Element-wise functions preserving positivity of matrices

Roland Hildebrand

LJK / CNRS

Kolloquium Graduiertenkolleg ALOP, Univ. Trier
January 23, 2017

Outline

- semi-definite matrices
- Hadamard functions preserving positivity
- representations of compact Lie groups

- maxcut polytope
- Nesterovs $\pi/2$ theorem

- copositive matrices
- triangle-free polytope
- representations of extreme rays

Positive semi-definite matrices

Definition

A real symmetric $n \times n$ matrix A is called positive semi-definite if $x^T A x \geq 0$ for all $x \in \mathbb{R}^n$.
The set of all positive semi-definite matrices forms the positive semi-definite cone $\mathcal{S}_+^n$.

- $\mathcal{S}_+^n$ is closed convex pointed
- $\mathcal{S}_+^n$ is symmetric (homogeneous and self-dual)
- used in semi-definite programming as the base cone of conic programs
- $A \in \mathcal{S}_+^n$ if and only if $\lambda_i(A) \geq 0$ for all i
- $A \in \mathcal{S}_+^n$ if and only if A is a Gram matrix of vectors in $\mathbb{R}^n$

- if $A \in \mathcal{S}_+^n$, then $A_{ii} \geq 0$ for all i
- $\operatorname{diag} A = \mathbf{1}$, then $A \in \mathcal{S}_+^n$ if and only if A is a Gram matrix of vectors on the unit sphere

Maps preserving positivity

submatrices $A \mapsto (A_{ij})_{i,j \in I \subset \{1,\dots,n\}}$

spectral functions

- $A \mapsto A^{-1}$ (for A invertible)
- $A \mapsto A^k$
- $A = U \operatorname{diag}(\lambda) U^T \mapsto f(A) = U \operatorname{diag}(f(\lambda)) U^T$, $f : \mathbb{R}_+ \to \mathbb{R}_+$

binary operations

- convex sums $(A, B) \mapsto \alpha A + \beta B$, $\alpha, \beta \geq 0$, $\alpha + \beta = 1$
- Kronecker product $(A, B) \mapsto A \otimes B$
- Hadamard product $(A, B) \mapsto A \circ B$

these preserve also the diag $= \mathbf{1}$ property

$$(A \otimes B)_{(i,k),(j,l)} = A_{ij} B_{kl}$$

$$(A \circ B)_{ij} = A_{ij} B_{ij}$$

Kronecker and Hadamard

$A \circ B$ is a principal submatrix of $A \otimes B =$

$$\begin{pmatrix}
A_{11}B_{11} & A_{11}B_{12} & A_{11}B_{13} & A_{12}B_{11} & A_{12}B_{12} & A_{12}B_{13} & A_{13}B_{11} & A_{13}B_{12} & A_{13}B_{13} \\
A_{11}B_{12} & A_{11}B_{22} & A_{11}B_{23} & A_{12}B_{12} & A_{12}B_{22} & A_{12}B_{23} & A_{13}B_{12} & A_{13}B_{22} & A_{13}B_{23} \\
A_{11}B_{13} & A_{11}B_{23} & A_{11}B_{33} & A_{12}B_{13} & A_{12}B_{23} & A_{12}B_{33} & A_{13}B_{13} & A_{13}B_{23} & A_{13}B_{33} \\
A_{12}B_{11} & A_{12}B_{12} & A_{12}B_{13} & A_{22}B_{11} & A_{22}B_{12} & A_{22}B_{13} & A_{23}B_{11} & A_{23}B_{12} & A_{23}B_{13} \\
A_{12}B_{12} & A_{12}B_{22} & A_{12}B_{23} & A_{22}B_{12} & A_{22}B_{22} & A_{22}B_{23} & A_{23}B_{12} & A_{23}B_{22} & A_{23}B_{23} \\
A_{12}B_{13} & A_{12}B_{23} & A_{12}B_{33} & A_{22}B_{13} & A_{22}B_{23} & A_{22}B_{33} & A_{23}B_{13} & A_{23}B_{23} & A_{23}B_{33} \\
A_{13}B_{11} & A_{13}B_{12} & A_{13}B_{13} & A_{23}B_{11} & A_{23}B_{12} & A_{23}B_{13} & A_{33}B_{11} & A_{33}B_{12} & A_{33}B_{13} \\
A_{13}B_{12} & A_{13}B_{22} & A_{13}B_{23} & A_{23}B_{12} & A_{23}B_{22} & A_{23}B_{23} & A_{33}B_{12} & A_{33}B_{22} & A_{33}B_{23} \\
A_{13}B_{13} & A_{13}B_{23} & A_{13}B_{33} & A_{23}B_{13} & A_{23}B_{23} & A_{23}B_{33} & A_{33}B_{13} & A_{33}B_{23} & A_{33}B_{33}
\end{pmatrix}$$

Hadamard functions

the k-th Hadamard power

$$A \mapsto A^{\circ k} = A \circ A \circ \cdots \circ A = (A_{ij}^k)_{ij}$$

is an element-wise function preserving positive semi-definiteness

generalizes to Hadamard functions

let $f : \mathbb{R} \to \mathbb{R}$ be a scalar function
define $f[A] = (f(A_{ij}))_{ij}$ be element-wise application of f on A

Corollary

Let $f : \mathbb{R} \to \mathbb{R}$ be an entire function with nonnegative Taylor coefficients. Then the Hadamard function $A \mapsto f[A]$ is positivity preserving.

partial sums of the Taylor series are positive semi-definite and converge to a positive semi-definite limit matrix

Unit diagonal case

we restrict to the subset of matrices $A \in \mathcal{S}_+^n$ with $\mathrm{diag}(A) = \mathbf{1}$

then $|A_{ij}| \leq 1$ and we may consider scalar functions $f : [-1, 1] \to \mathbb{R}$ (which may be normalized to $f(1) = 1$)

Theorem (Schönberg)

Let $f : [-1, 1] \to \mathbb{R}$ be continuous. Then f is positivity preserving (for all n) if and only if it is analytic, the Taylor series converges on the unit disc, and all Taylor coefficients are nonnegative.

Theorem (Schönberg, Crum)

Let $f : [-1, 1] \to \mathbb{R}$ be measurable. Then f is positivity preserving (for all n) if and only if it is analytic in $(-1, 1)$, the Taylor series converges on the unit disc, all Taylor coefficients are nonnegative, and $f(1) - \lim_{t \to 1} f(t) \geq |f(-1) - \lim_{t \to -1} f(t)|$.

the Hadamard powers generate extreme rays of the cone of positivity preserving functions

Finite size

$n = 2$: $f : [-1, 1] \to \mathbb{R}$ positivity preserving if and only if $f(1) \geq |f(x)|$ for all $x \in [-1, 1]$

$$A = \begin{pmatrix} 1 & a \\ a & 1 \end{pmatrix} \mapsto \begin{pmatrix} f(1) & f(a) \\ f(a) & f(1) \end{pmatrix}$$

$f[A] \in \mathcal{S}_+^2$ if and only if $|f(a)| \leq f(1)$

with normalization $f(1) = 1$ the positivity preserving functions are the unit ball of $L_\infty([-1, 1])$

f is extremal if and only if it is measurable with $|f(x)| \equiv 1$

$n \geq 3$: open problem

Finite rank

constrain matrix rank instead of matrix size

$S_+(n,k)$ — set of $n \times n$ real symmetric PSD matrices of rank $\leq k$ with diag $A = \mathbf{1}$

Definition

We call $f : [-1,1] \to \mathbb{R}$ rank k positivity preserving if $f[A] \in \mathcal{S}_+^n$ for all $n \geq 1$ and $A \in S_+(n,k)$.

Theorem (Schönberg)

Let $f : [-1,1] \to \mathbb{R}$ be continuous. Then f is rank k positivity preserving if and only if the Gegenbauer series (with parameter $\alpha = k/2 - 1$) of f has nonnegative coefficients. In this case the series is converging absolutely and uniformly.

Gegenbauer polynomials

the *Gegenbauer polynomials* or *ultraspherical polynomials* $C_l^{(\alpha)}(t)$ with parameter α are the orthogonal polynomials on $[-1,1]$ with weight $w(t) = (1-t^2)^{\alpha-1/2}$

$$\int_{-1}^{1} C_k^{(\alpha)}(t) C_l^{(\alpha)}(t)(1-t^2)^{\alpha-1/2}\,dt = \frac{\pi 2^{1-2\alpha}\Gamma(l+2\alpha)}{l!(l+\alpha)(\Gamma(l))^2}\delta_{kl}$$

every $f \in L_2([-1,1],w)$ can be expanded in a series

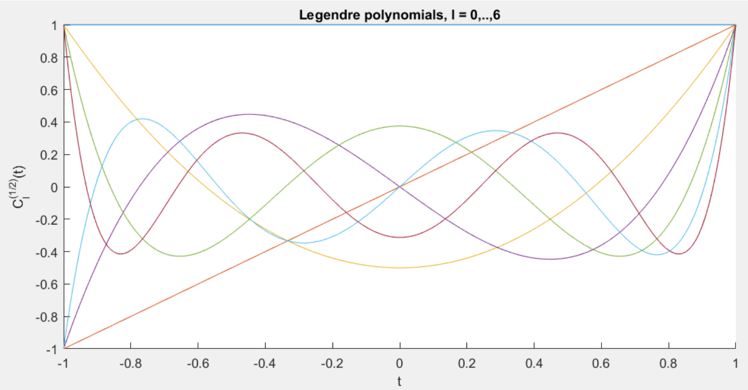
$$f(t) = \sum_{l=0}^{\infty} c_l(f) C_l^{(\alpha)}(t)$$

with coefficients

$$c_l = \frac{l!(l+\alpha)(\Gamma(l))^2}{\pi 2^{1-2\alpha}\Gamma(l+2\alpha)} \int_{-1}^{1} f(t) C_l^{(\alpha)}(t)(1-t^2)^{\alpha-1/2}\,dt$$

$k = 2$: Chebycheff polynomials $T_l(\cos\theta) = \cos(l\theta)$, weight $w(t) = (1 - t^2)^{-1/2}$

$k = 3$: Legendre polynomials, weight $w(t) \equiv 1$



$k \to \infty$: with an appropriate normalization $\lim_{\alpha\to\infty} C_l^{(\alpha)}(t) = t^n$ accordingly, the Gegenbauer coefficients tend to the Taylor coefficients

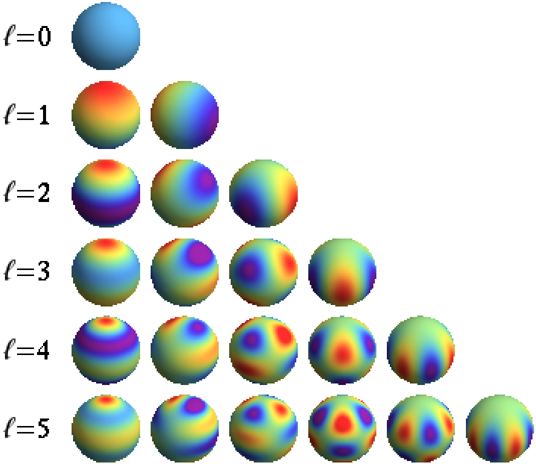
Mathematical background

- $A \in S_+(n,k) \Leftrightarrow A$ Gramian of vectors $x \in S^{k-1}$
- f rank k positivity preserving $\Leftrightarrow K(x,y) = f(\langle x, y \rangle)$ positive definite kernel on S^{k-1}
- $S^{k-1} = O(k)/O(k-1)$ is a homogeneous space: $O(k-1)$ isotropy subgroup of $x_0 \in S^{k-1}$
- $K(x,y) = K(gx, gy)$ for all $g \in O(k)$: kernel is bi-zonal

- $O(k)$ acts linearly on $L_2(S^{k-1})$
- Peter-Weyl theorem: this quasiregular representation decomposes into irreducible representations (harmonics)
- the quasiregular representation is multiplicity-free
- Berezin, Gelfand, Graev, Naimark: each irreducible subspace contains one zonal spherical function, i.e., which is invariant under the action of $O(k-1)$, $z(x) = z(\langle x, x_0 \rangle)$
- zonal harmonic of order l is the Gegenbauer polynomial $C_l^{(\alpha)}$

Spherical harmonics

$k = 3$:



Generalization to arbitrary groups

real symmetric matrices	general case
$O(k)$	compact Lie group G
$O(k-1)$	Lie subgroup H
S^{k-1}	homogeneous space G/H
$[-1,1]$	coset space $H \setminus G/H$
$C_l^{(\alpha)}$	zonal harmonic of order l
matrix $A \in S_+(n,k)$	matrix $A = ((g_j H)^{-1} g_i H)_{ij}$
function $f(\langle x, y \rangle)$	bi-zonal kernel $K(x,y)$
positivity preserving f	positive definite K

the quasiregular representation of G on $L_2(G/H)$ has to be multiplicity-free

Description of PD kernels

Theorem (Bochner)

Let $f : C \to \mathbb{C}$ be a continuous function. Then the following are equivalent:
i) the function f satisfies $f(Hg^{-1}H) = \overline{f(HgH)}$ for all $g \in G$, and for for every positive integer n and every n-tuple of points $g_1H, \ldots, g_nH \in M$ the matrix $(f((g_jH)^{-1}g_iH))_{i,j=1,\ldots,n}$ is PSD;
ii) the function f is a sum of zonal spherical functions with nonnegative real coefficients.
In this case, the corresponding Fourier series converges absolutely and uniformly to f.

Theorem (Crum,Devinatz)

Let $f : C \to \mathbb{C}$ be a measurable function satisfying i) above. Then $f = f_c + f_0$, where f_c, f_0 satisfy i), f_c is continuous, and f_0 is zero a.e.

Generalizations

- may replace real symmetric matrices by complex hermitian $(O(k) \mapsto U(k))$ or quaternionic hermitian $(O(k) \mapsto Sp(k))$ matrices
- the image of the positivity preserving map f has to be $\mathbb{C}$
- the positivity preserving property comes from the fact that the two-sided cosets $(g_j H)^{-1} g_i H$ can be parameterized by scalar products $\langle g_i x_0, g_j x_0 \rangle$ which form a Gramian

in the complex hermitian case the Gegenbauer polynomials are replaced by the generalized Zernike polynomials (Shapiro)

in the quaternionic case the zonal harmonics are still more complicated (Vilenkin, Klimyuk)

Maxcut polytope

denote $S_+(n,n) = \{A \succeq 0 \mid \mathrm{diag}(A) = \mathbf{1}\}$ by $\mathcal{SR}$

Definition

The maxcut polytope is the subset of $\mathcal{SR}$ given by

$$\mathcal{MC} = \mathrm{conv}\{xx^T \mid x \in \{-1,1\}^n\}.$$

- polytope with 2^{n-1} vertices
- symmetries $A \mapsto PAP^T$, $A \mapsto DAD$ with $P \in S_n$, D diagonal with $D^2 = I$
- optimisation over $\mathcal{MC}$ is a hard problem
- $\mathcal{SR}$ is the standard semi-definite relaxation overbounding $\mathcal{MC}$

Trigonometric approximation

Definition (Hirschfeld)

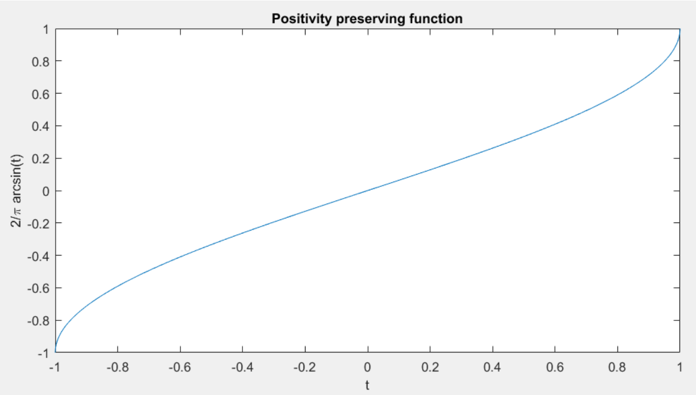
The non-convex set

$$\mathcal{TA} = \left\{ \frac{2}{\pi} \arcsin[A] \mid A \in \mathcal{SR} \right\}$$

is called the trigonometric approximation of the maxcut polytope.

$$\arcsin t = t + \frac{1}{2} \cdot \frac{t^3}{3} + \frac{1}{2} \cdot \frac{3}{4} \cdot \frac{t^5}{5} + \frac{1}{2} \cdot \frac{3}{4} \cdot \frac{5}{6} \cdot \frac{t^7}{7} + \dots$$

$f(t) = \frac{2}{\pi} \arcsin(t)$ is positivity preserving $\Rightarrow \mathcal{TA} \subset \mathcal{SR}$



Positivity preserving function
2/π arcsin(t)
t
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1

Inner approximation

Lemma (Nesterov)

Let $X \in \mathcal{SR}$ *and let* $\xi \sim \mathcal{N}(0, X)$. *Then* $\mathbb{E}(\operatorname{sgn}\xi)(\operatorname{sgn}\xi)^T = \frac{2}{\pi}\arcsin[X]$.

Corollary

$\operatorname{conv}\mathcal{TA} = \mathcal{MC}$

proof

- $(\operatorname{sgn}\xi)(\operatorname{sgn}\xi)^T \in \mathcal{MC} \Rightarrow \mathcal{TA} \subset \mathcal{MC}$
- $xx^T \in \mathcal{MC} \Rightarrow \frac{2}{\pi}\arcsin[xx^T] = xx^T$
- vertices of $\mathcal{MC}$ are in $\mathcal{TA}$

Nesterovs $\frac{\pi}{2}$ theorem

consider the problem $\max\{\langle C, A\rangle \mid A \in \mathcal{MC}\}$ with optimal value

$$\alpha_{opt} = \max_{B \in \mathcal{TA}} \langle C, B\rangle = \langle C, B^*\rangle$$

and upper bound

$$\alpha_{SDP} = \max_{A \in \mathcal{SR}} \langle C, A\rangle = \langle C, A^*\rangle$$

Theorem (Nesterov)

Let $C \succeq 0$, then $\alpha_{opt}(C) \geq \frac{2}{\pi}\alpha_{SDP}(C)$.

proof:

$$\alpha_{opt}(C) = \langle C, B^*\rangle \geq \langle C, \frac{2}{\pi}\arcsin[A^*]\rangle \geq \frac{2}{\pi}\langle C, A^*\rangle = \frac{2}{\pi}\alpha_{SDP}(C)$$

the second inequality holds because $f(t) = \arcsin(t) - t$ is positivity preserving

Sharpening of the bound

suppose $f(t) = \arcsin(t) - \gamma t$ is rank n positivity preserving

$$\alpha_{opt}(C) = \langle C, B^* \rangle \geq \langle C, \frac{2}{\pi} \arcsin[A^*] \rangle \geq \frac{2\gamma}{\pi} \langle C, A^* \rangle = \frac{2\gamma}{\pi} \alpha_{SDP}(C)$$

$$\max\{\gamma \mid f(t) = \arcsin(t) - \gamma t \text{ is rank } n \text{ positivity preserving}\}$$

is given by the first Gegenbauer coefficient of $\arcsin(t)$:

$$\gamma_{\max}(n) = \frac{\int_{-1}^{1} t \arcsin(t)\,(1-t^2)^{(n-3)/2}\,dt}{\int_{-1}^{1} t^2\,(1-t^2)^{(n-3)/2}\,dt} = \frac{\sqrt{\pi}\,\Gamma(\frac{n}{2}+1)\,\Gamma(n-1)}{2^{n-2}\,\Gamma(\frac{n-1}{2})\,\Gamma^2(\frac{n+1}{2})}$$

for $n = 1, 2, 3, \ldots$ we get $\frac{\pi}{2}, \frac{4}{\pi}, \frac{3\pi}{8}, \frac{32}{9\pi}, \frac{45\pi}{128}, \frac{256}{75\pi}, \ldots$

with recursion $\gamma_{\max}(n+1) = \gamma_{\max}^{-1}(n)\frac{n+1}{n}$

for big n $\gamma_{\max}(n) \approx 1 + \frac{1}{2n}$

Further properties of $\mathcal{TA}$

Theorem (Hirschfeld)

All 0,1,2-dimensional faces of $\mathcal{MC}$ are also faces of $\mathcal{TA}$.

Theorem (Hirschfeld)

Suppose that $g(t) = f^{-1}(\lambda f(t)) = \sin(\lambda \arcsin(t))$, with $f(t) = \frac{2}{\pi}\arcsin(t)$, is positivity preserving. Then $\mathcal{TA}$ is star-like with centre I, and for every $B \in \mathcal{TA}$

$$f^{-1}[\lambda B + (1-\lambda)I] \succeq (1 - \sin\frac{\pi\lambda}{2})I.$$

Lemma

The Gegenbauer expansion coefficients of $g(t)$ for half-integer values of the parameter α are nonnegative, and hence $g(t)$ is positivity preserving.

Copositive cone

Definition

A real symmetric $n \times n$ matrix A is called copositive if $x^T Ax \geq 0$ for all $x \in \mathbb{R}^n_+$.
The set of all copositive matrices forms the copositive cone $\mathcal{C}^n$.

let $\mathcal{C}^n_{\mathbf{1}}$ be the compact set of all $A \in \mathcal{C}^n$ with $\operatorname{diag} A = \mathbf{1}$

Definition

We call a function $f : [-1, \infty) \to \mathbb{R}$ n-copositivity preserving if $f[A] \in \mathcal{C}^n$ for all $A \in \mathcal{C}^n_{\mathbf{1}}$ and copositivity preserving if it is n-copositivity preserving for all $n \geq 1$.

Problem: Describe the cone of (n-)copositivity preserving functions?

Partial results

for $n = 2$ the n-copositivity preserving functions are those satisfying $f(1) \geq 0$ and $f(a) \geq -f(1)$ for all $a \geq -1$

Theorem (Hoffman, Pereira 1973)

The function $f(t) = \min(t, 1)$ is copositivity preserving.

Lemma

Let f be copositivity preserving and let f' be nonnegative. Then $f + f'$ is also copositivity preserving. In particular, every nonnegative function is copositivity preserving.

odd powers $f(t) = t^{2k+1}$ not copositivity preserving for $k \geq 1$

Triangle-free polytope

the vertices of the maxcut polytope are the matrices in $\mathcal{SR}$ with ± 1 entries

Theorem (Haynsworth, Hoffman 1969)

Let A be a symmetric $n \times n$ matrix with ± 1 entries and $\operatorname{diag}(A) = \mathbf{1}$. Let $G(A)$ be the graph on n vertices which has an edge (i,j) if and only if $A_{ij} = -1$. Then $A \in \mathcal{C}_{\mathbf{1}}^n$ if and only if $G(A)$ is triangle-free.

let $\mathcal{TF}$ be the convex hull of all matrices A as in the theorem such that $G(A)$ is triangle-free, the triangle-free polytope

then

$$\mathcal{MC} \subset \mathcal{TF} \subset \mathcal{C}_{\mathbf{1}}^n$$

Trigonometric approximation

define $\mathcal{TR} = \frac{2}{\pi}\arcsin[\mathcal{C}^n_{\mathbf{1}} \cap [-1,1]^{n\times n}]$

Problem: Does the inclusion $\mathcal{TR} \subset \mathcal{TF}$ hold?

there are families of extreme elements of $\mathcal{C}^n$ which become faces of $\mathcal{TF}$ under the element-wise map $f(t) = \frac{2}{\pi}\arcsin(t)$

extreme elements of $\mathcal{C}^5$

$$\begin{pmatrix} 1 & -\cos\xi_4 & \cos(\xi_4+\xi_5) & \cos(\xi_2+\xi_3) & -\cos\xi_3 \\ -\cos\xi_4 & 1 & -\cos\xi_5 & \cos(\xi_1+\xi_5) & \cos(\xi_3+\xi_4) \\ \cos(\xi_4+\xi_5) & -\cos\xi_5 & 1 & -\cos\xi_1 & \cos(\xi_1+\xi_2) \\ \cos(\xi_2+\xi_3) & \cos(\xi_1+\xi_5) & -\cos\xi_1 & 1 & -\cos\xi_2 \\ -\cos\xi_3 & \cos(\xi_3+\xi_4) & \cos(\xi_1+\xi_2) & -\cos\xi_2 & 1 \end{pmatrix}$$

$$\mapsto \begin{pmatrix} 1 & 2\delta_4-1 & 1-2\delta_4-2\delta_5 & 1-2\delta_2-2\delta_3 & 2\delta_3-1 \\ 2\delta_4-1 & 1 & 2\delta_5-1 & 1-2\delta_1-2\delta_5 & 1-2\delta_3-2\delta_4 \\ 1-2\delta_4-2\delta_5 & 2\delta_5-1 & 1 & 2\delta_1-1 & 1-2\delta_1-2\delta_2 \\ 1-2\delta_2-2\delta_3 & 1-2\delta_1-2\delta_5 & 2\delta_1-1 & 1 & 2\delta_2-1 \\ 2\delta_3-1 & 1-2\delta_3-2\delta_4 & 1-2\delta_1-2\delta_2 & 2\delta_2-1 & 1 \end{pmatrix}$$

$\xi_i = \pi\delta_i,\ \delta_i > 0,\ \sum_i \delta_i < 1$

Thank you