Schur polynomials and matrix positivity preservers

Alexander Belton (Lancaster University)

Dominique Guillot (University of Delaware)

Apoorva Khare (Stanford University)

Mihai Putinar (University of California at Santa Barbara and Newcastle University)

Abstract

A classical result by Schoenberg (1942) identifies all real-valued functions that preserve positive semidefiniteness (psd) when applied entrywise to matrices of arbitrary dimension. Schoenberg's work has continued to attract significant interest, including renewed recent attention due to applications in high-dimensional statistics. However, despite a great deal of effort in the area, an effective characterization of entrywise functions preserving positivity in a fixed dimension remains elusive to date. As a first step, we characterize new classes of polynomials preserving positivity in fixed dimension. The proof of our main result is representation theoretic, and employs Schur polynomials.

Entrywise functions preserving positivity

Definition. Let $\mathcal{P}_N(I) :=$ Hermitian $N \times N$ positive semidefinite matrices with entries in I.

Problem

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

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

Which functions $f[-] : \mathcal{P}_N(\mathbb{R}) \to \mathcal{P}_N(\mathbb{R})$ have this property? (Note: $f(x) = 1, x$ work.)

This problem has a long history, starting with Schur.

The *Schur product* (or Hadamard product) of two matrices is $A \circ B := (a_{jk} b_{jk})$.

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

If $A, B \in \mathcal{P}_N(\mathbb{C})$, then $A \circ B \in \mathcal{P}_N(\mathbb{C})$.

As a consequence of the Schur product theorem:

- $f(x) = x^2, x^3, \dots$ preserve positivity on $\mathcal{P}_N(\mathbb{C})$ for all N.
- If $c_k \geqslant 0 \; \forall k$ and $f(x) = \sum_{k=0}^{\infty} c_k x^k$ is convergent, then $f[-]$ preserves positivity.

Question (Pólya–Szegö, 1925): Are these functions all that work (in **every** dimension)?

Schoenberg's theorem

Pólya and Szegö's question was answered positively by Schoenberg.

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

Let $f : [-1,1] \to \mathbb{R}$. The following are equivalent:

1. The entrywise map preserves positivity, $f[-] : \mathcal{P}_N([-1,1]) \to \mathcal{P}_N(\mathbb{R})$ for all $N \geqslant 1$.
2. The function f is analytic on $[-1,1]$ and has non-negative Taylor coefficients.

Note: Schoenberg's result characterizes functions preserving positivity for matrices of **all** dimensions.

Challenging refinement: Classify entrywise maps preserving positivity in fixed dimension. (This is known for $N = 1, 2$, but is open to date for $N \geqslant 3$.)

Modern motivations: regularization of covariance matrices

Understanding relationships between variables is a fundamental problem in many fields (e.g. bioinformatics, climate science, finance, etc.).

Given a sample $x_1, \dots, x_n \in \mathbb{R}^p$, estimate the **covariance matrix** $\Sigma := (\mathrm{Cov}(X_j, X_k))_{j,k=1}^p$. This is very challenging in high dimensions.

- *Previous approaches:* Classical estimator $S =$ sample covariance matrix is a poor estimator of Σ in modern "large p, small n" problems.

  Compressed sensing methods (Daubechies, Tao, Candes, …) use convex optimization, and work well for dimensions up to $\sim 10,000$. However, they are not scalable to modern-day problems with $100,000+$ variables (disease detection, climate science, finance…).
- *Promising recent approach:* Small entries in S suggest independence of the corresponding variables. Thus it is natural to threshold these entries. E.g.,

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

  Can be significant if $p = 100,000$ and only $\sim 1\%$ of the entries of the true Σ are nonzero.
- More generally, apply a function $f : \mathbb{R} \to \mathbb{R}$ to elements of S [Bickel–Levina 2008; Hero–Rajaratnam 2011; Rothman–Levina–Zhu 2009].
- Highly scalable. Analysis on the cone – no optimization. Regularized matrix $f[S]$ further used in applications, where (estimates of) Σ are required in PCA, CCA, MANOVA, etc.

Question: When does this procedure $f[-]$ preserve positivity? (Critical for applications.)

Main result: polynomials preserving positivity

Finding a useful characterization of functions preserving positivity in fixed dimension remains a difficult open problem.

Horn's necessary condition

Theorem (Horn, *Trans. AMS* 1969; Guillot–Khare–Rajaratnam, *Trans. AMS* 2016)

Fix $N \geqslant 1$ and suppose that $f \in C^{N-1}(I)$, where $I := (0, \rho)$ and $0 < \rho \leqslant \infty$. Suppose $f[-]$ preserves positivity on $\mathcal{P}_N(I)$. Then $f^{(k)}(x) \geqslant 0$ for all $x \in I$ and $0 \leqslant k \leqslant N-1$.

Main result

Now suppose $c_0, \dots, c_{N-1}, c_N \neq 0$, and

$$f(z) = \sum_{j=0}^{N-1} c_j z^j + c_N z^N$$

preserves positivity on $\mathcal{P}_N((-\rho, \rho))$. Then by Horn's result, $c_0, \dots, c_{N-1} > 0$.

Can c_N be negative? What is a sharp bound?
(Open since Horn's 1969 paper.)

Theorem 1 ([1], *Adv. Math.* 2016)

Fix $\rho > 0$ and integers $N \geqslant 1$, $M \geqslant 0$. Let $f(z) := \sum_{j=0}^{N-1} c_j z^j + c' z^M$ be a polynomial with real coefficients. Let $\mathbf{c} := (c_0, \dots, c_{N-1})$ and define

$$\mathcal{C}(\mathbf{c}; z^M; N, \rho) := \sum_{j=0}^{N-1} \binom{M}{j}^2 \binom{M-j-1}{N-j-1}^2 \frac{\rho^{M-j}}{c_j}. \tag{1}$$

Then the following are equivalent.

1. $f[-]$ preserves positivity on $\mathcal{P}_N(\overline{D}(0, \rho))$.
2. Either $c_0, \dots, c_{N-1}, c' \geqslant 0$, or $c_0, \dots, c_{N-1} > 0$ and $c' \geqslant -\mathcal{C}(\mathbf{c}; z^M; N, \rho)^{-1}$.
3. $f[-]$ preserves positivity on the real rank-one matrices, with entries in $(0, \rho)$.

Some consequences:

- Characterization of polynomials of degree $\leqslant N$ that preserve positivity on $\mathcal{P}_N$.
- Provides the first construction of polynomials that preserve positivity on $\mathcal{P}_N$, but not on $\mathcal{P}_{N+1}$. Thus Horn's result is sharp.
- Surprisingly, preserving positivity on $\mathcal{P}_N(\overline{D}(0, \rho))$ is equivalent to preserving positivity on the much smaller set of real rank-one matrices.
- Also implies a sufficient condition for an arbitrary analytic function to preserve positivity.

A determinantal identity of Jacobi–Trudi type

A **Schur polynomial** is the unique extension to $\mathbb{F}^N$ of

$$s_{(n_N, \dots, n_1)}(x_1, \dots, x_N) := \frac{\det(x_i^{n_j + j - 1})}{\det(x_i^{j-1})} \qquad (n_N \geqslant \cdots \geqslant n_1)$$

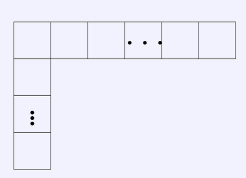
for pairwise distinct $x_i \in \mathbb{F}$. The denominator is precisely the Vandermonde determinant $\Delta_N(x_1, \dots, x_N) := \det(x_i^{j-1}) = \prod_{1 \leqslant i < j \leqslant N} (x_j - x_i)$.

Theorem 2 ([1], *Adv. Math.* 2016)

Let $M \geqslant N \geqslant 1$ be integers, and $c_0, \dots, c_{N-1} \in \mathbb{F}^\times$ be non-zero scalars in any field $\mathbb{F}$. Define the polynomial

$$p_t(z) := t(c_0 + \cdots + c_{N-1} z^{N-1}) - z^M,$$

and the hook partition $\mu(M, N, j) := (M - N + 1, 1, \dots, 1, 0, \dots, 0)$



$(N - j - 1$ ones, j zeros).

Then the following identity holds for $\mathbf{u}, \mathbf{v} \in \mathbb{F}^N$:

$$\det p_t[\mathbf{u}\mathbf{v}^T] = t^{N-1} \Delta_N(\mathbf{u}) \Delta_N(\mathbf{v}) \prod_{j=0}^{N-1} c_j \times \Big(t - \sum_{j=0}^{N-1} \frac{s_{\mu(M,N,j)}(\mathbf{u}) s_{\mu(M,N,j)}(\mathbf{v})}{c_j} \Big). \tag{2}$$

The proof uses the Cauchy–Binet formula to expand $\det \sum_{j=1}^m c_{n_j} A^{\circ n_j}$ for $A = \mathbf{u}\mathbf{v}^T$.

Proof of main result (sketch)

Clearly $(1) \implies (3)$. We now show $(3) \implies (2)$, assuming $c_0, \dots, c_{N-1} > 0 > c'$ and $M \geqslant N$.

Note that (1), (3) can be reformulated via *linear matrix inequalities*:

$$f[A] \in \mathcal{P}_N \qquad \Longleftrightarrow \qquad A^{\circ M} \leqslant t \cdot \sum_{j=0}^{N-1} c_j A^{\circ j}.$$

Question: *How small can $t = |c'|^{-1}$ be?* Note by Equation (2):

$$0 \leqslant \det p_t[\mathbf{u}\mathbf{u}^T] = t^{N-1} \Delta_N(\mathbf{u})^2 c_0 \cdots c_{N-1} \Big(t - \sum_{j=0}^{N-1} \frac{s_{\mu(M,N,j)}(\mathbf{u})^2}{c_j} \Big). \tag{3}$$

Set $u_k := \sqrt{\rho}(1 - t'\epsilon_k)$, with pairwise distinct $\epsilon_k \in (0,1)$, and $t' \in (0,1)$. Thus, $\Delta_N(\mathbf{u}) \neq 0$. Taking the limit as $t' \to 0^+$, since the final term in (3) must be non-negative, it follows that

$$t \geqslant \sum_{j=0}^{N-1} \frac{s_{\mu(M,N,j)}(\sqrt{\rho}, \dots, \sqrt{\rho})^2}{c_j} = \sum_{j=0}^{N-1} s_{\mu(M,N,j)}(1, \dots, 1)^2 \cdot \frac{\rho^{M-j}}{c_j} = \mathcal{C}(\mathbf{c}; z^M; N, \rho).$$

That $s_{\mu(M,N,j)}(1, \dots, 1) = \binom{M}{j}\binom{M-j-1}{N-j-1}$ follows using the Weyl Dimension Formula in type A; or the dual Jacobi–Trudi (Von Nägelsbach–Kostka) identity. □

A refined analysis of the proof shows that in fact, $f[A] \in \mathcal{P}_N$ is generically positive definite.

Theorem 3 ([1], *FPSAC* 2016)

Suppose $N > 1$, and $A \in \mathcal{P}_N(\overline{D}(0, \rho))$ has a row with pairwise distinct entries. Define $f(z) := c_0 + \cdots + c_{N-1} z^{N-1} - \mathcal{C}(\mathbf{c}; z^M; N, \rho)^{-1} z^M$. Then $f[A]$ is positive definite.

This uses the connection to Schur polynomials and semi-standard Young tableaux (see [1]).

Alternative variational approach: Rayleigh quotients via Schur polynomials

Given $\mathbf{c} = (c_0, \dots, c_{N-1})$, define $h_{\mathbf{c}}(z) := \sum_{j=0}^{N-1} c_j z^j$. Want the smallest constant $t > 0$ with

$$A^{\circ M} \leqslant t \cdot h_{\mathbf{c}}[A],$$

for all $A \in \mathcal{P}_N(\overline{D}(0, \rho))$, or for all rank-one psd matrices A.

Strategy:

1. First produce optimal constant $\Psi_{\mathbf{c},M}(A)$ for a single matrix: $A^{\circ M} \leqslant \Psi_{\mathbf{c},M}(A) \cdot h_{\mathbf{c}}[A]$.
2. Maximize $\Psi_{\mathbf{c},M}(A)$ over $A \in \mathcal{P}_N(\overline{D}(0, \rho))$.

The first step is addressed by the following result.

Theorem 4 ([1], *Adv. Math.* 2016)

With notation as above, the optimal constant for A equals a Rayleigh quotient:

$$\Psi_{\mathbf{c},M}(A) = \sup_{v \in S^{2N-1} \cap (\ker h_{\mathbf{c}}[A])^\perp} \frac{v^* A^{\circ M} v}{v^* h_{\mathbf{c}}[A] v} = \varrho(h_{\mathbf{c}}[A]^{\dagger/2} A^{\circ M} h_{\mathbf{c}}[A]^{\dagger/2}), \tag{4}$$

where $\varrho(C), C^\dagger$ denote the spectral radius and Moore-Penrose inverse of C, respectively.

The proof uses the theory of Kronecker normal forms.

Novel connections: Rayleigh quotients to Schur polynomials

If $A = \mathbf{u}\mathbf{u}^*$, the Rayleigh quotient $\Psi_{\mathbf{c},M}(A)$ can be written using Schur polynomials!

Theorem 5 ([1], *FPSAC* 2016)

If $\mathbf{u} \in \mathbb{C}^N$ has distinct coordinates and $A = \mathbf{u}\mathbf{u}^*$, then $h_{\mathbf{c}}[A]$ is invertible, and

$$\Psi_{\mathbf{c},M}(\mathbf{u}\mathbf{u}^*) = (\mathbf{u}^{\circ M})^* h_{\mathbf{c}}[\mathbf{u}\mathbf{u}^*]^{-1} \mathbf{u}^{\circ M} = \sum_{j=0}^{N-1} \frac{|s_{\mu(M,N,j)}(\mathbf{u})|^2}{c_j}. \tag{5}$$

Notice, this result immediately implies Main Theorem $(3) \implies (2)$.

The proof of $(2) \implies (1)$ is more involved (see [1] for details).

Reference

[1] A. Belton, D. Guillot, A. Khare, and M. Putinar. Matrix positivity preservers in fixed dimension. I. *Adv. Math.*, 298:325–368, 2016. (Extended abstract published in *DMTCS Proceedings* **BC**, pp. 155–166.)