Entrywise positivity preservers:

covariance estimation, symmetric function identities, novel graph invariant

LAMA Lecture – ILAS 2019, Rio

Apoorva Khare

IISc and APRG (Bangalore, India)

(Partly based on joint works with Alexander Belton, Dominique Guillot, Mihai Putinar, Bala Rajaratnam, and Terence Tao)

Introduction

Positivity (and preserving it) studied in many settings in the literature.

Introduction

Positivity (and preserving it) studied in many settings in the literature.

Different flavors of positivity:

- Positive semidefinite $N \times N$ real symmetric matrices:
  $u^T A u \geqslant 0 \ \forall u.$
  Equivalently: A has all non-negative eigenvalues, or all non-negative principal minors. (Examples: Correlation and covariance matrices.)

Introduction

Positivity (and preserving it) studied in many settings in the literature.

Different flavors of positivity:

- Positive semidefinite $N \times N$ real symmetric matrices:
  $u^T A u \geqslant 0 \ \forall u$.
  Equivalently: A has all non-negative eigenvalues, or all non-negative principal minors. (Examples: Correlation and covariance matrices.)
- Positive definite sequences/Toeplitz matrices (measures on S^1)
- Moment sequences/Hankel matrices (measures on $\mathbb{R}$)
- Hilbert space kernels
- Positive definite functions on metric spaces, topological (semi)groups

Introduction

Positivity (and preserving it) studied in many settings in the literature.

Different flavors of positivity:

- Positive semidefinite $N \times N$ real symmetric matrices:
  $u^T A u \geqslant 0 \; \forall u.$
  Equivalently: A has all non-negative eigenvalues, or all non-negative principal minors. (Examples: Correlation and covariance matrices.)
- Positive definite sequences/Toeplitz matrices (measures on S^1)
- Moment sequences/Hankel matrices (measures on $\mathbb{R}$)
- Hilbert space kernels
- Positive definite functions on metric spaces, topological (semi)groups

Question: Classify the positivity preservers in these settings.

Studied for the better part of a century.

Positivity and Analysis

Entrywise functions preserving positivity

Given $N \geqslant 1$ and $I \subset \mathbb{R}$, let $\mathbb{P}_N(I)$ denote the $N \times N$ positive semidefinite matrices, with entries in I. (Say $\mathbb{P}_N = \mathbb{P}_N(\mathbb{R})$.)

Problem: Given a function $f : I \to \mathbb{R}$, when is it true that

$$f[A] := (f(a_{ij})) \in \mathbb{P}_N \text{ for all } A \in \mathbb{P}_N(I)?$$

(Long history!)

Entrywise functions preserving positivity

Given $N \geqslant 1$ and $I \subset \mathbb{R}$, let $\mathbb{P}_N(I)$ denote the $N \times N$ positive semidefinite matrices, with entries in I. (Say $\mathbb{P}_N = \mathbb{P}_N(\mathbb{R})$.)

Problem: Given a function $f : I \to \mathbb{R}$, when is it true that

$$f[A] := (f(a_{ij})) \in \mathbb{P}_N \text{ for all } A \in \mathbb{P}_N(I)?$$

(Long history!) The *Hadamard product* (or Schur, or entrywise product) of two matrices is given by: $A \circ B = (a_{ij} b_{ij})$.

Schur Product Theorem (Schur, *J. Reine Angew. Math.* 1911)

If $A, B \in \mathbb{P}_N$, then $A \circ B \in \mathbb{P}_N$.

Entrywise functions preserving positivity

Given $N \geqslant 1$ and $I \subset \mathbb{R}$, let $\mathbb{P}_N(I)$ denote the $N \times N$ positive semidefinite matrices, with entries in I. (Say $\mathbb{P}_N = \mathbb{P}_N(\mathbb{R})$.)

Problem: Given a function $f : I \to \mathbb{R}$, when is it true that

$$f[A] := (f(a_{ij})) \in \mathbb{P}_N \text{ for all } A \in \mathbb{P}_N(I)?$$

(Long history!) The *Hadamard product* (or Schur, or entrywise product) of two matrices is given by: $A \circ B = (a_{ij} b_{ij})$.

Schur Product Theorem (Schur, *J. Reine Angew. Math.* 1911)

If $A, B \in \mathbb{P}_N$, then $A \circ B \in \mathbb{P}_N$.

Pólya–Szegö: As a consequence,

- $f(x) = x^2, x^3, \ldots, x^k$ preserves positivity on $\mathbb{P}_N$ for all N, k.

Entrywise functions preserving positivity

Given $N \geqslant 1$ and $I \subset \mathbb{R}$, let $\mathbb{P}_N(I)$ denote the $N \times N$ positive semidefinite matrices, with entries in I. (Say $\mathbb{P}_N = \mathbb{P}_N(\mathbb{R})$.)

Problem: Given a function $f : I \to \mathbb{R}$, when is it true that

$$f[A] := (f(a_{ij})) \in \mathbb{P}_N \text{ for all } A \in \mathbb{P}_N(I)?$$

(Long history!) The *Hadamard product* (or Schur, or entrywise product) of two matrices is given by: $A \circ B = (a_{ij} b_{ij})$.

Schur Product Theorem (Schur, *J. Reine Angew. Math.* 1911)

If $A, B \in \mathbb{P}_N$, then $A \circ B \in \mathbb{P}_N$.

Pólya–Szegö: As a consequence,

- $f(x) = x^2, x^3, \ldots, x^k$ preserves positivity on $\mathbb{P}_N$ for all N, k.
- $f(x) = \sum_{k=0}^{l} c_k x^k$ preserves positivity if $c_k \geqslant 0$.

Entrywise functions preserving positivity

Given $N \geqslant 1$ and $I \subset \mathbb{R}$, let $\mathbb{P}_N(I)$ denote the $N \times N$ positive semidefinite matrices, with entries in I. (Say $\mathbb{P}_N = \mathbb{P}_N(\mathbb{R})$.)

Problem: Given a function $f : I \to \mathbb{R}$, when is it true that

$$f[A] := (f(a_{ij})) \in \mathbb{P}_N \text{ for all } A \in \mathbb{P}_N(I)?$$

(Long history!) The *Hadamard product* (or Schur, or entrywise product) of two matrices is given by: $A \circ B = (a_{ij} b_{ij})$.

Schur Product Theorem (Schur, *J. Reine Angew. Math.* 1911)

If $A, B \in \mathbb{P}_N$, then $A \circ B \in \mathbb{P}_N$.

Pólya–Szegö: As a consequence,

- $f(x) = x^2, x^3, \dots, x^k$ preserves positivity on $\mathbb{P}_N$ for all N, k.
- $f(x) = \sum_{k=0}^{l} c_k x^k$ preserves positivity if $c_k \geqslant 0$.
- Taking limits: if $f(x) = \sum_{k=0}^{\infty} c_k x^k$ is convergent and $c_k \geqslant 0$, then $f[-]$ preserves positivity.

Entrywise functions preserving positivity

Given $N \geqslant 1$ and $I \subset \mathbb{R}$, let $\mathbb{P}_N(I)$ denote the $N \times N$ positive semidefinite matrices, with entries in I. (Say $\mathbb{P}_N = \mathbb{P}_N(\mathbb{R})$.)

Problem: Given a function $f : I \to \mathbb{R}$, when is it true that

$$f[A] := (f(a_{ij})) \in \mathbb{P}_N \text{ for all } A \in \mathbb{P}_N(I)?$$

(Long history!) The *Hadamard product* (or Schur, or entrywise product) of two matrices is given by: $A \circ B = (a_{ij}b_{ij})$.

Schur Product Theorem (Schur, *J. Reine Angew. Math.* 1911)

If $A, B \in \mathbb{P}_N$, then $A \circ B \in \mathbb{P}_N$.

Pólya–Szegö: As a consequence,

- $f(x) = x^2, x^3, \dots, x^k$ preserves positivity on $\mathbb{P}_N$ for all N, k.
- $f(x) = \sum_{k=0}^{l} c_k x^k$ preserves positivity if $c_k \geqslant 0$.
- Taking limits: if $f(x) = \sum_{k=0}^{\infty} c_k x^k$ is convergent and $c_k \geqslant 0$, then $f[-]$ preserves positivity.
- Anything else?

Schoenberg's theorem

Question (Pólya–Szegö, 1925): Anything else?

Schoenberg's theorem

Question (Pólya–Szegö, 1925): Anything else? Remarkably, the answer is **no**, if we want to preserve positivity in *all* dimensions.

Theorem (Schoenberg, *Duke Math. J.* 1942; Rudin, *Duke Math. J.* 1959)

Suppose $I = (-1,1)$ and $f : I \to \mathbb{R}$. The following are equivalent:

1. $f[A] \in \mathbb{P}_N$ *for all* $A \in \mathbb{P}_N(I)$ *and all* N.
2. f *is analytic on* I *and has nonnegative Maclaurin coefficients. In other words,* $f(x) = \sum_{k=0}^{\infty} c_k x^k$ *on* $(-1,1)$ *with all* $c_k \geqslant 0$.

Schoenberg's theorem

Question (Pólya–Szegö, 1925): Anything else? Remarkably, the answer is **no**, if we want to preserve positivity in *all* dimensions.

Theorem (Schoenberg, *Duke Math. J.* 1942; Rudin, *Duke Math. J.* 1959)

Suppose $I = (-1,1)$ and $f : I \to \mathbb{R}$. The following are equivalent:

1. $f[A] \in \mathbb{P}_N$ *for all* $A \in \mathbb{P}_N(I)$ *and all* N.
2. f *is analytic on* I *and has nonnegative Maclaurin coefficients. In other words,* $f(x) = \sum_{k=0}^{\infty} c_k x^k$ *on* $(-1,1)$ *with all* $c_k \geqslant 0$.

Such functions f are said to be **absolutely monotonic** on $(0,1)$.

Toeplitz and Hankel matrices

Motivations: Rudin was motivated by harmonic analysis and Fourier analysis on locally compact groups. On $G = S^1$, he studied preservers of *positive definite sequences* $(a_n)_{n \in \mathbb{Z}}$. This means the Toeplitz kernel $(a_{i-j})_{i,j \geqslant 0}$ is positive semidefinite.

Toeplitz and Hankel matrices

Motivations: Rudin was motivated by harmonic analysis and Fourier analysis on locally compact groups. On $G = S^1$, he studied preservers of *positive definite sequences* $(a_n)_{n \in \mathbb{Z}}$. This means the Toeplitz kernel $(a_{i-j})_{i,j \geqslant 0}$ is positive semidefinite.

- In [*Duke Math. J.* 1959] Rudin showed: f preserves positive definite sequences (Toeplitz matrices) if and only if f is absolutely monotonic. Suffices to work with measures with 3-point supports.

Toeplitz and Hankel matrices

Motivations: Rudin was motivated by harmonic analysis and Fourier analysis on locally compact groups. On $G = S^1$, he studied preservers of *positive definite sequences* $(a_n)_{n \in \mathbb{Z}}$. This means the Toeplitz kernel $(a_{i-j})_{i,j \geqslant 0}$ is positive semidefinite.

- In [*Duke Math. J.* 1959] Rudin showed: f preserves positive definite sequences (Toeplitz matrices) if and only if f is absolutely monotonic. Suffices to work with measures with 3-point supports.

- Important parallel notion: **moment sequences**.
  Given positive measures μ on $[-1,1]$, with moment sequences
  $$\mathbf{s}(\mu) := (s_k(\mu))_{k \geqslant 0}, \qquad \text{where } s_k(\mu) := \int_{\mathbb{R}} x^k \, d\mu,$$
  classify the moment-sequence transformers: $f(s_k(\mu)) = s_k(\sigma_\mu), \ \forall k \geqslant 0$.

Toeplitz and Hankel matrices

Motivations: Rudin was motivated by harmonic analysis and Fourier analysis on locally compact groups. On $G = S^1$, he studied preservers of *positive definite sequences* $(a_n)_{n \in \mathbb{Z}}$. This means the Toeplitz kernel $(a_{i-j})_{i,j \geqslant 0}$ is positive semidefinite.

- In [*Duke Math. J.* 1959] Rudin showed: f preserves positive definite sequences (Toeplitz matrices) if and only if f is absolutely monotonic. Suffices to work with measures with 3-point supports.
- Important parallel notion: **moment sequences**.
  Given positive measures μ on $[-1,1]$, with moment sequences
  $$\mathbf{s}(\mu) := (s_k(\mu))_{k \geqslant 0}, \qquad \text{where } s_k(\mu) := \int_{\mathbb{R}} x^k \, d\mu,$$
  classify the moment-sequence transformers: $f(s_k(\mu)) = s_k(\sigma_\mu), \ \forall k \geqslant 0.$
- With Belton–Guillot–Putinar $\rightsquigarrow$ a parallel result to Rudin:

Toeplitz and Hankel matrices (cont.)

Let $0 < \rho \leqslant \infty$ be a scalar, and set $I = (-\rho, \rho)$.

Theorem (Rudin, *Duke Math. J.* 1959)

Given a function $f : I \to \mathbb{R}$, the following are equivalent:

1. $f[-]$ *preserves the set of positive definite sequences with entries in* I.
2. $f[-]$ *preserves positivity on Toeplitz matrices of all sizes and rank* $\leqslant 3$.

Toeplitz and Hankel matrices (cont.)

Let $0 < \rho \leqslant \infty$ be a scalar, and set $I = (-\rho, \rho)$.

Theorem (Rudin, *Duke Math. J.* 1959)

Given a function $f : I \to \mathbb{R}$, the following are equivalent:

1. $f[-]$ *preserves the set of positive definite sequences with entries in* I.
2. $f[-]$ *preserves positivity on Toeplitz matrices of all sizes and rank* $\leqslant 3$.
3. f *is analytic on* I *and has nonnegative Maclaurin coefficients.*
   In other words, $f(x) = \sum_{k=0}^{\infty} c_k x^k$ *on* $(-1, 1)$ *with all* $c_k \geqslant 0$.

Toeplitz and Hankel matrices (cont.)

Let $0 < \rho \leqslant \infty$ be a scalar, and set $I = (-\rho, \rho)$.

Theorem (Rudin, *Duke Math. J.* 1959)

Given a function $f : I \to \mathbb{R}$, the following are equivalent:

1. *$f[-]$ preserves the set of positive definite sequences with entries in I.*
2. *$f[-]$ preserves positivity on Toeplitz matrices of all sizes and rank $\leqslant 3$.*
3. *f is analytic on I and has nonnegative Maclaurin coefficients.*
   In other words, $f(x) = \sum_{k=0}^{\infty} c_k x^k$ on $(-1,1)$ with all $c_k \geqslant 0$.

Theorem (Belton–Guillot–K.–Putinar, 2016)

Given a function $f : I \to \mathbb{R}$, the following are equivalent:

1. *$f[-]$ preserves the set of moment sequences with entries in I.*
2. *$f[-]$ preserves positivity on Hankel matrices of all sizes and rank $\leqslant 3$.*
3. *f is analytic on I and has nonnegative Maclaurin coefficients.*

Positive semidefinite kernels

- These two results greatly weaken the hypotheses of Schoenberg's theorem – only need to consider positive semidefinite matrices of rank $\leqslant 3$.
- Note, such matrices are precisely the Gram matrices of vectors in a 3-dimensional Hilbert space. Hence Rudin (essentially) showed:

  Let $\mathcal{H}$ be a real Hilbert space of dimension $\geqslant 3$. If $f[-]$ preserves positivity on all Gram matrices in $\mathcal{H}$, then f is a power series on $\mathbb{R}$ with non-negative Maclaurin coefficients.

Positive semidefinite kernels

- These two results greatly weaken the hypotheses of Schoenberg's theorem – only need to consider positive semidefinite matrices of rank $\leqslant 3$.

- Note, such matrices are precisely the Gram matrices of vectors in a 3-dimensional Hilbert space. Hence Rudin (essentially) showed:

  Let $\mathcal{H}$ be a real Hilbert space of dimension $\geqslant 3$. If $f[-]$ preserves positivity on all Gram matrices in $\mathcal{H}$, then f is a power series on $\mathbb{R}$ with non-negative Maclaurin coefficients.

- But such functions are precisely the *positive semidefinite kernels on $\mathcal{H}$*! (Results of Pinkus et al.) Such kernels are important in modern day machine learning, via RKHS.

Positive semidefinite kernels

- These two results greatly weaken the hypotheses of Schoenberg's theorem – only need to consider positive semidefinite matrices of rank $\leqslant 3$.
- Note, such matrices are precisely the Gram matrices of vectors in a 3-dimensional Hilbert space. Hence Rudin (essentially) showed:

  Let $\mathcal{H}$ be a real Hilbert space of dimension $\geqslant 3$. If $f[-]$ preserves positivity on all Gram matrices in $\mathcal{H}$, then f is a power series on $\mathbb{R}$ with non-negative Maclaurin coefficients.
- But such functions are precisely the *positive semidefinite kernels on $\mathcal{H}$*! (Results of Pinkus et al.) Such kernels are important in modern day machine learning, via RKHS.
- Thus, Rudin (1959) classified positive semidefinite kernels on $\mathbb{R}^3$, which is relevant in machine learning. (Now also via our parallel 'Hankel' result.)

Schoenberg's theorem in several variables

Let $I = (-\rho, \rho)$ for some $0 < \rho \leqslant \infty$ as above. Also fix $m \geqslant 1$.
Given matrices $A_1, \dots, A_m \in \mathbb{P}_N(I)$ and $f : I^m \to \mathbb{R}$, define

$$f[A_1, \dots, A_m]_{ij} := f(a_{ij}^{(1)}, \dots, a_{ij}^{(m)}), \qquad \forall i, j = 1, \dots, N.$$

Theorem (FitzGerald–Micchelli–Pinkus, *Linear Alg. Appl.* 1995)

Given $f : \mathbb{R}^m \to \mathbb{R}$, the following are equivalent:

Schoenberg's theorem in several variables

Let $I = (-\rho, \rho)$ for some $0 < \rho \leqslant \infty$ as above. Also fix $m \geqslant 1$.
Given matrices $A_1, \dots, A_m \in \mathbb{P}_N(I)$ and $f : I^m \to \mathbb{R}$, define

$$f[A_1, \dots, A_m]_{ij} := f(a_{ij}^{(1)}, \dots, a_{ij}^{(m)}), \qquad \forall i, j = 1, \dots, N.$$

Theorem (FitzGerald–Micchelli–Pinkus, *Linear Alg. Appl.* 1995)

Given $f : \mathbb{R}^m \to \mathbb{R}$, the following are equivalent:

1. $f[A_1, \dots, A_m] \in \mathbb{P}_N$ *for all* $A_j \in \mathbb{P}_N(I)$ *and all* N.
2. *The function f is real entire and absolutely monotonic: for all* $\mathbf{x} \in \mathbb{R}^m$,
$$f(\mathbf{x}) = \sum_{\alpha \in \mathbb{Z}_+^m} c_\alpha \mathbf{x}^\alpha, \qquad \text{where } c_\alpha \geqslant 0 \ \forall \alpha \in \mathbb{Z}_+^m.$$

($(2) \Rightarrow (1)$ by Schur Product Theorem.)

Schoenberg's theorem in several variables

Let $I = (-\rho, \rho)$ for some $0 < \rho \leqslant \infty$ as above. Also fix $m \geqslant 1$.
Given matrices $A_1, \dots, A_m \in \mathbb{P}_N(I)$ and $f : I^m \to \mathbb{R}$, define

$$f[A_1, \dots, A_m]_{ij} := f(a_{ij}^{(1)}, \dots, a_{ij}^{(m)}), \qquad \forall i, j = 1, \dots, N.$$

Theorem (FitzGerald–Micchelli–Pinkus, *Linear Alg. Appl.* 1995)

Given $f : \mathbb{R}^m \to \mathbb{R}$, the following are equivalent:

1. $f[A_1, \dots, A_m] \in \mathbb{P}_N$ *for all* $A_j \in \mathbb{P}_N(I)$ *and all* N.
2. *The function f is real entire and absolutely monotonic: for all* $\mathbf{x} \in \mathbb{R}^m$,
$$f(\mathbf{x}) = \sum_{\alpha \in \mathbb{Z}_+^m} c_\alpha \mathbf{x}^\alpha, \qquad \text{where } c_\alpha \geqslant 0 \ \forall \alpha \in \mathbb{Z}_+^m.$$

($(2) \Rightarrow (1)$ by Schur Product Theorem.) The test set can again be reduced:

Theorem (Belton–Guillot–K.–Putinar, 2016)

The above two hypotheses are further equivalent to:

3. $f[-]$ *preserves positivity on m-tuples of Hankel matrices of rank* $\leqslant 3$.

Positivity and Metric geometry

Distance geometry

How did the study of positivity and its preservers begin?

Distance geometry

How did the study of positivity and its preservers begin?

In the 1900s, the notion of a *metric space* emerged from the works of Fréchet and Hausdorff. . .

- Now ubiquitous in science (mathematics, physics, economics, statistics, computer science. . .).

Distance geometry

How did the study of positivity and its preservers begin?

In the 1900s, the notion of a *metric space* emerged from the works of Fréchet and Hausdorff...

- Now ubiquitous in science (mathematics, physics, economics, statistics, computer science...).
- Fréchet [*Math. Ann.* 1910]. If (X, d) is a metric space with $|X| = n + 1$, then (X, d) isometrically embeds into $(\mathbb{R}^n, \ell_\infty)$.
- This avenue of work led to the exploration of metric space embeddings.

  Natural question: *Which metric spaces isometrically embed into* ***Euclidean space****?*

Euclidean metric spaces and positive matrices

Which metric spaces isometrically embed into a Euclidean space?

Euclidean metric spaces and positive matrices

Which metric spaces isometrically embed into a Euclidean space?

- Menger [*Amer. J. Math.* 1931] and Fréchet [*Ann. of Math.* 1935] provided characterizations.

Euclidean metric spaces and positive matrices

Which metric spaces isometrically embed into a Euclidean space?

- Menger [*Amer. J. Math.* 1931] and Fréchet [*Ann. of Math.* 1935] provided characterizations.
- Reformulated by Schoenberg, using... matrix positivity!

Theorem (Schoenberg, *Ann. of Math.* 1935)

Fix integers $n, r \geqslant 1$, *and a finite metric space* (X, d), *where* $X = \{x_0, \dots, x_n\}$. *Then* (X, d) *isometrically embeds into* $\mathbb{R}^r$ *(with the Euclidean distance/norm) but not into* $\mathbb{R}^{r-1}$ *if and only if the* $n \times n$ *matrix*

$$A := (d(x_0, x_i)^2 + d(x_0, x_j)^2 - d(x_i, x_j)^2)_{i,j=1}^n$$

is positive semidefinite of rank r.

This is how Schoenberg connected metric geometry and matrix positivity.

Distance transforms: positive definite functions

- In the preceding result, the matrix

  $A = (d(x_0, x_i)^2 + d(x_0, x_j)^2 - d(x_i, x_j)^2)_{i,j=1}^n$ is positive semidefinite,

  if and only if the matrix $A'_{(n+1)\times(n+1)} := (-d(x_i, x_j)^2)_{i,j=0}^n$ is

  conditionally positive semidefinite: $u^T A' u \geqslant 0$ whenever $\sum_{j=0}^n u_j = 0$.

- Early instance of how (conditionally) positive matrices emerged from metric geometry.

Distance transforms: positive definite functions

- In the preceding result, the matrix

  $A = (d(x_0, x_i)^2 + d(x_0, x_j)^2 - d(x_i, x_j)^2)_{i,j=1}^n$ is positive semidefinite,

  if and only if the matrix $A'_{(n+1)\times(n+1)} := (-d(x_i, x_j)^2)_{i,j=0}^n$ is

  conditionally positive semidefinite: $u^T A' u \geqslant 0$ whenever $\sum_{j=0}^n u_j = 0$.
- Early instance of how (conditionally) positive matrices emerged from metric geometry.

Now we move to *transforms* of positive matrices. Note that:

- Applying the function $-x^2$ *entrywise* sends any distance matrix from Euclidean space, to a conditionally positive semidefinite matrix A'.

Distance transforms: positive definite functions

- In the preceding result, the matrix

  $A = (d(x_0,x_i)^2 + d(x_0,x_j)^2 - d(x_i,x_j)^2)_{i,j=1}^n$ is positive semidefinite,

  if and only if the matrix $A'_{(n+1)\times(n+1)} := (-d(x_i,x_j)^2)_{i,j=0}^n$ is

  conditionally positive semidefinite: $u^T A' u \geqslant 0$ whenever $\sum_{j=0}^n u_j = 0$.
- Early instance of how (conditionally) positive matrices emerged from metric geometry.

Now we move to *transforms* of positive matrices. Note that:

- Applying the function $-x^2$ *entrywise* sends any distance matrix from Euclidean space, to a conditionally positive semidefinite matrix A'.
- Similarly, which entrywise maps send distance matrices to positive semidefinite matrices?

Distance transforms: positive definite functions

- In the preceding result, the matrix

  $A = (d(x_0,x_i)^2 + d(x_0,x_j)^2 - d(x_i,x_j)^2)_{i,j=1}^n$ is positive semidefinite,

  if and only if the matrix $A'_{(n+1)\times(n+1)} := (-d(x_i,x_j)^2)_{i,j=0}^n$ is

  conditionally positive semidefinite: $u^T A' u \geqslant 0$ whenever $\sum_{j=0}^n u_j = 0$.
- Early instance of how (conditionally) positive matrices emerged from metric geometry.

Now we move to *transforms* of positive matrices. Note that:

- Applying the function $-x^2$ *entrywise* sends any distance matrix from Euclidean space, to a conditionally positive semidefinite matrix A'.
- Similarly, which entrywise maps send distance matrices to positive semidefinite matrices? Such maps are called *positive definite functions*.

Distance transforms: positive definite functions

- In the preceding result, the matrix

  $A = (d(x_0, x_i)^2 + d(x_0, x_j)^2 - d(x_i, x_j)^2)_{i,j=1}^n$ is positive semidefinite,

  if and only if the matrix $A'_{(n+1)\times(n+1)} := (-d(x_i, x_j)^2)_{i,j=0}^n$ is

  conditionally positive semidefinite: $u^T A' u \geqslant 0$ whenever $\sum_{j=0}^n u_j = 0$.
- Early instance of how (conditionally) positive matrices emerged from metric geometry.

Now we move to *transforms* of positive matrices. Note that:

- Applying the function $-x^2$ *entrywise* sends any distance matrix from Euclidean space, to a conditionally positive semidefinite matrix A'.
- Similarly, which entrywise maps send distance matrices to positive semidefinite matrices? Such maps are called *positive definite functions*.
- Schoenberg was interested in embedding metric spaces into Euclidean spheres. This embeddability turns out to involve a *single* p.d. function!

Distance transforms: positive definite functions

- In the preceding result, the matrix

  $A = (d(x_0, x_i)^2 + d(x_0, x_j)^2 - d(x_i, x_j)^2)_{i,j=1}^n$ is positive semidefinite,

  if and only if the matrix $A'_{(n+1)\times(n+1)} := (-d(x_i, x_j)^2)_{i,j=0}^n$ is

  conditionally positive semidefinite: $u^T A' u \geqslant 0$ whenever $\sum_{j=0}^n u_j = 0$.
- Early instance of how (conditionally) positive matrices emerged from metric geometry.

Now we move to *transforms* of positive matrices. Note that:

- Applying the function $-x^2$ *entrywise* sends any distance matrix from Euclidean space, to a conditionally positive semidefinite matrix A'.
- Similarly, which entrywise maps send distance matrices to positive semidefinite matrices? Such maps are called *positive definite functions*.
- Schoenberg was interested in embedding metric spaces into Euclidean spheres. This embeddability turns out to involve a *single* p.d. function! This is the cosine function.

Positive definite functions on spheres

Notice that the Hilbert sphere S^∞ (hence every subspace such as S^{r-1}) has a rotation-invariant distance – *arc-length* along a great circle:

$$d(x,y) := \sphericalangle(x,y) = \arccos\langle x,y\rangle, \qquad x,y \in S^\infty.$$

Now applying $\cos[-]$ entrywise to any distance matrix on S^∞ yields:

$$\cos[(d(x_i,x_j))_{i,j\geqslant 0}] = (\langle x_i,x_j\rangle)_{i,j\geqslant 0},$$

and this is a Gram matrix, so $\cos(\cdot)$ is positive definite on S^∞.

Positive definite functions on spheres

Notice that the Hilbert sphere S^∞ (hence every subspace such as S^{r-1}) has a rotation-invariant distance – *arc-length* along a great circle:

$$d(x,y) := \sphericalangle(x,y) = \arccos\langle x, y\rangle, \qquad x, y \in S^\infty.$$

Now applying $\cos[-]$ entrywise to any distance matrix on S^∞ yields:

$$\cos[(d(x_i,x_j))_{i,j\geqslant 0}] = (\langle x_i, x_j\rangle)_{i,j\geqslant 0},$$

and this is a Gram matrix, so $\cos(\cdot)$ is positive definite on S^∞.

Schoenberg then classified *all* continuous f such that $f \circ \cos(\cdot)$ is p.d.:

Theorem (Schoenberg, *Duke Math. J.* 1942)

Suppose $f : [-1,1] \to \mathbb{R}$ is continuous, and $r \geqslant 2$. Then $f(\cos \cdot)$ is positive definite on the unit sphere $S^{r-1} \subset \mathbb{R}^r$ if and only if

$$f(\cdot) = \sum_{k\geqslant 0} a_k C_k^{(\frac{r-2}{2})}(\cdot) \qquad \text{for some } a_k \geqslant 0,$$

where $C_k^{(\lambda)}(\cdot)$ are the ultraspherical / Gegenbauer / Chebyshev polynomials.

Positive definite functions on spheres

Notice that the Hilbert sphere S^∞ (hence every subspace such as S^{r-1}) has a rotation-invariant distance – *arc-length* along a great circle:

$$d(x,y) := \sphericalangle(x,y) = \arccos\langle x,y\rangle, \qquad x,y \in S^\infty.$$

Now applying $\cos[-]$ entrywise to any distance matrix on S^∞ yields:

$$\cos[(d(x_i,x_j))_{i,j\geqslant 0}] = (\langle x_i,x_j\rangle)_{i,j\geqslant 0},$$

and this is a Gram matrix, so $\cos(\cdot)$ is positive definite on S^∞.

Schoenberg then classified *all* continuous f such that $f \circ \cos(\cdot)$ is p.d.:

Theorem (Schoenberg, *Duke Math. J.* 1942)

Suppose $f : [-1,1] \to \mathbb{R}$ is continuous, and $r \geqslant 2$. Then $f(\cos\cdot)$ is positive definite on the unit sphere $S^{r-1} \subset \mathbb{R}^r$ if and only if

$$f(\cdot) = \sum_{k\geqslant 0} a_k C_k^{(\frac{r-2}{2})}(\cdot) \qquad \text{for some } a_k \geqslant 0,$$

where $C_k^{(\lambda)}(\cdot)$ are the ultraspherical / Gegenbauer / Chebyshev polynomials.

Also follows from Bochner's work on compact homogeneous spaces [*Ann. of Math.* 1941] – but Schoenberg proved it directly with less 'heavy' machinery.

From spheres to correlation matrices

- Any Gram matrix of vectors $x_j \in S^{r-1}$ is the same as a rank $\leqslant r$ correlation matrix $A = (a_{ij})_{i,j=1}^n$, i.e.,

$$A = \begin{pmatrix} 1 & & * & \\ & 1 & & \\ * & & 1 & \\ & & & 1 \end{pmatrix} = \begin{pmatrix} — & x_1^T & — \\ — & x_2^T & — \\ & \vdots & \\ — & x_n^T & — \end{pmatrix} \begin{pmatrix} | & | & & | \\ x_1 & x_2 & \dots & x_n \\ | & | & & | \end{pmatrix} = (\langle x_i, x_j \rangle)_{i,j=1}^n.$$

From spheres to correlation matrices

- Any Gram matrix of vectors $x_j \in S^{r-1}$ is the same as a rank $\leqslant r$ correlation matrix $A = (a_{ij})_{i,j=1}^n$, i.e.,

$$A = \begin{pmatrix} 1 & & * & \\ & 1 & & \\ & * & 1 & \\ & & & 1 \end{pmatrix} = \begin{pmatrix} — & x_1^T & — \\ — & x_2^T & — \\ & \vdots & \\ — & x_n^T & — \end{pmatrix} \begin{pmatrix} | & | & & | \\ x_1 & x_2 & \dots & x_n \\ | & | & & | \end{pmatrix} = (\langle x_i, x_j \rangle)_{i,j=1}^n.$$

- So,

$$\begin{aligned} f(\cos \cdot) \text{ positive definite on } S^{r-1} \quad &\Longleftrightarrow \quad (f(\cos d(x_i, x_j)))_{i,j=1}^n \in \mathbb{P}_n \\ &\Longleftrightarrow \quad (f(\langle x_i, x_j \rangle))_{i,j=1}^n \in \mathbb{P}_n \\ &\Longleftrightarrow \quad (f(a_{ij}))_{i,j=1}^n \in \mathbb{P}_n \ \forall n \geqslant 1, \end{aligned}$$

From spheres to correlation matrices

- Any Gram matrix of vectors $x_j \in S^{r-1}$ is the same as a rank $\leqslant r$ correlation matrix $A = (a_{ij})_{i,j=1}^n$, i.e.,

$$A = \begin{pmatrix} 1 & & * & \\ & 1 & & \\ * & & 1 & \\ & & & 1 \end{pmatrix} = \begin{pmatrix} - & x_1^T & - \\ - & x_2^T & - \\ & \vdots & \\ - & x_n^T & - \end{pmatrix} \begin{pmatrix} | & | & & | \\ x_1 & x_2 & \dots & x_n \\ | & | & & | \end{pmatrix} = (\langle x_i, x_j \rangle)_{i,j=1}^n.$$

- So,

$$\begin{aligned} f(\cos \cdot) \text{ positive definite on } S^{r-1} \quad &\iff \quad (f(\cos d(x_i, x_j)))_{i,j=1}^n \in \mathbb{P}_n \\ &\iff \quad (f(\langle x_i, x_j \rangle))_{i,j=1}^n \in \mathbb{P}_n \\ &\iff \quad (f(a_{ij}))_{i,j=1}^n \in \mathbb{P}_n \ \forall n \geqslant 1, \end{aligned}$$

i.e., f preserves positivity on correlation matrices of rank $\leqslant r$.

From spheres to correlation matrices

- Any Gram matrix of vectors $x_j \in S^{r-1}$ is the same as a rank $\leqslant r$ correlation matrix $A = (a_{ij})_{i,j=1}^n$, i.e.,

$$A = \begin{pmatrix} 1 & & * & \\ & 1 & & \\ * & & 1 & \\ & & & 1 \end{pmatrix} = \begin{pmatrix} — & x_1^T & — \\ — & x_2^T & — \\ & \vdots & \\ — & x_n^T & — \end{pmatrix} \begin{pmatrix} | & | & & | \\ x_1 & x_2 & \dots & x_n \\ | & | & & | \end{pmatrix} = (\langle x_i, x_j \rangle)_{i,j=1}^n.$$

- So,

$$\begin{aligned} f(\cos \cdot) \text{ positive definite on } S^{r-1} \quad & \Longleftrightarrow \quad (f(\cos d(x_i, x_j)))_{i,j=1}^n \in \mathbb{P}_n \\ & \Longleftrightarrow \quad (f(\langle x_i, x_j \rangle))_{i,j=1}^n \in \mathbb{P}_n \\ & \Longleftrightarrow \quad (f(a_{ij}))_{i,j=1}^n \in \mathbb{P}_n \ \forall n \geqslant 1, \end{aligned}$$

  i.e., f preserves positivity on correlation matrices of rank $\leqslant r$.

- If instead $r = \infty$, such a result would classify the entrywise positivity preservers on all correlation matrices.

From spheres to correlation matrices

- Any Gram matrix of vectors $x_j \in S^{r-1}$ is the same as a rank $\leqslant r$ correlation matrix $A = (a_{ij})_{i,j=1}^n$, i.e.,

$$A = \begin{pmatrix} 1 & & * & \\ & 1 & & \\ * & & 1 & \\ & & & 1 \end{pmatrix} = \begin{pmatrix} — & x_1^T & — \\ — & x_2^T & — \\ & \vdots & \\ — & x_n^T & — \end{pmatrix} \begin{pmatrix} | & | & & | \\ x_1 & x_2 & \dots & x_n \\ | & | & & | \end{pmatrix} = (\langle x_i, x_j \rangle)_{i,j=1}^n.$$

- So,

$$\begin{aligned} f(\cos \cdot) \text{ positive definite on } S^{r-1} \quad &\iff \quad (f(\cos d(x_i, x_j)))_{i,j=1}^n \in \mathbb{P}_n \\ &\iff \quad (f(\langle x_i, x_j \rangle))_{i,j=1}^n \in \mathbb{P}_n \\ &\iff \quad (f(a_{ij}))_{i,j=1}^n \in \mathbb{P}_n \ \forall n \geqslant 1, \end{aligned}$$

  i.e., f preserves positivity on correlation matrices of rank $\leqslant r$.

- If instead $r = \infty$, such a result would classify the entrywise positivity preservers on all correlation matrices. Interestingly, 70 years later the subject has acquired renewed interest because of its immediate impact in high-dimensional covariance estimation, in several applied fields.

Schoenberg's theorem on positivity preservers

And indeed, Schoenberg did make the leap from S^{r-1} to S^∞:

Theorem (Schoenberg, *Duke Math. J.* 1942)

Suppose $f : [-1,1] \to \mathbb{R}$ *is continuous. Then* $f(\cos \cdot)$ *is positive definite on the Hilbert sphere* $S^\infty \subset \mathbb{R}^\infty = \ell^2$ *if and only if*

$$f(\cos\theta) = \sum_{k \geqslant 0} c_k \cos^k \theta,$$

where $c_k \geqslant 0 \ \forall k$ *are such that* $\sum_k c_k < \infty$.

Schoenberg's theorem on positivity preservers

And indeed, Schoenberg did make the leap from S^{r-1} to S^∞:

Theorem (Schoenberg, *Duke Math. J.* 1942)

Suppose $f : [-1,1] \to \mathbb{R}$ is continuous. Then $f(\cos \cdot)$ is positive definite on the Hilbert sphere $S^\infty \subset \mathbb{R}^\infty = \ell^2$ if and only if

$$f(\cos\theta) = \sum_{k \geqslant 0} c_k \cos^k \theta,$$

where $c_k \geqslant 0 \ \forall k$ are such that $\sum_k c_k < \infty$.

(By the Schur product theorem, $\cos^k \theta$ is positive definite on S^∞.)

Freeing this result from the sphere context, one obtains Schoenberg's theorem on entrywise positivity preservers.

Schoenberg's theorem on positivity preservers

And indeed, Schoenberg did make the leap from S^{r-1} to S^∞:

Theorem (Schoenberg, *Duke Math. J.* 1942)

Suppose $f : [-1,1] \to \mathbb{R}$ is continuous. Then $f(\cos \cdot)$ is positive definite on the Hilbert sphere $S^\infty \subset \mathbb{R}^\infty = \ell^2$ if and only if

$$f(\cos\theta) = \sum_{k \geqslant 0} c_k \cos^k \theta,$$

where $c_k \geqslant 0 \ \forall k$ are such that $\sum_k c_k < \infty$.

(By the Schur product theorem, $\cos^k \theta$ is positive definite on S^∞.)

Freeing this result from the sphere context, one obtains Schoenberg's theorem on entrywise positivity preservers.

For more information: *A panorama of positivity* – arXiv, Dec. 2018. (Survey, 80+ pp., by A. Belton, D. Guillot, A.K., and M. Putinar.)

Positivity and Statistics

Modern motivation: covariance estimation

Schoenberg's result has recently attracted renewed attention, owing to the statistics of big data.

Modern motivation: covariance estimation

Schoenberg's result has recently attracted renewed attention, owing to the statistics of big data.

- Major challenge in science: detect structure in vast amount of data.
- Covariance/correlation is a fundamental measure of dependence between random variables:

$$\Sigma = (\sigma_{ij})_{i,j=1}^{p}, \qquad \sigma_{ij} = \mathrm{Cov}(X_i, X_j) = \mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j].$$

Modern motivation: covariance estimation

Schoenberg's result has recently attracted renewed attention, owing to the statistics of big data.

- Major challenge in science: detect structure in vast amount of data.
- Covariance/correlation is a fundamental measure of dependence between random variables:
$$\Sigma = (\sigma_{ij})_{i,j=1}^p, \qquad \sigma_{ij} = \mathrm{Cov}(X_i, X_j) = \mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j].$$
- **Important question:** Estimate Σ from data $x_1, \dots, x_n \in \mathbb{R}^p$.

Modern motivation: covariance estimation

Schoenberg's result has recently attracted renewed attention, owing to the statistics of big data.

- Major challenge in science: detect structure in vast amount of data.
- Covariance/correlation is a fundamental measure of dependence between random variables:
$$\Sigma = (\sigma_{ij})_{i,j=1}^p, \qquad \sigma_{ij} = \mathrm{Cov}(X_i, X_j) = \mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j].$$
- **Important question:** Estimate Σ from data $x_1, \dots, x_n \in \mathbb{R}^p$.
- In modern-day settings (small samples, ultra-high dimension), covariance estimation can be very challenging.
- Classical estimators (e.g. sample covariance matrix (MLE)):
$$S = \frac{1}{n} \sum_{j=1}^n (x_j - \overline{x})(x_j - \overline{x})^T$$
perform poorly, are singular/ill-conditioned, etc.

Modern motivation: covariance estimation

Schoenberg's result has recently attracted renewed attention, owing to the statistics of big data.

- Major challenge in science: detect structure in vast amount of data.
- Covariance/correlation is a fundamental measure of dependence between random variables:
$$\Sigma = (\sigma_{ij})_{i,j=1}^p, \qquad \sigma_{ij} = \mathrm{Cov}(X_i, X_j) = \mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j].$$
- **Important question:** Estimate Σ from data $x_1, \ldots, x_n \in \mathbb{R}^p$.
- In modern-day settings (small samples, ultra-high dimension), covariance estimation can be very challenging.
- Classical estimators (e.g. sample covariance matrix (MLE)):
$$S = \frac{1}{n} \sum_{j=1}^n (x_j - \overline{x})(x_j - \overline{x})^T$$
perform poorly, are singular/ill-conditioned, etc.
- Require some form of *regularization* – and resulting matrix has to be positive semidefinite (in the parameter space) for applications.

Motivation from high-dimensional statistics

Graphical models: Connections between statistics and combinatorics.

Let $X_1, \dots, X_p$ be a collection of random variables.

- Very large vectors: rare that all X_j depend strongly on each other.
- Many variables are (conditionally) independent; not used in prediction.

Motivation from high-dimensional statistics

Graphical models: Connections between statistics and combinatorics.

Let $X_1, \ldots, X_p$ be a collection of random variables.

- Very large vectors: rare that all X_j depend strongly on each other.
- Many variables are (conditionally) independent; not used in prediction.
- Leverage the independence/conditional independence structure to reduce dimension – translates to zeros in covariance/inverse covariance matrix.

Motivation from high-dimensional statistics

Graphical models: Connections between statistics and combinatorics.

Let $X_1, \dots, X_p$ be a collection of random variables.

- Very large vectors: rare that all X_j depend strongly on each other.
- Many variables are (conditionally) independent; not used in prediction.
- Leverage the independence/conditional independence structure to reduce dimension – translates to zeros in covariance/inverse covariance matrix.
- **Modern approach:** Compressed sensing methods (Daubechies, Donoho, Candes, Tao, ...) use convex optimization to obtain a sparse estimate of Σ (e.g., ℓ^1-penalized likelihood methods).
- State-of-the-art for ~ 20 years.

  Works well for dimensions of a few thousands.

Motivation from high-dimensional statistics

Graphical models: Connections between statistics and combinatorics.

Let $X_1, \dots, X_p$ be a collection of random variables.

- Very large vectors: rare that all X_j depend strongly on each other.
- Many variables are (conditionally) independent; not used in prediction.
- Leverage the independence/conditional independence structure to reduce dimension – translates to zeros in covariance/inverse covariance matrix.
- **Modern approach:** Compressed sensing methods (Daubechies, Donoho, Candes, Tao, ...) use convex optimization to obtain a sparse estimate of Σ (e.g., ℓ^1-penalized likelihood methods).
- State-of-the-art for ~ 20 years.

  Works well for dimensions of a few thousands.
- Not scalable to modern-day problems with $100,000+$ variables (disease detection, climate sciences, finance...).

Thresholding and regularization

Thresholding covariance/correlation matrices

$$\text{True } \Sigma = \begin{pmatrix} 1 & 0.2 & 0 \\ 0.2 & 1 & 0.5 \\ 0 & 0.5 & 1 \end{pmatrix} \quad S = \begin{pmatrix} 0.95 & 0.18 & 0.02 \\ 0.18 & 0.96 & 0.47 \\ 0.02 & 0.47 & 0.98 \end{pmatrix}$$

Thresholding and regularization

Thresholding covariance/correlation matrices

$$\text{True } \Sigma = \begin{pmatrix} 1 & 0.2 & 0 \\ 0.2 & 1 & 0.5 \\ 0 & 0.5 & 1 \end{pmatrix} \qquad S = \begin{pmatrix} 0.95 & 0.18 & 0.02 \\ 0.18 & 0.96 & 0.47 \\ 0.02 & 0.47 & 0.98 \end{pmatrix}$$

Natural to *threshold* small entries (thinking the variables are independent):

$$\tilde{S} = \begin{pmatrix} 0.95 & 0.18 & \mathbf{0} \\ 0.18 & 0.96 & 0.47 \\ \mathbf{0} & 0.47 & 0.98 \end{pmatrix}$$

Thresholding and regularization

Thresholding covariance/correlation matrices

$$\text{True } \Sigma = \begin{pmatrix} 1 & 0.2 & 0 \\ 0.2 & 1 & 0.5 \\ 0 & 0.5 & 1 \end{pmatrix} \qquad S = \begin{pmatrix} 0.95 & 0.18 & 0.02 \\ 0.18 & 0.96 & 0.47 \\ 0.02 & 0.47 & 0.98 \end{pmatrix}$$

Natural to *threshold* small entries (thinking the variables are independent):

$$\tilde{S} = \begin{pmatrix} 0.95 & 0.18 & \mathbf{0} \\ 0.18 & 0.96 & 0.47 \\ \mathbf{0} & 0.47 & 0.98 \end{pmatrix}$$

Can be significant if $p = 1,000,000$ and only, say, $\sim 1\%$ of the entries of the true Σ are nonzero.

Entrywise functions – regularization

More generally, we could apply a function $f : \mathbb{R} \to \mathbb{R}$ to the elements of the matrix S – *regularization*:

Entrywise functions – regularization

More generally, we could apply a function $f : \mathbb{R} \to \mathbb{R}$ to the elements of the matrix S – *regularization*:

$$\widehat{\Sigma} = f[S] := \begin{pmatrix} f(\sigma_{11}) & f(\sigma_{12}) & \dots & f(\sigma_{1N}) \\ f(\sigma_{21}) & f(\sigma_{22}) & \dots & f(\sigma_{2N}) \\ \vdots & \vdots & \ddots & \vdots \\ f(\sigma_{N1}) & f(\sigma_{N2}) & \dots & f(\sigma_{NN}) \end{pmatrix}$$

(Example on previous slide is $f_\epsilon(x) = x \cdot \mathbf{1}_{|x|>\epsilon}$ for some $\epsilon > 0$.)

Entrywise functions – regularization

More generally, we could apply a function $f : \mathbb{R} \to \mathbb{R}$ to the elements of the matrix S – *regularization*:

$$\widehat{\Sigma} = f[S] := \begin{pmatrix} f(\sigma_{11}) & f(\sigma_{12}) & \dots & f(\sigma_{1N}) \\ f(\sigma_{21}) & f(\sigma_{22}) & \dots & f(\sigma_{2N}) \\ \vdots & \vdots & \ddots & \vdots \\ f(\sigma_{N1}) & f(\sigma_{N2}) & \dots & f(\sigma_{NN}) \end{pmatrix}$$

(Example on previous slide is $f_\epsilon(x) = x \cdot \mathbf{1}_{|x|>\epsilon}$ for some $\epsilon > 0$.)

- Highly scalable. Analysis on the cone – no optimization.

Entrywise functions – regularization

More generally, we could apply a function $f : \mathbb{R} \to \mathbb{R}$ to the elements of the matrix S – *regularization*:

$$\widehat{\Sigma} = f[S] := \begin{pmatrix} f(\sigma_{11}) & f(\sigma_{12}) & \dots & f(\sigma_{1N}) \\ f(\sigma_{21}) & f(\sigma_{22}) & \dots & f(\sigma_{2N}) \\ \vdots & \vdots & \ddots & \vdots \\ f(\sigma_{N1}) & f(\sigma_{N2}) & \dots & f(\sigma_{NN}) \end{pmatrix}$$

(Example on previous slide is $f_\epsilon(x) = x \cdot \mathbf{1}_{|x|>\epsilon}$ for some $\epsilon > 0$.)

- Highly scalable. Analysis on the cone – no optimization.
- Regularized matrix $f[S]$ further used in applications, where (estimates of) Σ required in procedures such as PCA, CCA, MANOVA, etc.

Entrywise functions – regularization

More generally, we could apply a function $f : \mathbb{R} \to \mathbb{R}$ to the elements of the matrix S – *regularization*:

$$\widehat{\Sigma} = f[S] := \begin{pmatrix} f(\sigma_{11}) & f(\sigma_{12}) & \dots & f(\sigma_{1N}) \\ f(\sigma_{21}) & f(\sigma_{22}) & \dots & f(\sigma_{2N}) \\ \vdots & \vdots & \ddots & \vdots \\ f(\sigma_{N1}) & f(\sigma_{N2}) & \dots & f(\sigma_{NN}) \end{pmatrix}$$

(Example on previous slide is $f_\epsilon(x) = x \cdot \mathbf{1}_{|x|>\epsilon}$ for some $\epsilon > 0$.)

- Highly scalable. Analysis on the cone – no optimization.
- Regularized matrix $f[S]$ further used in applications, where (estimates of) Σ required in procedures such as PCA, CCA, MANOVA, etc.
- **Question:** When does this procedure preserve positive (semi)definiteness? Critical for applications since $\Sigma \in \mathbb{P}_N$.

Entrywise functions – regularization

More generally, we could apply a function $f : \mathbb{R} \to \mathbb{R}$ to the elements of the matrix S – *regularization*:

$$\widehat{\Sigma} = f[S] := \begin{pmatrix} f(\sigma_{11}) & f(\sigma_{12}) & \dots & f(\sigma_{1N}) \\ f(\sigma_{21}) & f(\sigma_{22}) & \dots & f(\sigma_{2N}) \\ \vdots & \vdots & \ddots & \vdots \\ f(\sigma_{N1}) & f(\sigma_{N2}) & \dots & f(\sigma_{NN}) \end{pmatrix}$$

(Example on previous slide is $f_\epsilon(x) = x \cdot \mathbf{1}_{|x|>\epsilon}$ for some $\epsilon > 0$.)

- Highly scalable. Analysis on the cone – no optimization.
- Regularized matrix $f[S]$ further used in applications, where (estimates of) Σ required in procedures such as PCA, CCA, MANOVA, etc.
- **Question:** When does this procedure preserve positive (semi)definiteness? Critical for applications since $\Sigma \in \mathbb{P}_N$.

Problem: For what functions $f : \mathbb{R} \to \mathbb{R}$, does $f[-]$ preserve $\mathbb{P}_N$?

Preserving positivity in fixed dimension

Schoenberg's result characterizes functions preserving positivity for matrices of *all* dimensions: $f[A] \in \mathbb{P}_N$ for all $A \in \mathbb{P}_N$ and **all** N.

Preserving positivity in fixed dimension

Schoenberg's result characterizes functions preserving positivity for matrices of *all* dimensions: $f[A] \in \mathbb{P}_N$ for all $A \in \mathbb{P}_N$ and **all** N.

Similar/related problems studied by many others, including:

- Bharali, Bhatia, Christensen, FitzGerald, Helson, Hiai, Holtz,
- Horn, Jain, Kahane, Karlin, Katznelson, Loewner, Menegatto,
- Micchelli, Pinkus, Pólya, Ressel, Vasudeva, Willoughby, ...

Preserving positivity in fixed dimension

Schoenberg's result characterizes functions preserving positivity for matrices of *all* dimensions: $f[A] \in \mathbb{P}_N$ for all $A \in \mathbb{P}_N$ and **all** N.

Similar/related problems studied by many others, including:

- Bharali, Bhatia, Christensen, FitzGerald, Helson, Hiai, Holtz,
- Horn, Jain, Kahane, Karlin, Katznelson, Loewner, Menegatto,
- Micchelli, Pinkus, Pólya, Ressel, Vasudeva, Willoughby, ...

Preserving positivity for fixed N:

- Natural refinement of original problem of Schoenberg.
- In applications: dimension of the problem is known.
  Unnecessarily restrictive to preserve positivity in all dimensions.

Preserving positivity in fixed dimension

Schoenberg's result characterizes functions preserving positivity for matrices of *all* dimensions: $f[A] \in \mathbb{P}_N$ for all $A \in \mathbb{P}_N$ and **all** N.

Similar/related problems studied by many others, including:

- Bharali, Bhatia, Christensen, FitzGerald, Helson, Hiai, Holtz,
- Horn, Jain, Kahane, Karlin, Katznelson, Loewner, Menegatto,
- Micchelli, Pinkus, Pólya, Ressel, Vasudeva, Willoughby, ...

Preserving positivity for fixed N:

- Natural refinement of original problem of Schoenberg.
- In applications: dimension of the problem is known.
  Unnecessarily restrictive to preserve positivity in all dimensions.
- Known for $N = 2$ (Vasudeva, *IJPAM* 1979):

$$f \textit{ is nondecreasing and } f(x)f(y) \geqslant f(\sqrt{xy})^2 \textit{ on } (0,\infty).$$

Preserving positivity in fixed dimension

Schoenberg's result characterizes functions preserving positivity for matrices of *all* dimensions: $f[A] \in \mathbb{P}_N$ for all $A \in \mathbb{P}_N$ and **all** N.

Similar/related problems studied by many others, including:

- Bharali, Bhatia, Christensen, FitzGerald, Helson, Hiai, Holtz,
- Horn, Jain, Kahane, Karlin, Katznelson, Loewner, Menegatto,
- Micchelli, Pinkus, Pólya, Ressel, Vasudeva, Willoughby, ...

Preserving positivity for fixed N:

- Natural refinement of original problem of Schoenberg.
- In applications: dimension of the problem is known. Unnecessarily restrictive to preserve positivity in all dimensions.
- Known for $N = 2$ (Vasudeva, *IJPAM* 1979):

  f is nondecreasing and $f(x)f(y) \geqslant f(\sqrt{xy})^2$ on $(0, \infty)$.
- **Open** for $N \geqslant 3$.

Problems motivated by applications

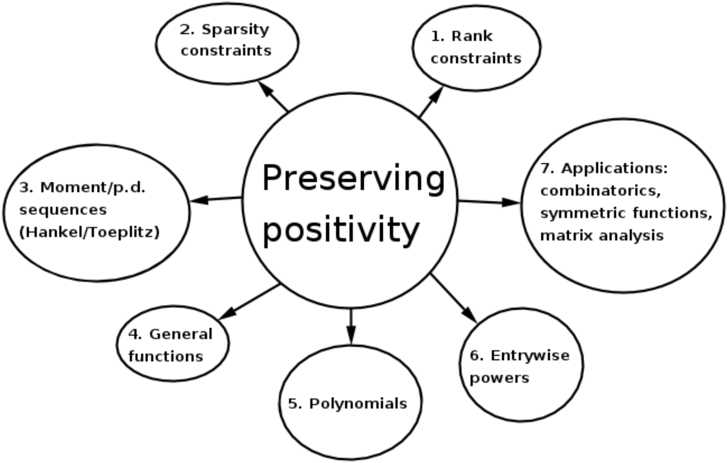
- We revisit this problem with modern applications in mind.

Problems motivated by applications

- We revisit this problem with modern applications in mind.
- Applications motivate many new exciting problems:

Problems motivated by applications

- We revisit this problem with modern applications in mind.
- Applications motivate many new exciting problems:



Positivity and Symmetric functions

Preserving positivity in fixed dimension

Question: Find a power series with a negative coefficient, which preserves positivity on $\mathbb{P}_N$ for some $N \geqslant 3$.

Preserving positivity in fixed dimension

Question: Find a power series with a negative coefficient, which preserves positivity on $\mathbb{P}_N$ for some $N \geqslant 3$. (**Open** since Schoenberg's *Duke* 1942 paper.)

Preserving positivity in fixed dimension

Question: Find a power series with a negative coefficient, which preserves positivity on $\mathbb{P}_N$ for some $N \geqslant 3$. (**Open** since Schoenberg's *Duke* 1942 paper.)

For fixed $N \geqslant 3$ and general f, only known necessary condition is due to Horn:

Theorem (Horn, *Trans. AMS* 1969; Guillot–K.–Rajaratnam, *Trans. AMS* 2017)

*Fix $I = (0, \rho)$ for $0 < \rho \leqslant \infty$, and $f : I \to \mathbb{R}$. Suppose $f[A] \in \mathbb{P}_N$ for all $A \in \mathbb{P}_N(I)$ Hankel of rank $\leqslant 2$, with N **fixed**.*

Preserving positivity in fixed dimension

Question: Find a power series with a negative coefficient, which preserves positivity on $\mathbb{P}_N$ for some $N \geqslant 3$. (**Open** since Schoenberg's *Duke* 1942 paper.)

For fixed $N \geqslant 3$ and general f, only known necessary condition is due to Horn:

Theorem (Horn, *Trans. AMS* 1969; Guillot–K.–Rajaratnam, *Trans. AMS* 2017)

Fix $I = (0, \rho)$ *for* $0 < \rho \leqslant \infty$, *and* $f : I \to \mathbb{R}$. *Suppose* $f[A] \in \mathbb{P}_N$ *for all* $A \in \mathbb{P}_N(I)$ *Hankel of rank* $\leqslant 2$, *with* N ***fixed***. *Then* $f \in C^{N-3}(I)$, *and*

$$f, f', f'', \cdots, f^{(N-3)} \geqslant 0 \text{ on } I.$$

Preserving positivity in fixed dimension

Question: Find a power series with a negative coefficient, which preserves positivity on $\mathbb{P}_N$ for some $N \geqslant 3$. (**Open** since Schoenberg's *Duke* 1942 paper.)

For fixed $N \geqslant 3$ and general f, only known necessary condition is due to Horn:

Theorem (Horn, *Trans. AMS* 1969; Guillot–K.–Rajaratnam, *Trans. AMS* 2017)

*Fix $I = (0, \rho)$ for $0 < \rho \leqslant \infty$, and $f : I \to \mathbb{R}$. Suppose $f[A] \in \mathbb{P}_N$ for all $A \in \mathbb{P}_N(I)$ Hankel of rank $\leqslant 2$, with N **fixed**. Then $f \in C^{N-3}(I)$, and*

$$f, f', f'', \cdots, f^{(N-3)} \geqslant 0 \text{ on } I.$$

If $f \in C^{N-1}(I)$ then this also holds for $f^{(N-2)}, f^{(N-1)}$.

- Implies Schoenberg–Rudin result for matrices with positive entries.

Preserving positivity in fixed dimension

Question: Find a power series with a negative coefficient, which preserves positivity on $\mathbb{P}_N$ for some $N \geqslant 3$. (**Open** since Schoenberg's *Duke* 1942 paper.)

For fixed $N \geqslant 3$ and general f, only known necessary condition is due to Horn:

Theorem (Horn, *Trans. AMS* 1969; Guillot–K.–Rajaratnam, *Trans. AMS* 2017)

*Fix $I = (0, \rho)$ for $0 < \rho \leqslant \infty$, and $f : I \to \mathbb{R}$. Suppose $f[A] \in \mathbb{P}_N$ for all $A \in \mathbb{P}_N(I)$ Hankel of rank $\leqslant 2$, with N **fixed**. Then $f \in C^{N-3}(I)$, and*

$$f, f', f'', \cdots, f^{(N-3)} \geqslant 0 \text{ on } I.$$

If $f \in C^{N-1}(I)$ then this also holds for $f^{(N-2)}, f^{(N-1)}$.

- Implies Schoenberg–Rudin result for matrices with positive entries.
- E.g., let $N \in \mathbb{N}$ and $c_0, \dots, c_{N-1} \neq 0$. If $f(z) = \sum_{j=0}^{N-1} c_j z^j + c_N z^N$ preserves positivity on $\mathbb{P}_N$, then $c_0, \dots, c_{N-1} > 0$.

Preserving positivity in fixed dimension

Question: Find a power series with a negative coefficient, which preserves positivity on $\mathbb{P}_N$ for some $N \geqslant 3$. (**Open** since Schoenberg's *Duke* 1942 paper.)

For fixed $N \geqslant 3$ and general f, only known necessary condition is due to Horn:

Theorem (Horn, *Trans. AMS* 1969; Guillot–K.–Rajaratnam, *Trans. AMS* 2017)

*Fix $I = (0, \rho)$ for $0 < \rho \leqslant \infty$, and $f : I \to \mathbb{R}$. Suppose $f[A] \in \mathbb{P}_N$ for all $A \in \mathbb{P}_N(I)$ Hankel of rank $\leqslant 2$, with N **fixed**. Then $f \in C^{N-3}(I)$, and*

$$f, f', f'', \cdots, f^{(N-3)} \geqslant 0 \text{ on } I.$$

If $f \in C^{N-1}(I)$ then this also holds for $f^{(N-2)}, f^{(N-1)}$.

- Implies Schoenberg–Rudin result for matrices with positive entries.
- E.g., let $N \in \mathbb{N}$ and $c_0, \dots, c_{N-1} \neq 0$. If $f(z) = \sum_{j=0}^{N-1} c_j z^j + c_N z^N$ preserves positivity on $\mathbb{P}_N$, then $c_0, \dots, c_{N-1} > 0$.

Can c_N be negative?

Preserving positivity in fixed dimension

Question: Find a power series with a negative coefficient, which preserves positivity on $\mathbb{P}_N$ for some $N \geqslant 3$. (**Open** since Schoenberg's *Duke* 1942 paper.)

For fixed $N \geqslant 3$ and general f, only known necessary condition is due to Horn:

Theorem (Horn, *Trans. AMS* 1969; Guillot–K.–Rajaratnam, *Trans. AMS* 2017)

Fix $I = (0, \rho)$ for $0 < \rho \leqslant \infty$, and $f : I \to \mathbb{R}$. Suppose $f[A] \in \mathbb{P}_N$ for all $A \in \mathbb{P}_N(I)$ Hankel of rank $\leqslant 2$, with N ***fixed****. Then $f \in C^{N-3}(I)$, and*

$$f, f', f'', \cdots, f^{(N-3)} \geqslant 0 \text{ on } I.$$

If $f \in C^{N-1}(I)$ then this also holds for $f^{(N-2)}, f^{(N-1)}$.

- Implies Schoenberg–Rudin result for matrices with positive entries.
- E.g., let $N \in \mathbb{N}$ and $c_0, \dots, c_{N-1} \neq 0$. If $f(z) = \sum_{j=0}^{N-1} c_j z^j + c_N z^N$ preserves positivity on $\mathbb{P}_N$, then $c_0, \dots, c_{N-1} > 0$.

Can c_N be negative? Sharp bound? (Not known to date.)

Polynomials preserving positivity in fixed dimension

More generally, the first N *nonzero* Maclaurin coefficients must be positive.
Can the next one be negative?

Polynomials preserving positivity in fixed dimension

More generally, the first N *nonzero* Maclaurin coefficients must be positive. Can the next one be negative?

Theorem (K.–Tao 2017; Belton–Guillot–K.–Putinar, *Adv. Math.* 2016)

Fix $\rho > 0$ and integers $0 \leqslant n_0 < \cdots < n_{N-1} < M$, and let

$$f(z) = \sum_{j=0}^{N-1} c_j z^{n_j} + c' z^M$$

be a polynomial with real coefficients.

Polynomials preserving positivity in fixed dimension

More generally, the first N *nonzero* Maclaurin coefficients must be positive. Can the next one be negative?

Theorem (K.–Tao 2017; Belton–Guillot–K.–Putinar, *Adv. Math.* 2016)

Fix $\rho > 0$ *and integers* $0 \leqslant n_0 < \cdots < n_{N-1} < M$*, and let*

$$f(z) = \sum_{j=0}^{N-1} c_j z^{n_j} + c' z^M$$

be a polynomial with real coefficients.

Then the following are equivalent.

1. $f[-]$ *preserves positivity on* $\mathbb{P}_N((0,\rho))$*.*
2. *The coefficients* c_j *satisfy either* $c_0, \dots, c_{N-1}, c' \geqslant 0$*,*

Polynomials preserving positivity in fixed dimension

More generally, the first N *nonzero* Maclaurin coefficients must be positive. Can the next one be negative?

Theorem (K.–Tao 2017; Belton–Guillot–K.–Putinar, *Adv. Math.* 2016)

Fix $\rho > 0$ and integers $0 \leqslant n_0 < \cdots < n_{N-1} < M$, and let

$$f(z) = \sum_{j=0}^{N-1} c_j z^{n_j} + c' z^M$$

be a polynomial with real coefficients.

Then the following are equivalent.

1. *$f[-]$ preserves positivity on $\mathbb{P}_N((0,\rho))$.*
2. *The coefficients c_j satisfy either $c_0, \dots, c_{N-1}, c' \geqslant 0$, or $c_0, \dots, c_{N-1} > 0$ and $c' \geqslant -\mathcal{C}^{-1}$, where*

$$\mathcal{C} := \sum_{j=0}^{N-1} \frac{\rho^{M-n_j}}{c_j} \prod_{i=0, i \neq j}^{N-1} \frac{(M-n_i)^2}{(n_j-n_i)^2}.$$

Polynomials preserving positivity in fixed dimension

More generally, the first N *nonzero* Maclaurin coefficients must be positive. Can the next one be negative?

Theorem (K.–Tao 2017; Belton–Guillot–K.–Putinar, *Adv. Math.* 2016)

Fix $\rho > 0$ *and integers* $0 \leqslant n_0 < \cdots < n_{N-1} < M$, *and let*

$$f(z) = \sum_{j=0}^{N-1} c_j z^{n_j} + c' z^M$$

be a polynomial with real coefficients.

Then the following are equivalent.

1. $f[-]$ *preserves positivity on* $\mathbb{P}_N((0,\rho))$.
2. *The coefficients* c_j *satisfy either* $c_0, \dots, c_{N-1}, c' \geqslant 0$, *or* $c_0, \dots, c_{N-1} > 0$ *and* $c' \geqslant -\mathcal{C}^{-1}$, *where*
$$\mathcal{C} := \sum_{j=0}^{N-1} \frac{\rho^{M-n_j}}{c_j} \prod_{i=0, i \neq j}^{N-1} \frac{(M-n_i)^2}{(n_j-n_i)^2}.$$
3. $f[-]$ *preserves positivity on rank-one Hankel matrices in* $\mathbb{P}_N((0,\rho))$.

Consequences

1. Quantitative version of Schoenberg's theorem in fixed dimension: polynomials that preserve positivity on $\mathbb{P}_N$, but not on $\mathbb{P}_{N+1}$.

Consequences

1. Quantitative version of Schoenberg's theorem in fixed dimension: polynomials that preserve positivity on $\mathbb{P}_N$, but not on $\mathbb{P}_{N+1}$.
2. When $M = N$, the theorem provides an exact characterization of polynomials of degree at most N that preserve positivity on $\mathbb{P}_N$.

Consequences

1. Quantitative version of Schoenberg's theorem in fixed dimension: polynomials that preserve positivity on $\mathbb{P}_N$, but not on $\mathbb{P}_{N+1}$.
2. When $M = N$, the theorem provides an exact characterization of polynomials of degree at most N that preserve positivity on $\mathbb{P}_N$.
3. The result holds verbatim for sums of *real powers*.

Consequences

1. Quantitative version of Schoenberg's theorem in fixed dimension: polynomials that preserve positivity on $\mathbb{P}_N$, but not on $\mathbb{P}_{N+1}$.
2. When $M = N$, the theorem provides an exact characterization of polynomials of degree at most N that preserve positivity on $\mathbb{P}_N$.
3. The result holds verbatim for sums of *real powers*.
4. Surprisingly, the sharp bound on the negative threshold

$$\mathcal{C} := \sum_{j=0}^{N-1} \frac{\rho^{M-j}}{c_j} \prod_{i=0, i \neq j}^{N-1} \frac{(M - n_i)^2}{(n_j - n_i)^2}.$$

is obtained on rank 1 matrices with positive entries.

Consequences

1. Quantitative version of Schoenberg's theorem in fixed dimension: polynomials that preserve positivity on $\mathbb{P}_N$, but not on $\mathbb{P}_{N+1}$.
2. When $M = N$, the theorem provides an exact characterization of polynomials of degree at most N that preserve positivity on $\mathbb{P}_N$.
3. The result holds verbatim for sums of *real powers*.
4. Surprisingly, the sharp bound on the negative threshold
$$\mathcal{C} := \sum_{j=0}^{N-1} \frac{\rho^{M-j}}{c_j} \prod_{i=0, i \neq j}^{N-1} \frac{(M - n_i)^2}{(n_j - n_i)^2}.$$
is obtained on rank 1 matrices with positive entries.
5. The proofs involve a deep result on *Schur positivity*.

Consequences

1. Quantitative version of Schoenberg's theorem in fixed dimension: polynomials that preserve positivity on $\mathbb{P}_N$, but not on $\mathbb{P}_{N+1}$.
2. When $M = N$, the theorem provides an exact characterization of polynomials of degree at most N that preserve positivity on $\mathbb{P}_N$.
3. The result holds verbatim for sums of *real powers*.
4. Surprisingly, the sharp bound on the negative threshold
$$\mathcal{C} := \sum_{j=0}^{N-1} \frac{\rho^{M-j}}{c_j} \prod_{i=0, i\neq j}^{N-1} \frac{(M-n_i)^2}{(n_j-n_i)^2}.$$
is obtained on rank 1 matrices with positive entries.
5. The proofs involve a deep result on *Schur positivity*.
6. **Further applications:** Schubert cell-type stratifications, connections to Rayleigh quotients, thresholds for analytic functions and Laplace transforms, additional novel symmetric function identities, ...

Schur polynomials

Key ingredient in proof – representation theory / symmetric functions:

Given a decreasing N-tuple $n_{N-1} > n_{N-2} > \cdots > n_0 \geqslant 0$, the corresponding **Schur polynomial** over a field $\mathbb{F}$ is the unique polynomial extension to $\mathbb{F}^N$ of

$$s_{(n_{N-1},\dots,n_0)}(x_1,\dots,x_N) := \frac{\det(x_i^{n_{j-1}})}{\det(x_i^{j-1})}$$

for pairwise distinct $x_i \in \mathbb{F}$.

Schur polynomials

Key ingredient in proof – representation theory / symmetric functions:

Given a decreasing N-tuple $n_{N-1} > n_{N-2} > \cdots > n_0 \geqslant 0$, the corresponding **Schur polynomial** over a field $\mathbb{F}$ is the unique polynomial extension to $\mathbb{F}^N$ of

$$s_{(n_{N-1},\dots,n_0)}(x_1,\dots,x_N) := \frac{\det(x_i^{n_{j-1}})}{\det(x_i^{j-1})}$$

for pairwise distinct $x_i \in \mathbb{F}$. Note that the denominator is precisely the Vandermonde determinant

$$V(\mathbf{x}) = V(x_1,\dots,x_N) := \det(x_i^{j-1}) = \prod_{1\leqslant i<j\leqslant N} (x_j - x_i).$$

Schur polynomials

Key ingredient in proof – representation theory / symmetric functions:

Given a decreasing N-tuple $n_{N-1} > n_{N-2} > \cdots > n_0 \geqslant 0$, the corresponding **Schur polynomial** over a field $\mathbb{F}$ is the unique polynomial extension to $\mathbb{F}^N$ of

$$s_{(n_{N-1},\dots,n_0)}(x_1,\dots,x_N) := \frac{\det(x_i^{n_{j-1}})}{\det(x_i^{j-1})}$$

for pairwise distinct $x_i \in \mathbb{F}$. Note that the denominator is precisely the Vandermonde determinant

$$V(\mathbf{x}) = V(x_1,\dots,x_N) := \det(x_i^{j-1}) = \prod_{1\leqslant i<j\leqslant N} (x_j - x_i).$$

Example: If $N = 2$ and $\mathbf{n} = (m < n)$, then

$$s_{\mathbf{n}}(x_1,x_2) = \frac{x_1^n x_2^m - x_1^m x_2^n}{x_1 - x_2} = (x_1x_2)^m(x_1^{n-m-1}+x_1^{n-m-2}x_2+\cdots+x_2^{n-m-1}).$$

Basis of homogeneous symmetric polynomials in $x_1,\dots,x_N$.

By-product: novel symmetric function identity

- Well-known identity of Cauchy: if $f_0(t) = 1/(1-t) = \sum_{k \geqslant 0} t^k$, then

$$\det f_0[\mathbf{u}\mathbf{v}^T] = V(\mathbf{u})V(\mathbf{v}) \sum_{\mathbf{n}} s_{\mathbf{n}}(\mathbf{u}) s_{\mathbf{n}}(\mathbf{v}),$$

where $\mathbf{n}$ runs over all decreasing integer tuples with at most N parts.

By-product: novel symmetric function identity

- Well-known identity of Cauchy: if $f_0(t) = 1/(1-t) = \sum_{k \geqslant 0} t^k$, then
$$\det f_0[\mathbf{u}\mathbf{v}^T] = V(\mathbf{u})V(\mathbf{v}) \sum_{\mathbf{n}} s_{\mathbf{n}}(\mathbf{u}) s_{\mathbf{n}}(\mathbf{v}),$$
where $\mathbf{n}$ runs over all decreasing integer tuples with at most N parts.
- Frobenius extended this to all $f_c(t) = (1-ct)/(1-t)$ for a scalar c.

By-product: novel symmetric function identity

- Well-known identity of Cauchy: if $f_0(t) = 1/(1-t) = \sum_{k \geqslant 0} t^k$, then

$$\det f_0[\mathbf{u}\mathbf{v}^T] = V(\mathbf{u})V(\mathbf{v}) \sum_{\mathbf{n}} s_{\mathbf{n}}(\mathbf{u}) s_{\mathbf{n}}(\mathbf{v}),$$

  where $\mathbf{n}$ runs over all decreasing integer tuples with at most N parts.

- Frobenius extended this to all $f_c(t) = (1-ct)/(1-t)$ for a scalar c.

- We show this for *every* power series
  $\rightsquigarrow$ obtained by generalizing a matrix positivity computation of Loewner:

Theorem (K., 2018)

Fix a commutative unital ring R and let t be an indeterminate. Let $f(t) := \sum_{M \geqslant 0} f_M t^M \in R[[t]]$ be an arbitrary formal power series. Given vectors $\mathbf{u}, \mathbf{v} \in R^N$ for some $N \geqslant 1$, we have:

$$\det f[t\mathbf{u}\mathbf{v}^T] = V(\mathbf{u})V(\mathbf{v}) \sum_{M \geqslant \binom{N}{2}} t^M \sum_{\mathbf{n}=(n_{N-1},\dots,n_0) \vdash M} s_{\mathbf{n}}(\mathbf{u}) s_{\mathbf{n}}(\mathbf{v}) \prod_{k=0}^{N-1} f_{n_k}.$$

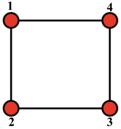
Positivity and Combinatorics

Matrices with zeros according to graphs

- In many applications, rare for all variables to depend strongly on each other – simplifies prediction.
- Many variables are (conditionally) independent – domain-specific knowledge in applications.

Matrices with zeros according to graphs

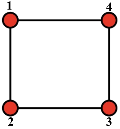
- In many applications, rare for all variables to depend strongly on each other – simplifies prediction.
- Many variables are (conditionally) independent – domain-specific knowledge in applications. Leverage the (conditional) independence structure to reduce dimension.



$$\begin{pmatrix} * & * & 0 & * \\ * & * & * & 0 \\ 0 & * & * & * \\ * & 0 & * & * \end{pmatrix}$$

Matrices with zeros according to graphs

- In many applications, rare for all variables to depend strongly on each other – simplifies prediction.
- Many variables are (conditionally) independent – domain-specific knowledge in applications. Leverage the (conditional) independence structure to reduce dimension.

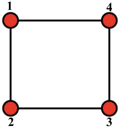


$$\begin{pmatrix} * & * & 0 & * \\ * & * & * & 0 \\ 0 & * & * & * \\ * & 0 & * & * \end{pmatrix}$$

- Natural to encode dependencies via a graph, where lack of an edge signifies conditional independence (given other variables).

Matrices with zeros according to graphs

- In many applications, rare for all variables to depend strongly on each other – simplifies prediction.
- Many variables are (conditionally) independent – domain-specific knowledge in applications. Leverage the (conditional) independence structure to reduce dimension.



$$\begin{pmatrix} * & * & 0 & * \\ * & * & * & 0 \\ 0 & * & * & * \\ * & 0 & * & * \end{pmatrix}$$

- Natural to encode dependencies via a graph, where lack of an edge signifies conditional independence (given other variables).

Study matrices with zeros according to graphs:

Given a graph $G = (V, E)$ on N vertices, and $I \subset \mathbb{R}$, define

$$\mathbb{P}_G(I) := \{A = (a_{ij}) \in \mathbb{P}_N(I) : a_{ij} = 0 \text{ if } i \neq j,\ (i,j) \notin E\}.$$

Note: a_{ij} can be zero if $(i, j) \in E$.

Preserving positivity with sparsity constraints

Given a subset $I \subset \mathbb{R}$ and a graph $G = (V, E)$, define for $A \in \mathbb{P}_G(I)$:

$$(f_G[A])_{ij} := \begin{cases} f(a_{ij}) & \text{if } i = j \text{ or } (i,j) \in E, \\ 0 & \text{otherwise.} \end{cases}$$

Preserving positivity with sparsity constraints

Given a subset $I \subset \mathbb{R}$ and a graph $G = (V, E)$, define for $A \in \mathbb{P}_G(I)$:

$$(f_G[A])_{ij} := \begin{cases} f(a_{ij}) & \text{if } i = j \text{ or } (i,j) \in E, \\ 0 & \text{otherwise.} \end{cases}$$

Can we characterize the functions f such that
$f_G[A] \in \mathbb{P}_G$ for every $A \in \mathbb{P}_G(I)$?

Preserving positivity with sparsity constraints

Given a subset $I \subset \mathbb{R}$ and a graph $G = (V, E)$, define for $A \in \mathbb{P}_G(I)$:

$$(f_G[A])_{ij} := \begin{cases} f(a_{ij}) & \text{if } i = j \text{ or } (i,j) \in E, \\ 0 & \text{otherwise.} \end{cases}$$

Can we characterize the functions f such that $f_G[A] \in \mathbb{P}_G$ for every $A \in \mathbb{P}_G(I)$?

- Previously known characterization for individual graphs: only for K_2, i.e., $\mathbb{P}_2$ – Vasudeva (1979).
- Only known characterization for sequence of graphs: $\{K_N : N \in \mathbb{N}\}$ [Schoenberg, Rudin]. Yields absolutely monotonic functions.

Preserving positivity with sparsity constraints

Given a subset $I \subset \mathbb{R}$ and a graph $G = (V, E)$, define for $A \in \mathbb{P}_G(I)$:

$$(f_G[A])_{ij} := \begin{cases} f(a_{ij}) & \text{if } i = j \text{ or } (i,j) \in E, \\ 0 & \text{otherwise.} \end{cases}$$

Can we characterize the functions f such that $f_G[A] \in \mathbb{P}_G$ for every $A \in \mathbb{P}_G(I)$?

- Previously known characterization for individual graphs: only for K_2, i.e., $\mathbb{P}_2$ – Vasudeva (1979).
- Only known characterization for sequence of graphs: $\{K_N : N \in \mathbb{N}\}$ [Schoenberg, Rudin]. Yields absolutely monotonic functions.
- (Guillot–K.–Rajaratnam, 2016:) Characterization for any collection of trees.

Preserving positivity with sparsity constraints

Given a subset $I \subset \mathbb{R}$ and a graph $G = (V, E)$, define for $A \in \mathbb{P}_G(I)$:

$$(f_G[A])_{ij} := \begin{cases} f(a_{ij}) & \text{if } i = j \text{ or } (i,j) \in E, \\ 0 & \text{otherwise.} \end{cases}$$

Can we characterize the functions f such that $f_G[A] \in \mathbb{P}_G$ for every $A \in \mathbb{P}_G(I)$?

- Previously known characterization for individual graphs: only for K_2, i.e., $\mathbb{P}_2$ – Vasudeva (1979).
- Only known characterization for sequence of graphs: $\{K_N : N \in \mathbb{N}\}$ [Schoenberg, Rudin]. Yields absolutely monotonic functions.
- (Guillot–K.–Rajaratnam, 2016:) Characterization for any collection of trees.

We now explain how *powers* preserving positivity $\rightsquigarrow$ a novel graph invariant.

Powers preserving positivity: Working example

Distinguished family of functions: the power maps $x^\alpha, \alpha \in \mathbb{R},\ x \geqslant 0$.
(Here, $0^\alpha := 0$.)

Powers preserving positivity: Working example

Distinguished family of functions: the power maps $x^\alpha, \alpha \in \mathbb{R},\ x \geqslant 0$.
(Here, $0^\alpha := 0$.)

Example: Suppose $T_5 = \begin{pmatrix} 1 & 0.6 & 0.5 & 0 & 0 \\ 0.6 & 1 & 0.6 & 0.5 & 0 \\ 0.5 & 0.6 & 1 & 0.6 & 0.5 \\ 0 & 0.5 & 0.6 & 1 & 0.6 \\ 0 & 0 & 0.5 & 0.6 & 1 \end{pmatrix}$.

Powers preserving positivity: Working example

Distinguished family of functions: the power maps $x^\alpha, \alpha \in \mathbb{R},\ x \geqslant 0$.
(Here, $0^\alpha := 0$.)

Example: Suppose $T_5 = \begin{pmatrix} 1 & 0.6 & 0.5 & 0 & 0 \\ 0.6 & 1 & 0.6 & 0.5 & 0 \\ 0.5 & 0.6 & 1 & 0.6 & 0.5 \\ 0 & 0.5 & 0.6 & 1 & 0.6 \\ 0 & 0 & 0.5 & 0.6 & 1 \end{pmatrix}$.

Raise each entry to the αth power for some $\alpha > 0$.
When is the resulting matrix positive semidefinite?

Powers preserving positivity: Working example

Distinguished family of functions: the power maps $x^\alpha, \alpha \in \mathbb{R},\ x \geqslant 0$.
(Here, $0^\alpha := 0$.)

Example: Suppose $T_5 = \begin{pmatrix} 1 & 0.6 & 0.5 & 0 & 0 \\ 0.6 & 1 & 0.6 & 0.5 & 0 \\ 0.5 & 0.6 & 1 & 0.6 & 0.5 \\ 0 & 0.5 & 0.6 & 1 & 0.6 \\ 0 & 0 & 0.5 & 0.6 & 1 \end{pmatrix}$.

Raise each entry to the αth power for some $\alpha > 0$.
When is the resulting matrix positive semidefinite?

Intriguing "phase transition" discovered by FitzGerald and Horn:

Powers preserving positivity: Working example

Distinguished family of functions: the power maps $x^\alpha, \alpha \in \mathbb{R},\ x \geqslant 0$.
(Here, $0^\alpha := 0$.)

Example: Suppose $T_5 = \begin{pmatrix} 1 & 0.6 & 0.5 & 0 & 0 \\ 0.6 & 1 & 0.6 & 0.5 & 0 \\ 0.5 & 0.6 & 1 & 0.6 & 0.5 \\ 0 & 0.5 & 0.6 & 1 & 0.6 \\ 0 & 0 & 0.5 & 0.6 & 1 \end{pmatrix}$.

Raise each entry to the αth power for some $\alpha > 0$.
When is the resulting matrix positive semidefinite?

Intriguing "phase transition" discovered by FitzGerald and Horn:

Theorem (FitzGerald–Horn, *J. Math. Anal. Appl.* 1977)

Let $N \geqslant 2$. Then $f(x) = x^\alpha$ preserves positivity on $\mathbb{P}_N([0,\infty))$ if and only if $\alpha \in \mathbb{N} \cup [N-2, \infty)$.

Powers preserving positivity: Working example

Distinguished family of functions: the power maps $x^\alpha, \alpha \in \mathbb{R},\ x \geqslant 0$.
(Here, $0^\alpha := 0$.)

Example: Suppose $T_5 = \begin{pmatrix} 1 & 0.6 & 0.5 & 0 & 0 \\ 0.6 & 1 & 0.6 & 0.5 & 0 \\ 0.5 & 0.6 & 1 & 0.6 & 0.5 \\ 0 & 0.5 & 0.6 & 1 & 0.6 \\ 0 & 0 & 0.5 & 0.6 & 1 \end{pmatrix}$.

Raise each entry to the αth power for some $\alpha > 0$.
When is the resulting matrix positive semidefinite?

Intriguing "phase transition" discovered by FitzGerald and Horn:

Theorem (FitzGerald–Horn, *J. Math. Anal. Appl.* 1977)

Let $N \geqslant 2$. Then $f(x) = x^\alpha$ preserves positivity on $\mathbb{P}_N([0,\infty))$ if and only if $\alpha \in \mathbb{N} \cup [N-2, \infty)$. The threshold $N-2$ is called the ***critical exponent****.*

Powers preserving positivity: Working example

Distinguished family of functions: the power maps $x^\alpha, \alpha \in \mathbb{R},\ x \geqslant 0$.
(Here, $0^\alpha := 0$.)

Example: Suppose $T_5 = \begin{pmatrix} 1 & 0.6 & 0.5 & 0 & 0 \\ 0.6 & 1 & 0.6 & 0.5 & 0 \\ 0.5 & 0.6 & 1 & 0.6 & 0.5 \\ 0 & 0.5 & 0.6 & 1 & 0.6 \\ 0 & 0 & 0.5 & 0.6 & 1 \end{pmatrix}$.

Raise each entry to the αth power for some $\alpha > 0$.
When is the resulting matrix positive semidefinite?

Intriguing "phase transition" discovered by FitzGerald and Horn:

Theorem (FitzGerald–Horn, *J. Math. Anal. Appl.* 1977)

*Let $N \geqslant 2$. Then $f(x) = x^\alpha$ preserves positivity on $\mathbb{P}_N([0,\infty))$ if and only if $\alpha \in \mathbb{N} \cup [N-2, \infty)$. The threshold $N-2$ is called the **critical exponent**.*

So for T_5 as above, all powers $\alpha \in \mathbb{N} \cup [3, \infty)$ work.

Powers preserving positivity: Working example

Distinguished family of functions: the power maps $x^\alpha, \alpha \in \mathbb{R},\ x \geqslant 0$.
(Here, $0^\alpha := 0$.)

Example: Suppose $T_5 = \begin{pmatrix} 1 & 0.6 & 0.5 & 0 & 0 \\ 0.6 & 1 & 0.6 & 0.5 & 0 \\ 0.5 & 0.6 & 1 & 0.6 & 0.5 \\ 0 & 0.5 & 0.6 & 1 & 0.6 \\ 0 & 0 & 0.5 & 0.6 & 1 \end{pmatrix}$.

Raise each entry to the αth power for some $\alpha > 0$.
When is the resulting matrix positive semidefinite?

Intriguing "phase transition" discovered by FitzGerald and Horn:

Theorem (FitzGerald–Horn, *J. Math. Anal. Appl.* 1977)

*Let $N \geqslant 2$. Then $f(x) = x^\alpha$ preserves positivity on $\mathbb{P}_N([0,\infty))$ if and only if $\alpha \in \mathbb{N} \cup [N-2, \infty)$. The threshold $N-2$ is called the **critical exponent**.*

So for T_5 as above, all powers $\alpha \in \mathbb{N} \cup [3, \infty)$ work. Can we do better?

Critical exponent of a graph

Exploit the sparsity structure of $\mathbb{P}_G$.

Problem: Compute the set of powers preserving positivity on $\mathbb{P}_G$:

$$\mathcal{H}_G := \{\alpha \geqslant 0 : A^{\circ\alpha} \in \mathbb{P}_G \text{ for all } A \in \mathbb{P}_G([0,\infty))\}$$

$$CE(G) := \text{ smallest } \alpha_0 \text{ s.t. } x^\alpha \text{ preserves positivity on } \mathbb{P}_G, \forall \alpha \geqslant \alpha_0.$$

Critical exponent of a graph

Exploit the sparsity structure of $\mathbb{P}_G$.

Problem: Compute the set of powers preserving positivity on $\mathbb{P}_G$:

$$\mathcal{H}_G := \{\alpha \geqslant 0 : A^{\circ\alpha} \in \mathbb{P}_G \text{ for all } A \in \mathbb{P}_G([0,\infty))\}$$

$$CE(G) := \text{ smallest } \alpha_0 \text{ s.t. } x^\alpha \text{ preserves positivity on } \mathbb{P}_G, \forall \alpha \geqslant \alpha_0.$$

- By FitzGerald–Horn, $CE(G)$ always exists and is $\leqslant |V(G)| - 2$.
  Call this the *critical exponent of the graph* G.

Critical exponent of a graph

Exploit the sparsity structure of $\mathbb{P}_G$.

Problem: Compute the set of powers preserving positivity on $\mathbb{P}_G$:

$$\mathcal{H}_G := \{\alpha \geqslant 0 : A^{\circ\alpha} \in \mathbb{P}_G \text{ for all } A \in \mathbb{P}_G([0,\infty))\}$$

$$CE(G) := \text{ smallest } \alpha_0 \text{ s.t. } x^\alpha \text{ preserves positivity on } \mathbb{P}_G, \forall \alpha \geqslant \alpha_0.$$

- By FitzGerald–Horn, $CE(G)$ always exists and is $\leqslant |V(G)| - 2$.
  Call this the *critical exponent of the graph* G.
- FitzGerald–Horn studied the case $G = K_r$: $\quad CE(K_r) = r - 2$.

Critical exponent of a graph

Exploit the sparsity structure of $\mathbb{P}_G$.

Problem: Compute the set of powers preserving positivity on $\mathbb{P}_G$:

$$\mathcal{H}_G := \{\alpha \geqslant 0 : A^{\circ\alpha} \in \mathbb{P}_G \text{ for all } A \in \mathbb{P}_G([0,\infty))\}$$

$$CE(G) := \text{ smallest } \alpha_0 \text{ s.t. } x^\alpha \text{ preserves positivity on } \mathbb{P}_G, \forall \alpha \geqslant \alpha_0.$$

- By FitzGerald–Horn, $CE(G)$ always exists and is $\leqslant |V(G)| - 2$. Call this the *critical exponent of the graph* G.
- FitzGerald–Horn studied the case $G = K_r$: $CE(K_r) = r - 2$.
- Guillot–K.–Rajaratnam [*Trans. AMS* 2016] studied trees: $CE(T) = 1$.

Critical exponent of a graph

Exploit the sparsity structure of $\mathbb{P}_G$.

Problem: Compute the set of powers preserving positivity on $\mathbb{P}_G$:

$$\mathcal{H}_G := \{\alpha \geqslant 0 : A^{\circ\alpha} \in \mathbb{P}_G \text{ for all } A \in \mathbb{P}_G([0,\infty))\}$$

$$CE(G) := \text{ smallest } \alpha_0 \text{ s.t. } x^\alpha \text{ preserves positivity on } \mathbb{P}_G, \forall \alpha \geqslant \alpha_0.$$

- By FitzGerald–Horn, $CE(G)$ always exists and is $\leqslant |V(G)| - 2$. Call this the *critical exponent of the graph* G.

- FitzGerald–Horn studied the case $G = K_r$: $\quad CE(K_r) = r - 2$.
- Guillot–K.–Rajaratnam [*Trans. AMS* 2016] studied trees: $\quad CE(T) = 1$.

- How do $CE(G)$ and $\mathcal{H}_G$ depend on the geometry of G? Compute $CE(G)$ for a family containing complete graphs and trees?

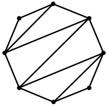
Chordal graphs – powers preserving positivity

Trees have no cycles of length $n \geqslant 3$.

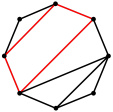
Chordal graphs – powers preserving positivity

Trees have no cycles of length $n \geqslant 3$.

Definition: G is *chordal* if it does not contain induced cycles of length $n \geqslant 4$.



Chordal

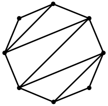


Not Chordal

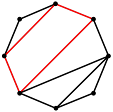
Chordal graphs – powers preserving positivity

Trees have no cycles of length $n \geqslant 3$.

Definition: G is *chordal* if it does not contain induced cycles of length $n \geqslant 4$.



Chordal



Not Chordal

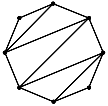
Theorem (Guillot–K.–Rajaratnam, *J. Combin. Theory Ser. A* 2016)

Let $K_r^{(1)}$ be the 'almost complete' graph on r nodes – missing one edge. Let $r = r(G)$ be the largest integer such that either K_r or $K_r^{(1)}$ is an induced subgraph of G.

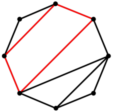
Chordal graphs – powers preserving positivity

Trees have no cycles of length $n \geqslant 3$.

Definition: G is *chordal* if it does not contain induced cycles of length $n \geqslant 4$.



Chordal



Not Chordal

Theorem (Guillot–K.–Rajaratnam, *J. Combin. Theory Ser. A* 2016)

Let $K_r^{(1)}$ be the 'almost complete' graph on r nodes – missing one edge. Let $r = r(G)$ be the largest integer such that either K_r or $K_r^{(1)}$ is an induced subgraph of G.

If G is chordal with $|V| \geqslant 2$, then $\mathcal{H}_G = \mathbb{N} \cup [r-2, \infty)$.

In particular, $CE(G) = r - 2$.

Unites complete graphs, trees, band graphs, split graphs. . .

Non-chordal graphs

Example: Band graphs with bandwidth d: $CE(G) = \min(d, n-2)$.

So for $T_5 = \begin{pmatrix} 1 & 0.6 & 0.5 & 0 & 0 \\ 0.6 & 1 & 0.6 & 0.5 & 0 \\ 0.5 & 0.6 & 1 & 0.6 & 0.5 \\ 0 & 0.5 & 0.6 & 1 & 0.6 \\ 0 & 0 & 0.5 & 0.6 & 1 \end{pmatrix}$ as above, all powers $\geqslant 2 = d$ work.

Non-chordal graphs

Example: Band graphs with bandwidth d: $CE(G) = \min(d, n-2)$.

So for $T_5 = \begin{pmatrix} 1 & 0.6 & 0.5 & 0 & 0 \\ 0.6 & 1 & 0.6 & 0.5 & 0 \\ 0.5 & 0.6 & 1 & 0.6 & 0.5 \\ 0 & 0.5 & 0.6 & 1 & 0.6 \\ 0 & 0 & 0.5 & 0.6 & 1 \end{pmatrix}$ as above, all powers $\geqslant 2 = d$ work.

Other graphs? (Talk by *Dominique Guillot* in MS18-iii.)

Non-chordal graphs

Example: Band graphs with bandwidth d: $CE(G) = \min(d, n-2)$.

So for $T_5 = \begin{pmatrix} 1 & 0.6 & 0.5 & 0 & 0 \\ 0.6 & 1 & 0.6 & 0.5 & 0 \\ 0.5 & 0.6 & 1 & 0.6 & 0.5 \\ 0 & 0.5 & 0.6 & 1 & 0.6 \\ 0 & 0 & 0.5 & 0.6 & 1 \end{pmatrix}$ as above, all powers $\geqslant 2 = d$ work.

Other graphs? (Talk by *Dominique Guillot* in MS18-iii.)
$CE(G)$ **in terms of other graph invariants?** Not clear.

American Institute of Mathematics
NSF
SERB INDIA
Taylor & Francis
DARPA
ICMS International Centre for Mathematical Sciences
Infosys Foundation
Indian Institute of Science भारतीय विज्ञान संस्थान
ILAS International Linear Algebra Society

Selected publications

D. Guillot, A. Khare, and B. Rajaratnam:

[1] *Preserving positivity for rank-constrained matrices*, Trans. AMS, 2017.

[2] *Preserving positivity for matrices with sparsity constraints*, Tr. AMS, 2016.

[3] *Critical exponents of graphs*, J. Combin. Theory Ser. A, 2016.

[4] *Complete characterization of Hadamard powers preserving Loewner positivity, monotonicity, and convexity*, J. Math. Anal. Appl., 2015.

A. Belton, D. Guillot, A. Khare, and M. Putinar:

[5] *Matrix positivity preservers in fixed dimension. I*, Advances in Math., 2016.

[6] *Moment-sequence transforms*, Preprint, 2016.

[7] *A panorama of positivity* (survey), Shimorin volume + Ransford-60 proc.

---

[8] *On the sign patterns of entrywise positivity preservers in fixed dimension*, (With T. Tao) Preprint, 2017.

[9] *Smooth entrywise positivity preservers, a Horn–Loewner master theorem, and Schur polynomials*, Preprint, 2018.