

form a measurable partition of $\mathbb{R}$ (since $a_j$ is a measurable function).

Keep $k$ fixed and let $A_{j,1}^{(k)}, \ldots, A_{j,N(k)}^{(k)}$ be a measurable partition containing the $A_{j,k}$. Take an arbitrary fixed $x_n^{(k)} \in A_n^{(k)}$ for each $n \in \{1, \ldots, N(k)\}$, define the function $a_j^{(k)}$ by

\[ (2.16) \quad a_j^{(k)}(x) = a_j(x_n^{(k)}) \text{ if } x \in A_n^{(k)}, \]

and let $B_j^{(k)}$ be the discrete distribution function with jumps $B_j(A_n^{(k)})$ at the points $x_n^{(k)}$.

The $a_j^{(k)}$ are measurable stepfunctions and for all $x$ and $j$

\[ (2.17) \quad \lim_{k \to \infty} |a_j(x) - a_j^{(k)}(x)| \leq \lim_{k \to \infty} \frac{1}{k} = 0. \]
Furthermore

\[ H^{(k)}(y|x) := \sum_{j=1}^{k} a_j^{(k)}(x)B_j^{(k)}(y) = \sum_{j=1}^{k} a_j(x_n^{(k)})B_j(\bigcup_{\{n \mid x_n^{(k)} \leq y\}} A_n^{(k)}) \]

if \( x \in A_n^{(k)} \)

is a transition function concentrated on the finite set \( \{x_1^{(k)}, \ldots, x_{N(k)}^{(k)}\} \).

Let \( C^{(k)} = (c_{ij}^{(k)}) \) be the corresponding kernel matrix. We have

\[ c_{ij}^{(k)} = \frac{N(k)}{n!} a_j^{(k)}(x_n^{(k)})b_i^{(k)}(A_n^{(k)}) = \int a_j^{(k)}(x)dB_i^{(k)}(x) = \int a_j^{(k)}(x)dB_i(x). \]

Using Lebesgue's dominated convergence theorem (\(|a_j(x)| \leq L, \int LdB_i < \infty\)) we find for all \( i, j \in \{1, \ldots, r\} \)

\[ \lim_{k \to \infty} c_{ij}^{(k)} = \lim_{k \to \infty} \int a_j^{(k)}(x)dB_i(x) = \int a_j(x)dB_i(x) = c_{ij}. \]

**Proposition 2.8:** The trace of a kernel matrix is nonnegative.

**Proof:** Combine Proposition 2.7 and Corollary 2.4.

3. Eigenvalues

A. Stochastic matrices

Let \( \bar{M}_n \) denote the set of all complex numbers \( \lambda \) that are eigenvalues of stochastic matrices of order \( n \). The problem of determining \( \bar{M}_n \) (or, slightly more general, the set of all eigenvalues of nonnegative matrices of order \( n \)) was posed by Kolmogorov, partly solved by Dmitriyev and Dynkin in 1946 [1], and finally completely solved by Karpelewitsj in 1951 [3].

\( \bar{M}_n \) turns out to be a closed, star-shaped subset of the unit disk. The only points of \( \bar{M}_n \) on the unit circle are the points \( e^{2\pi i \frac{k}{n}} (k \leq n) \). \( \bar{M}_n \) is
symmetrical with respect to the real axis.

The boundary of $\tilde{M}_n$ between 1 and $e^{\frac{2\pi i}{n}}$ is a straight line, and further consists of polynomial arcs. See [3] for the explicit formulas. The basic observation in [1] and [3] is that $\lambda \in \tilde{M}_n$ if and only if there exists a convex $k$-angular polygon $Q$ ($k \leq n$), which is mapped into itself when multiplied by $\lambda$.

Kernel matrices $C$ are in general not nonnegative, and it is not clear how the arguments used in [1] and [3] can be extended to get information about $\tilde{C}_r$, the set of all complex numbers $\lambda$ that are eigenvalues of kernel matrices $C \in \mathcal{C}_r$. If we go back to the transition distribution function $H(y|x)$ we have to study eigenfunctions $\varphi$ instead of eigenvectors. ($\varphi$ is called an eigenfunction of $H(y|x)$ if $\int \varphi(y)dH(y|x) = \lambda \varphi(x)$. It can be shown that all eigenfunctions have the form $\varphi(x) = \sum_{j=1}^{k} \theta_j a_j(x)$, where $\theta$ is an eigenvector of $C$). From section 2 it is clear that $\tilde{C}_r$ is also contained in the unit disk and that $\tilde{M}_r \subset \tilde{C}_r$. 

\textit{Figure 3.1.}
Proposition 3.1: \( \mathcal{C}_2 \) is the set of all \( 2 \times 2 \) matrices that are similar to \( 2 \times 2 \) transition matrices.

Proof: Let \( C \in \mathcal{C}_2 \). By Proposition 1.9 \( C \) has an eigenvalue 1 and one other real eigenvalue \( \lambda \) with \( |\lambda| \leq 1 \).

If \( \lambda < 1 \) then \( C \) is similar to \( \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix} \), its Jordan normal form, and this matrix in its turn is similar (via the transformation matrix \( T = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \)) to the matrix

\[
\begin{pmatrix}
\frac{1 + \lambda}{2} & \frac{1 - \lambda}{2} \\
\frac{1 - \lambda}{2} & \frac{1 + \lambda}{2}
\end{pmatrix}
\]

which is a transition matrix.

If \( \lambda = 1 \) then the Jordan normal form cannot be \( \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \) (since then \( C^n \) would be unbounded for \( n \to \infty \)), so it is \( \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \). \( \square \)

Corollary 3.2: \( \mathcal{C}_2 = \mathcal{M}_2 = [-1,1] \subset \mathbb{R} \).

If \( M \) is finite transition matrix of order \( n \) and rank 2, the rows of \( M \) considered as points in \( \mathbb{R}^n \) lie on a straight line in \( \mathbb{R}^n \). Each of these \( n \) points is a convex combination of the two extreme points on this line. If we take these extreme points to compose the \( 2 \times n \) matrix \( B \) (cf. the beginning of section 2), all entries of the corresponding matrix \( A \) are nonnegative (and at most 1). It is clear that now \( BA \) is a transition matrix. This provides an other proof for the finite case of Proposition 3.1 (problem 99 in Statistica Neerlandica 34 (1980), solution by J.Th. Runnenburg).

C. \( \mathcal{C}_3 \)

Only nonreal eigenvalues are of interest to us if we want information about \( \mathcal{C}_3 \). Let \( 1, x + iy \) and \( x - iy \) (\( y > 0 \)) be the three eigenvalues.
of a kernel matrix $C \in \mathfrak{C}_3$. The sum of the eigenvalues is nonnegative (Corollary 2.4) hence $x \geq -\frac{1}{4}$. This leads to

**Proposition 3.3**: If $u + iv$ is a complex number with $v \neq 0$ such that $\text{Re}(u + iv)^n < -\frac{1}{4}$ then $u + iv \notin \bar{C}_3$.

**Proof**: If $u + iv$ is an eigenvalue of a matrix $C \in \mathfrak{C}_3$, then $(u + iv)^n$ is an eigenvalue of the matrix $C^n \in \mathfrak{C}_3$, so that $\text{Re}(u + iv)^n$ is at least $-\frac{1}{4}$.

Proposition 3.3 enables us to exclude a part of the unit disk in searching the area of $\bar{C}_3$.

\[\text{figure 3.2.}\]

$\bar{C}_3$ is contained in the nonshaded area (except the part $[-1,-\frac{1}{4}]$ of the real axis) and, on the other hand, contains the triangle $(\bar{H}_3 \subseteq \bar{C}_3)$.

Let $z_n = 2 \pi \frac{i}{n} (n = 1,2,...)$, the solution of the equation $z^n = -\frac{1}{4}$. 
Set \( z_n = x_n + iy_n \) and \( t := \frac{n}{n} \), then

\[
x_n \frac{2^{-t} \cos t}{t}, \quad y_n = \frac{2^{-t} \sin t}{t}.
\]

To see the behaviour of the sequence \((z_n)\) near 1 we take \( t \) as a continuous parameter and let \( t \) tend to 0 (from above). We find

\[
\left. \frac{dy}{dx} \right|_{x=1} = \left. \frac{dy}{dt} \cdot \frac{dt}{dx} \right|_{t=0}
\]

\[
= \left. \left[ (2^{-t} \cos t - 2^{-t} \log 2 \sin t) (2^{-t} \sin t - 2^{-t} \log 2 \cos t) \right]^{-1} \right|_{t=0}
\]

\[
= - \frac{1}{\log 2} = -1.44\ldots
\]

So, in particular, we see that \( \left. \frac{dy}{dx} \right|_{x=1} \) is finite.

In order to obtain an inner bound for \( \zeta_3 \) we considered the following problem: find a transition matrix \( M \) of order 4 and rank 3, with eigenvalues \( x \pm itx \) for a specified \( t \) such that \( x \) is maximal (for \( t > 0 \)) or minimal (for \( t < 0 \)). Since for every matrix \((p_{ij})_{i,j=1}^n\) with eigenvalues \( \lambda_1, \ldots, \lambda_n \) the following relations hold for the first two invariants of the matrix (see e.g. [2], sec 4.3)

\[
(3.2) \quad \sum_{i=1}^n \lambda_i = \sum_{i=1}^n p_{ii}
\]

\[
(3.3) \quad \sum_{i \neq j} \lambda_i \lambda_j = \sum_{i \neq j} \left| \begin{array}{cc} p_{ii} & p_{ij} \\ p_{ji} & p_{jj} \end{array} \right|
\]

the above problem can be formulated as follows (\( t > 0 \), fixed):
maximize \( x = \frac{1}{4}(p_{11} + p_{22} + p_{33} + p_{44} - 1) \) under the conditions

\[
\begin{align*}
& p_{ij} \geq 0; \\
& \sum_{j=1}^{4} p_{ij} = 1 \text{ for } i = 1, \ldots, 4; \\
& 2x + x^2 + t^2 = \sum_{i \neq j} p_{ii} p_{jj}.
\end{align*}
\]

The maximization has been carried out by computer. The result is given in figure 3.3. The arc between the points \( e^{2\pi i/3} \) and \( \frac{i}{2} + \frac{1}{2}i \) is given by (for \(-\frac{1}{2} \leq x \leq \frac{1}{2}\))

\[
(3.4) \quad y^2 \leq \frac{1}{4}(x - \frac{1}{2})^2 + \frac{1}{4}.
\]

To see this, let the eigenvalues be 1, 0 and \( x + iy \), and take \( x \) fixed (i.e. \( p_{11} + p_{22} + p_{33} + p_{44} \) is fixed). From (3.2) we now deduce

\[
(3.5) \quad y^2 \leq \sum_{i \neq j} p_{ii} p_{jj} - x^2 - 2x \leq 6p^2 - x^2 - 2x,
\]

where \( p = \frac{1}{4}(p_{11} + p_{22} + p_{33} + p_{44}) = \frac{1}{4}(2x + 1) \). This leads directly to (3.4). For \(-\frac{1}{2} \leq x \leq \frac{1}{2}\) the upper bound for \( y^2 \) is attained if we take

\[
M = \begin{pmatrix}
  p & a & g & 0 \\
  0 & p & 1-p & 0 \\
  0 & 0 & p & 1-p \\
  1-p & 0 & 0 & p
\end{pmatrix}
\]

with \( a = p(1 - p) + \frac{p^4}{(1 - p)^2} \) and \( g = \frac{(1 - p)^4 - p^4}{(1 - p)^2} \). The first row is now a linear combination of the 2nd, 3rd and 4th row, so that \( M \) has rank 3. This fails for \( x > \frac{1}{2} \) (then \( p > \frac{1}{2} \) and \( g \) becomes negative). And,
in general, no transition matrix with all main diagonal entries greater than 1 can have an eigenvalue 0 (see e.g. [2], Sec. 6.8, Gershgorin's theorem).

Figure 3.3.

At present this is about all the information we have about $C_3$. We are trying to get more numerical information by considering the situation of a $5 \times 5$ transition matrix of rank 3. Starting from a $n \times n$ transition matrix $M$ we obtain, instead of (3.5), the inequality

\[ y^2 \leq \frac{p^2}{n} - x^2 - 2x \]

where now $p = \frac{1}{n} (2x + 1)$, so that we find

\[ y^2 \leq (1 - \frac{2}{n})x^2 - \frac{2}{n} x + \frac{1}{2}(1 - \frac{1}{n}). \]

For $n \to \infty$ this yields the inequality $\text{Re}(x + iy)^2 \geq -\frac{1}{4}$. It is doubtful whether these values can actually be attained.

Using the foregoing results about $C_3$ we can show that several of the pleasant properties of the set $\mu_\tau$ of stochastic matrices are not inherited by $\bar{\mu}_\tau$. 
First, $\mathcal{C}_r$ is not closed under matrix multiplication. Take for example

\[ C_1 = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & 0 \\ 0 & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & -\frac{1}{4} & 1 \end{pmatrix} \quad \text{and} \quad C_2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \]

$C_1$ is a kernel matrix corresponding to

\[ M = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{4} & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & 0 & 0 & \frac{1}{4} \end{pmatrix} \]

and $C_2$ is a kernel matrix as it is itself a transition matrix of full rank. We have however

\[ C_1 C_2 = \begin{pmatrix} 0 & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & 0 & \frac{1}{4} \\ 1 & \frac{1}{4} & -\frac{1}{4} \end{pmatrix} \]

with a negative trace, so that (by Proposition 2.8) $C_1 C_2 \not\in \mathcal{E}_3$.

The same two kernel matrices may serve as a counterexample to see that $\mathcal{E}_r$ is not convex. The eigenvalues of $C_\alpha = \alpha C_1 + (1 - \alpha) C_2$ are $\lambda_1 = 1$ and

\[ \lambda_{2,3} = \alpha - \frac{1}{4} + \frac{1}{2} i \sqrt{3 - 2\alpha} . \]

Relation (3.4) now reads

\[ \frac{1}{2} (3 - 2\alpha) \leq \frac{1}{2} (\alpha - 1)^2 + \frac{1}{4} . \]

This is true for $\alpha = 0$ and for $\alpha = 1$, but not for any $\alpha$ between 0 and 1.
4. References


