Matrix Analysis, Pólya frequency functions, and Preservers of (Total) Positivity

Apoorva Khare
Indian Institute of Science

1. Introduction

This text arose out of the course notes for *Math 341: Matrix Analysis and Positivity*, a one-semester course offered in Spring 2018 and Fall 2019 at the Indian Institute of Science (IISc). Owing to the subsequent inclusion of additional topics, the text has now grown to cover roughly a two-semester course in analysis and matrix positivity preservers – or, more broadly, composition operators preserving various kinds of positive kernels. Thus in this text, we briefly describe some notions of positivity in matrix theory, followed by our main focus: a detailed study of the operations that preserve these notions (and, in the process, an understanding of some aspects of real functions). Several different notions of positivity in analysis, studied for classical and modern reasons, are touched upon in the text:

- Positive semidefinite and positive definite matrices.
- Entrywise positive matrices.
- A common strengthening of the first two notions, which involves totally positive (TP) and totally non-negative (TN) matrices.
- Settings somewhat outside matrix theory. For instance, consider discrete data associated with positive measures on locally compact abelian groups G. E.g., for $G = \mathbb{R}$, one obtains moment sequences, which are intimately related to positive semidefinite Hankel matrices. For $G = S^1$, the circle group, one obtains Fourier–Stieltjes sequences, which are connected to positive semidefinite Toeplitz matrices. (See works of Carathéodory, Hamburger, Hausdorff, Herglotz, and Stieltjes, among others.)
- More classically, functions and kernels with positivity structures have long been studied in analysis, including on locally compact groups and metric spaces (see Bochner, Schoenberg, von Neumann, Pólya). Distinguished examples include positive definite functions and Pólya frequency functions and sequences.

The text begins by discussing the above notions, focussing on their properties and some results in matrix theory. The next two parts then study, in detail, the preservers of several of these notions of positivity. Among other things, this journey involves going through many beautiful classical results by leading experts in analysis during the first half of the twentieth century. Apart from also covering several different tools required in proving these results, an interesting outcome also is that several classes of "positive" matrices repeatedly get highlighted by way of studying positivity preservers – these include generalized Vandermonde matrices, Hankel moment matrices and kernels, and Toeplitz kernels on the line or the integers (aka Pólya frequency functions and sequences).

In this text, we will study the post-composition transforms that preserve (total) positivity on various classes of kernels. When the kernel has finite domain – i.e., is a matrix – then this amounts to studying *entrywise* preservers of various notions of positivity. The question of why entrywise calculus was studied – as compared to the usual holomorphic functional calculus – has a rich and classical history in the analysis literature, beginning with the work of Schoenberg, Rudin, Loewner, and Horn (these results are proved in Part 3 of the text), but also drawing upon earlier works of Menger, Schur, Bochner, and others. (In fact, the entrywise calculus was introduced, and the first such result proved, by Schur in 1911.) Interestingly, this entrywise calculus also arises in modern-day applications from high-dimensional covariance estimation; we elaborate on this in Section 13.1, and briefly also in Section 14. Furthermore, this evergreen area of mathematics continues to be studied in the literature, drawing techniques from – and also contributing to – symmetric function theory, statistics and graphical models, combinatorics, and linear algebra (in addition to analysis).

As a historical curiosity, the course and this text arose in a sense out of research carried out in significant measure by mathematicians at Stanford University (including their students) over the years. This includes Loewner, Karlin, and their students: FitzGerald, Horn, Micchelli, and Pinkus. Less directly, there was also Katznelson, who had previously worked with Helson, Kahane, and Rudin, leading to Rudin's strengthening of Schoenberg's theorem. (Coincidentally, Pólya and Szegő, who made the original observation on entrywise preservers of positivity using the Schur product theorem, were again colleagues at Stanford.) On a personal note, the author's contributions to this area also have their origins in his time spent at Stanford University, collaborating with Alexander Belton, Dominique Guillot, Mihai Putinar, Bala Rajaratnam, and Terence Tao (though the collaboration with the last-named colleague was carried out almost entirely at IISc).

We now discuss the course, the notes that led to this text, and their mathematical contents. The notes were scribed by the students taking the course in Spring 2018 at IISc, followed by extensive "homogenization" by the author – and, in several sections, addition of material. Each section was originally intended to cover the notes of roughly one 90-minute lecture, or occasionally two; that said, some material has subsequently been moved around for logical, mathematical, and expositional reasons. The notes, and the course itself, require an understanding of basic linear algebra and analysis, with a bit of measure theory as well. Beyond these basic topics, we have tried to keep these notes as self-contained as possible, with full proofs. To that end, we have included proofs of "preliminary" results, including:

(i) results of Schoenberg, Menger, von Neumann, Fréchet, and others connecting metric geometry and positive definite functions to matrix positivity;
(ii) results in Euclidean geometry, including on triangulation, Heron's formula for the area of a triangle, and connecting Cayley–Menger matrices to simplicial volumes;
(iii) Boas–Widder and Bernstein's theorems on functions with positive forward differences;
(iv) Sierpińsky's result: mid-convexity and measurability imply continuity;
(v) an extension to normed linear spaces, of (a special case of) a classical result of Ostrowski on mid-convexity and local boundedness implying continuity;
(vi) Whitney's density of totally positive matrices inside totally non-negative matrices;
(vii) Descartes' rule of signs – several variants;
(viii) a follow-up to Descartes, by Laguerre, on variation diminution in power series, and its follow-up by Fekete involving Pólya frequency sequences;
(ix) Fekete's result on totally positive matrices via positive contiguous minors;
(x) (on a related note:) results on real and complex polynomials, their "compositions", and zeros of these: by Gauss–Lucas, Hermite–Kakeya–Obrechkoff, Hermite–Biehler, Routh–Hurwitz, Hermite–Poulain, Laguerre, Maló, Jensen, Schur, Weisner, de Bruijn, and Pólya;
(xi) a detailed sketch of Pólya and Schur's characterizations of multiplier sequences;
(xii) Mercer's lemma, identifying positive semidefinite kernels with kernels of positive type;
(xiii) Perron's theorem for matrices with positive entries (the precursor to Perron–Frobenius);
(xiv) compound matrices and Kronecker's theorem on their spectra;
(xv) Sylvester's criterion and the Schur product theorem on positive (semi)definiteness (also, the Jacobi formula);
(xvi) the Jacobi complementary minor formula;
(xvii) the Rayleigh–Ritz theorem;
(xviii) a special case of Weyl's inequality on eigenvalues;
(xix) matrix identities by Andréief and Cauchy–Binet, and a continuous generalization;

(xx) the discreteness of zeros of real analytic functions (and a sketch of the continuity of roots of complex polynomials); and

(xxi) the equivalence of Cauchy's and Littlewood's definitions of Schur polynomials (and of the Jacobi–Trudi and von Nägelsbach–Kostka identities) via Lindstr̊om–Gessel–Viennot bijections.

Owing to considerations of time, we had to leave out some proofs. These include proofs of theorems by Hamburger/Hausdorff/Stieltjes, Fubini, Tonelli, Cauchy, Montel, Morera, and Hurwitz; a Schur positivity phenomenon for ratios of Schur polynomials; Lebesgue's dominated convergence theorem; as well as the closure of real analytic functions under composition. Most of these can be found in standard textbooks in mathematics. We also omit the proofs of several classical results on Laplace transforms and Pólya frequency functions, found in textbooks, in papers by Schoenberg and his co-authors, and in Karlin's comprehensive monograph on total positivity. Nevertheless, as the previous and current paragraphs indicate, these notes cover many classical results by past experts and acquaint the reader with a variety of tools in analysis (especially the study of real functions) and in matrix theory – many of these tools are not found in more "traditional" courses on these subjects.

This text is broadly divided into six parts, with detailed bibliographic notes following each part. In **Part 1**, the key objects of interest – namely, positive semidefinite / totally positive / totally non-negative matrices – are introduced, together with some basic results as well as some important classes of examples. (The analogous kernels are also studied.) In **Part 2**, we begin the study of functions acting entrywise on such matrices, and preserving the relevant notion of positivity. Here, we will mostly restrict ourselves to studying power functions that act on various sets of matrices of a *fixed* size. This is a long-studied question, including by Bhatia, Elsner, Fallat, FitzGerald, Hiai, Horn, Jain, Johnson, and Sokal; as well as by the author in collaboration with Guillot and Rajaratnam. In particular, an interesting highlight is the construction by Jain of individual (pairs of) matrices, which turn out to encode the entire set of entrywise powers preserving Loewner positivity, monotonicity, and convexity. We also obtain certain necessary conditions on general entrywise functions that preserve positivity, including multiplicative mid-convexity and continuity, as well as a classification of all functions preserving total non-negativity or total positivity in each fixed dimension. We explain some of the modern motivations, and we end with some unsolved problems.

Part 3 deals with some of the foundational results on matrix positivity preservers. After mentioning some of the early history – including work by Menger, Fréchet, Bochner, and Schoenberg – we classify the entrywise functions that preserve positive semidefiniteness (= positivity) in all dimensions, or total non-negativity on Hankel matrices of *all* sizes. This is a celebrated result of Schoenberg – later strengthened by Rudin – which is a converse to the Schur product theorem, and we prove a stronger version by using a rank-constrained test set. The proof given in these notes is different from the previous approaches of Schoenberg and Rudin, is essentially self-contained, and uses relatively less sophisticated machinery compared to the works of Schoenberg and Rudin. Moreover, it first proves a variant by Vasudeva for matrices with only positive entries, and it lends itself to a multivariate generalization (which will not be covered here). The starting point of these proofs is a necessary condition for entrywise preservers in a fixed dimension, proved by Loewner (and Horn) in the late 1960s. To this day, this result remains essentially the only known condition in a fixed dimension $n \geq 3$, and a proof of a (rank-constrained, as above) stronger version is also provided in these notes. In addition to techniques and ingredients introduced by the above authors, the text also borrows from the author's joint work with Belton, Guillot, and Putinar. This part ends with several appendices:

(1) The first appendix covers a result by Boas and Widder, which shows a converse "mean value theorem" for divided differences.
(2) We next cover recent work by Vishwakarma on an "off-diagonal" variant of the positivity preserver problem.
(3) The third and fourth appendices classify preservers of Loewner positivity, monotonicity, and convexity on "dimension-free matrices", or on kernels over infinite domains.
(4) The fifth appendix explores the theme of Euclidean distance geometry, with a focus on some classical results by Menger.

In **Parts 4 and 5**, we formulate the preserver problem in analysis terms, using composition operators on kernels. This allows one to consider such questions not only for matrices of a fixed or arbitrary size, but also over more general, infinite domains. It also makes available for use, the powerful analysis machinery developed by Bernstein, Pólya, Schoenberg, Widder, and others. Thus in Part 5, we provide characterizations of such composition operators preserving total positivity or non-negativity on structured kernels – specifically Toeplitz kernels on various sub-domains of $\mathbb{R}$. Two distinguished classes of such kernels are Pólya frequency functions and Pólya frequency sequences, i.e., Toeplitz kernels on $\mathbb{R}\times\mathbb{R}$ and $\mathbb{Z}\times\mathbb{Z}$, respectively.

Before solving the preserver problem in this paradigm, we begin by developing some of the results by Schoenberg and others on Pólya frequency functions and sequences. In fact we begin even more classically: with a host of root-location results for zeros of real and complex polynomials, which motivated the study of the Laguerre–Pólya class of entire functions and the Pólya–Schur classification of multiplier sequences. After presenting these, we briefly discuss modern offshoots of the Laguerre–Pólya–Schur program, followed by Schoenberg's results. (Similarly, in the next part we also briefly discuss the Wallach set in representation theory and probability.)

In Part 5, we return to the question studied in Part 2 above, of classifying the preservers of *all* TN or TP kernels, on $X \times Y$ for arbitrary totally ordered sets X, Y. Here we provide a complete resolution of this question. In a sense, the total positivity preserver problem is the culmination of all that has come before in this text; it uses many of the tools and techniques from the previous parts. These include (i) Vandermonde kernels, TP kernel completions of 2×2 matrices, and Fekete's result, from Part 1; (ii) entrywise powers preserving positivity (via a trick of Jain to use Toeplitz cosine matrices), and the classification of total-positivity preservers in finite dimension, from Part 2; (iii) the stronger Vasudeva theorem classifying entrywise positivity preservers on low-rank Hankel matrices, from Part 3; and (iv) preservers of Toeplitz kernels, including of Pólya frequency functions and sequences, from Part 4. We also develop as needed, set-theoretic tools and Whitney-type density results – as well as understanding the structure and preservers of continuous Hankel kernels defined on an interval. This latter requires results of Mercer, Bernstein, Hamburger, and Widder.

We remark that Part 4 contains – in addition to theorems found e.g. in Karlin's book – very recent results from 2020+ on Pólya frequency functions (and not merely their preservers), which in particular are not found in previous treatments of the subject. To name a few:

- a characterization of Pólya frequency functions of order p, for any $p \geq 3$;
- strengthenings of multiple results of Karlin (1955) and of Schoenberg (1964);
- a converse to the same result of Karlin (1955);
- a critical exponent phenomenon in total positivity;

- a closer look at a multiparameter family of density functions introduced by Hirschman and Widder (1949); and
- a connection between Pólya frequency functions and the Riemann hypothesis.

(Similarly, Part 2 of this text contains recent results on totally non-negative/positive matrices, which are not found in earlier books on total positivity.) Thus, in a sense Parts 4 and 5 can be viewed as "one possible sequel" to Karlin's book and to the body of work by Schoenberg on Pólya frequency functions, since they present novel material on these classes of functions, then use the "structural" results by the aforementioned authors about various families of Toeplitz kernels to classify the preservers of various sub-families of these objects. In addition to results of Schoenberg and his coauthors, as well as an authoritative survey of these results, compiled in Karlin's majestic monograph, both parts borrow from the author's 2020 works (one with Belton, Guillot, and Putinar).

In the final **Part 6**, we return to the study of entrywise functions preserving positivity in a fixed dimension. This is a challenging problem – it is still open in general, even for 3×3 matrices – and we restrict ourselves in this part to studying polynomial preservers. According to the Schur product theorem (1911), if the polynomial has all non-negative coefficients, then it is easily seen to be a preserver; but, interestingly, until 2016 not a single example was known of any other entrywise polynomial preserver of positivity in a fixed dimension $n \geq 3$. Very recently, this question has been answered to some degree of satisfaction by the author, in collaboration first with Belton, Guillot, and Putinar, and subsequently with Tao. The text ends by covering some of this recent progress, and it comes back full circle to Schur through symmetric function theory.

A quick note on the **logical structure**: Parts 1–5 are best read sequentially; note that some sections do not get used later in the text, e.g. Sections 8 and 31, and the Appendices. That said, Part 4, which involves non-preserver results on Pólya frequency functions and sequences, can be read from scratch, requiring only Sections 6 and 12.1 and Lemma 26.3 as pre-requisites. The final Part 6 can be read following Section 9 (also see the Schoenberg–Rudin theorem 16.2 and the Horn–Loewner theorem 17.1). We also point out the occurrence of **Historical notes** and **Further questions**, which serve to acquaint the reader with past work(er)s as well as related areas; and possible avenues for future work – and which can be accessed from the Index at the end. (See also the Bibliographic notes at the end of each part of the text.)

To conclude, thanks are owed to the scribes (listed below), as well as to Sayan Adhikari, Alexander Belton, Shabarish Chenakkod, Projesh Nath Choudhury, Julian R. D'Costa, Sudip Dolai, Dominique Guillot, Prakhar Gupta, Roger A. Horn, Sarvesh Ravichandran Iyer, Poornendu Kumar, Gadadhar Misra, Frank Oertel, Aaradhya Pandey, Vamsi Pritham Pingali, Paramita Pramanick, Mihai Putinar, Shubham Rastogi, Ritul, Aditya Guha Roy, Siddhartha Sahi, Kartik Singh, Naren Sundaravaradan, G.V. Krishna Teja, Akaki Tikaradze, Raghavendra Tripathi, Prateek Kumar Vishwakarma, Pranjal Warade, and Upamanyu Yaddanapudi for helpful suggestions that improved the text. I am, of course, deeply indebted to my collaborators for their support and all of their research efforts in positivity – but also for many stimulating discussions, which helped shape my thinking about the field as a whole and the structure of this text in particular. Finally, I am grateful to the University Grants Commission (UGC, Government of India), the Science and Engineering Research Board (SERB) and the Department of Science and Technology (DST) of the Government of India, the Council for Scientific and Industrial Research (CSIR, Government of India), the Infosys Foundation, and the Tata Trusts – for their support through a CAS-II grant, through a MATRICS grant and the Ramanujan and Swarnajayanti Fellowships and a DST-FIST grant, through a Shanti Swarup Bhatnagar Prize, through a Young Investigator Award, and through their Travel Grants, respectively.

Department of Mathematics, Indian Institute of Science
& Analysis and Probability Research Group
Bangalore, India

List of scribes: Each section contains material originally covered in a 90-minute lecture, and subsequently augmented by the author. These lectures (in Spring 2018) did not cover Sections 4, 8 in Part 1, Section 15 in Part 2, the entirety of Parts 4 and 5, and the Appendices. The text also includes two sets of "general remarks" by the author; see Sections 13.1 and 14.

Part 1: Preliminaries

(1) January 03: (Much of the above text.)
(2) January 08: Sarvesh Iyer
(3) January 10: Prateek Kumar Vishwakarma
(4) Apoorva Khare
(5) January 17: Prakhar Gupta
(6) January 19: Pranjal Warade
(7) January 24: Swarnalipa Datta, and Apoorva Khare
(8) Apoorva Khare

Part 2: Entrywise powers preserving (total) positivity in a fixed dimension

(9) January 31: Raghavendra Tripathi
(10) February 02: Pabitra Barman
(11) February 07: Kartick Ghosh
(12) February 09: K. Philip Thomas (and Feb 14: Pranab Sarkar), and Apoorva Khare
(13) February 28: Ratul Biswas (with an introduction by Apoorva Khare)
(14) March 02: Prateek Kumar Vishwakarma (with remarks by Apoorva Khare)
(15) Apoorva Khare

Part 3: Entrywise functions preserving positivity in all dimensions

(16) February 14, 16: Pranab Sarkar, Pritam Ganguly, and Apoorva Khare
(17) March 07: Shubham Rastogi
(18) March 09: Poornendu Kumar, Sarvesh Iyer, and Apoorva Khare
(19) March 14: Poornendu Kumar and Shubham Rastogi
(20) March 16: Lakshmi Kanta Mahata and Kartick Ghosh
(21) March 19, 21: Kartick Ghosh, Swarnalipa Datta, and Apoorva Khare
(22) March 23: Sarvesh Iyer and Raghavendra Tripathi
(23) Appendix A: Poornendu Kumar, Shubham Rastogi, and Apoorva Khare
(24) Appendix B: Prateek Kumar Vishwakarma and Apoorva Khare
(25–27) Appendices C–E: Apoorva Khare

Part 4: Pólya frequency functions and sequences

(28–35) Apoorva Khare

Part 5: Composition operators preserving totally positive kernels

(36–40) Apoorva Khare

Part 6: Entrywise polynomials preserving positivity in a fixed dimension

(41) March 26: Prakhar Gupta, Pranjal Warade, and Apoorva Khare
(42) March 28: Ratul Biswas and K. Philip Thomas
(43) April 02: Pritam Ganguly and Pranab Sarkar
(44) April 11: Pabitra Barman, Lakshmi Kanta Mahata, and Apoorva Khare
(45) Appendix F: Apoorva Khare

Contents

Index 327

Part 1:
Preliminaries

Part 1: Preliminaries

2. The cone of positive semidefinite matrices. Examples. Schur complements.

A *kernel* is a function $K : X \times Y \to \mathbb{R}$. Broadly speaking, the goal of this text is to understand: ***Which functions $F : \mathbb{R} \to \mathbb{R}$, when applied to kernels with some notion of positivity, preserve that notion?*** To do so, we first study the test sets of such kernels K themselves, and then the post-composition operators F that preserve these test sets. We begin by understanding such kernels when the domains X, Y are finite, i.e., matrices.

In this text, we will assume familiarity with linear algebra and a first course in calculus/analysis. To set notation: an uppercase letter with a two-integer subscript (such as $A_{m\times n}$) represents a matrix with m rows and n columns. If m, n are clear from context or unimportant, then they will be omitted. Three examples of real matrices are $0_{m\times n}, \mathbf{1}_{m\times n}, \mathrm{Id}_{n\times n}$, which are the (rectangular) matrix consisting of all zeros, all ones, and the identity matrix, respectively. The entries of a matrix A will be denoted a_{ij}, a_{jk}, etc. Vectors are denoted by lowercase letters (occasionally in bold), and are columnar in nature. All matrices, unless specified otherwise, are real; and similarly, all functions, unless specified otherwise, are defined on – and take values in – $\mathbb{R}^m$ for some $m \geq 1$. As is standard, we let $\mathbb{C}, \mathbb{R}, \mathbb{Q}, \mathbb{Z}, \mathbb{N}$ denote the complex numbers, reals, rationals, integers, and positive integers respectively. Given $S \subset \mathbb{R}$, let $S^{\geq 0} := S \cap [0, \infty)$.

2.1. **Preliminaries.** We begin with several basic definitions.

Definition 2.1. A matrix $A_{n\times n}$ is said to be *symmetric* if $a_{jk} = a_{kj}$ for all $1 \leq j, k \leq n$. A real symmetric matrix $A_{n\times n}$ is said to be *positive semidefinite)* if the real number $x^T A x$ is non-negative for all $x \in \mathbb{R}^n$ – in other words, the quadratic form given by A is positive semidefinite. If, furthermore, $x^T A x > 0$ for all $x \neq 0$ then A is said to be *positive definite*. Denote the set of (real symmetric) positive semidefinite matrices by $\mathbb{P}_n$.

We state the spectral theorem for symmetric (i.e., self-adjoint) operators without proof.

Theorem 2.2 (Spectral theorem for symmetric matrices)**.** *For $A_{n\times n}$ a real symmetric matrix, $A = U^T D U$ for some orthogonal matrix U (i.e., $U^T U = \mathrm{Id}$) and real diagonal matrix D. D contains all the eigenvalues of A (counting multiplicities) along its diagonal.*

As a consequence, $A = \sum_{j=1}^n \lambda_j v_j v_j^T$, where each v_j is an eigenvector for A with *real* eigenvalue λ_j, and the v_j (which are the columns of U^T) form an orthonormal basis of $\mathbb{R}^n$.

We also have the following related results, stated here without proof: the spectral theorem for two commuting matrices, and the singular value decomposition.

Theorem 2.3 (Spectral theorem for commuting symmetric matrices)**.** *Let $A_{n\times n}$ and $B_{n\times n}$ be two commuting real symmetric matrices. Then A and B are simultaneously diagonalizable, i.e., for some common orthogonal matrix U, $A = U^T D_1 U$ and $B = U^T D_2 U$ for D_1 and D_2 diagonal matrices (whose diagonal entries comprise the eigenvalues of A, B respectively).*

Theorem 2.4 (Singular value decomposition)**.** *Every real matrix $A_{m\times n} \neq 0$ decomposes as $A = P_{m\times m} \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix}_{m\times n} Q_{n\times n}$, where P, Q are orthogonal and Σ_r is a diagonal matrix with positive eigenvalues. The entries of Σ_r are called the* singular values *of A, and are the square roots of the non-zero eigenvalues of AA^T (or $A^T A$).*

2.2. **Criteria for positive (semi)definiteness.** We write down several equivalent criteria for positive (semi)definiteness. There are three initial criteria which are easy to prove, and a final criterion which requires separate treatment.

Theorem 2.5 (Criteria for positive (semi)definiteness)**.** *Given $A_{n\times n}$ a real symmetric matrix of rank $0 \leq r \leq n$, the following are equivalent:*

(1) A is positive semidefinite (respectively, positive definite).
(2) All eigenvalues of A are non-negative (respectively, positive).
(3) There exists a matrix $B \in \mathbb{R}^{r\times n}$ of rank r, such that $B^T B = A$. (In particular, if A is positive definite then B is square and non-singular.)

Proof. We prove only the positive semidefinite statements; minor changes show the corresponding positive definite variants. If (1) holds and λ is an eigenvalue – for an eigenvector x – then $x^T Ax = \lambda\|x\|^2 \geq 0$. Hence, $\lambda \geq 0$, proving (2). Conversely, if (2) holds then by the spectral theorem, $A = \sum_j \lambda_j v_j v_j^T$ with all $\lambda_j \geq 0$, so A is positive semidefinite:

$$x^T Ax = \sum_j \lambda_j x^T v_j v_j^T x = \sum_j \lambda_j (x^T v_j)^2 \geq 0, \qquad \forall x \in \mathbb{R}^n.$$

Next if (1) holds then write $A = U^T DU$ by the spectral theorem; note that $D = UAU^T$ has the same rank as A. Since D has non-negative diagonal entries d_{jj}, it has a square root $\sqrt{D}$, which is a diagonal matrix with diagonal entries $\sqrt{d_{jj}}$. Write $D = \begin{pmatrix} D'_{r\times r} & 0 \\ 0 & 0_{(n-r)\times(n-r)} \end{pmatrix}$, where D' is a diagonal matrix with positive diagonal entries. Correspondingly, write $U = \begin{pmatrix} P_{r\times r} & Q \\ R & S_{(n-r)\times(n-r)} \end{pmatrix}$. If we set $B := (\sqrt{D'}P \mid \sqrt{D'}Q)_{r\times n}$, then it is easily verified that

$$B^T B = \begin{pmatrix} P^T D'P & P^T D'Q \\ Q^T D'P & Q^T D'Q \end{pmatrix} = U^T DU = A.$$

Hence, (1) $\implies$ (3). Conversely, if (3) holds then $x^T Ax = \|Bx\|^2 \geq 0$ for all $x \in \mathbb{R}^n$. Hence, A is positive semidefinite. Moreover, we claim that B and $B^T B$ have the same null space and hence the same rank. Indeed, if $Bx = 0$ then $B^T Bx = 0$, while

$$B^T Bx = 0 \quad \implies \quad x^T B^T Bx = 0 \quad \implies \quad \|Bx\|^2 = 0 \quad \implies Bx = 0. \qquad \square$$

Corollary 2.6. *For any real symmetric matrix $A_{n\times n}$, the matrix $A - \lambda_{\min} \operatorname{Id}_{n\times n}$ is positive semidefinite, where $\lambda_{\min}$ denotes the smallest eigenvalue of A.*

We now state Sylvester's criterion for positive (semi)definiteness. (Incidentally, Sylvester is believed to have first introduced the use of "matrix" in mathematics, in the nineteenth century.) This requires some additional notation.

Definition 2.7. Given an integer $n \geq 1$, define $[n] := \{1, \dots, n\}$. Now given a matrix $A_{m\times n}$ and subsets $J \subset [m], K \subset [n]$, define $A_{J\times K}$ to be the submatrix of A with entries a_{jk} for $j \in J, k \in K$ (always considered to be arranged in increasing order in this text). If J, K have the same size then $\det A_{J\times K}$ is called a *minor* of A. If A is square and $J = K$ then $A_{J\times K}$ is called a *principal submatrix* of A, and $\det A_{J\times K}$ is a *principal minor*. The principal submatrix (and principal minor) are *leading* if $J = K = \{1, \dots, m\}$ for some $1 \leq m \leq n$.

Theorem 2.8 (Sylvester's criterion)**.** *A symmetric matrix is positive semidefinite (definite) if and only if all its principal minors are non-negative (positive).*

We will show Theorem 2.8 with the help of a few preliminary results.

Lemma 2.9. *If $A_{n\times n}$ is a positive semidefinite (respectively, positive definite) matrix, then so are all principal submatrices of A.*

Proof. Fix a subset $J \subset [n] = \{1, \dots, n\}$ (so $B := A_{J\times J}$ is the corresponding principal submatrix of A), and let $x \in \mathbb{R}^J$. Define $x' \in \mathbb{R}^n$ to be the vector, such that $x'_j = x_j$ for all $j \in J$ and 0 otherwise. It is easy to see that $x^T Bx = (x')^T Ax'$. Hence, B is positive (semi)definite if A is. □

As a corollary, all the principal minors of a positive semidefinite (positive definite) matrix are non-negative (positive) since the corresponding principal submatrices have non-negative (positive) eigenvalues and hence non-negative (positive) determinants. So one direction of Sylvester's criterion holds trivially.

Lemma 2.10. *Sylvester's criterion is true for positive definite matrices.*

Proof. We induct on the dimension of the matrix A. Suppose $n = 1$. Then A is just an ordinary real number, so its only principal minor is A itself, and so the result is trivial.

Now, suppose the result is true for matrices of dimension $\leq n - 1$. We claim that A has at least $n - 1$ positive eigenvalues. To see this, let $\lambda_1, \lambda_2 \leq 0$ be eigenvalues of A. Let W be the $n - 1$ dimensional subspace of $\mathbb{R}^n$ with last entry 0. If v_j are orthogonal eigenvectors for λ_j, $j = 1, 2$, then the span of the v_j must intersect W nontrivially, since the sum of dimensions of these two subspaces of $\mathbb{R}^n$ exceeds n. Define $u := c_1v_1 + c_2v_2 \in W$; then $u^T Au > 0$ by Lemma 2.9. However,

$$u^T Au = (c_1v_1^T + c_2v_2^T)A(c_1v_1 + c_2v_2) = c_1^2\lambda_1||v_1||^2 + c_2^2\lambda_2||v_2||^2 \leq 0$$

thereby giving a contradiction and proving the claim.

Now since the determinant of A is positive (it is the minor corresponding to A itself), it follows that all eigenvalues are positive, completing the proof. □

We will now prove the Jacobi formula, an important result in its own right. A corollary of this result will be used, along with the previous result and the idea that positive semidefinite matrices can be expressed as entrywise limits of positive definite matrices, to prove Sylvester's criterion for all positive semidefinite matrices.

Theorem 2.11 (Jacobi formula)**.** *Let $A_t : \mathbb{R} \to \mathbb{R}^{n\times n}$ be a matrix-valued differentiable function. Letting $\operatorname{adj}(A_t)$ denote the adjugate matrix of A_t, we have:*

$$\frac{d}{dt}(\det A_t) = \operatorname{tr}\left(\operatorname{adj}(A_t)\frac{dA_t}{dt}\right). \tag{2.12}$$

Proof. The first step is to compute the differential of the determinant. We claim that

$$d(\det)(A)(B) = \operatorname{tr}(\operatorname{adj}(A)B), \qquad \forall A, B \in \mathbb{R}^{n\times n}.$$

As a special case, at $A = \mathrm{Id}_{n\times n}$, the differential of the determinant is precisely the trace.

To show the claim, we need to compute the directional derivative

$$\lim_{\epsilon\to 0} \frac{\det(A + \epsilon B) - \det A}{\epsilon}.$$

The fraction is a polynomial in ϵ with vanishing constant term (e.g., set $\epsilon = 0$ to see this); and we need to compute the coefficient of the linear term. Expand $\det(A + \epsilon B)$ using the Laplace expansion as a sum over permutations $\sigma \in S_n$; now each individual summand $(-1)^\sigma \prod_{k=1}^n (a_{k\sigma(k)} + \epsilon b_{k\sigma(k)})$ splits as a sum of 2^n terms. (It may be illustrative to try and work out the $n = 3$ case by hand.) From these $2^n \cdot n!$ terms, choose the ones that are linear in ϵ. For each $1 \leq i, j \leq n$, there are precisely $(n-1)!$ terms corresponding to ϵb_{ij}; and added together, they equal the (i,j)th cofactor C_{ij} of A – which equals $\operatorname{adj}(A)_{ji}$. Thus, the coefficient of ϵ is

$$d(\det)(A)(B) = \sum_{i,j=1}^{n} C_{ij} b_{ij},$$

and this is precisely $\operatorname{tr}(\operatorname{adj}(A)B)$, as claimed.

More generally, the above argument shows that if $B(\epsilon)$ is any family of matrices, with limit $B(0)$ as $\epsilon \to 0$, then

$$\lim_{\epsilon \to 0} \frac{\det(A + \epsilon B(\epsilon)) - \det A}{\epsilon} = \operatorname{tr}(\operatorname{adj}(A)B(0)). \tag{2.13}$$

Returning to the proof of the theorem, for $\epsilon \in \mathbb{R}$ small and $t \in \mathbb{R}$ we write

$$A_{t+\epsilon} = A_t + \epsilon B(\epsilon)$$

where $B(\epsilon) \to B(0) := \frac{dA_t}{dt}$ as $\epsilon \to 0$, by definition. Now compute using (2.13):

$$\frac{d}{dt}(\det A_t) = \lim_{\epsilon \to 0} \frac{\det(A_t + \epsilon B(\epsilon)) - \det A_t}{\epsilon} = \operatorname{tr}(\operatorname{adj}(A_t)\frac{dA_t}{dt}). \qquad \square$$

With these results in hand, we can finish the proof of Sylvester's criterion for positive semidefinite matrices.

Proof of Theorem 2.8. For positive definite matrices, the result was proved in Lemma 2.10. Now suppose $A_{n\times n}$ is positive semidefinite. One direction follows by the remarks preceding Lemma 2.10. We show the converse by induction on n, with an easy argument for $n = 1$ similar to the positive definite case.

Now suppose the result holds for matrices of dimension $\leq n-1$ and let $A_{n\times n}$ have all principal minors non-negative. Let B be any principal submatrix of A, and define $f(t) := \det(B + t\,\mathrm{Id})$. Note that $f'(t) = \operatorname{tr}(\operatorname{adj}(B + t\,\mathrm{Id}))$ by the Jacobi formula (2.12).

We claim that $f'(t) > 0\ \forall t > 0$. Indeed, each diagonal entry of $\operatorname{adj}(B + t\,\mathrm{Id})$ is the determinant of a proper principal submatrix of $A + t\,\mathrm{Id}_{n\times n}$, say $A' + t\,\mathrm{Id}_{k\times k}$. This submatrix is positive definite since A' is positive semidefinite by the induction hypothesis, and so $x^T(A' + t\,\mathrm{Id}_{k\times k})x = x^T A' x + t\|x\|^2 > 0$ for $x \in \mathbb{R}^k$ and $t > 0$. Hence its determinant is positive, and so $f'(t) > 0$.

The claim implies: $f(t) > f(0) = \det B \geq 0\ \forall t > 0$. Thus all principal minors of $A + t\,\mathrm{Id}$ are positive, and by Sylvester's criterion for positive definite matrices, $A + t\,\mathrm{Id}$ is positive definite for all $t > 0$. Now note that $x^T A x = \lim_{t\to 0^+} x^T(A + t\,\mathrm{Id}_{n\times n})x$; therefore the non-negativity of the right-hand side implies that of the left-hand side for all $x \in \mathbb{R}^n$, completing the proof. $\square$

2.3. **Examples of positive semidefinite matrices.** We next discuss several examples.

2.3.1. *Gram matrices.*

Definition 2.14. For any finite set of vectors $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^m$, their *Gram matrix* is given by $\operatorname{Gram}((\mathbf{x}_j)_j) := (\langle \mathbf{x}_j, \mathbf{x}_k \rangle)_{1\leq j,k\leq n}$.

A *correlation matrix* is a positive semidefinite matrix with ones on the diagonal.

In fact, we need not use $\mathbb{R}^m$ here; any inner product space/Hilbert space is sufficient.

Proposition 2.15. *Given a real symmetric matrix $A_{n\times n}$, it is positive semidefinite if and only if there exist an integer $m > 0$ and vectors $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^m$, such that $A = \operatorname{Gram}((\mathbf{x}_j)_j)$.*

As a special case, correlation matrices precisely correspond to those Gram matrices for which the $\mathbf{x}_j$ are unit vectors. We also remark that a "continuous" version of this result is given by a well-known result of Mercer [261]. See Theorem 39.9.

Proof. If A is positive semidefinite, then by Theorem 2.5 we can write $A = B^T B$ for some matrix $B_{m\times n}$. It is now easy to check that A is the Gram matrix of the columns of B.

Conversely, if $A = \mathrm{Gram}(\mathbf{x}_1, \dots, \mathbf{x}_n)$ with all $\mathbf{x}_j \in \mathbb{R}^m$, then to show that A is positive semidefinite, we compute for any $\mathbf{u} = (u_1, \dots, u_n)^T \in \mathbb{R}^n$:

$$\mathbf{u}^T A \mathbf{u} = \sum_{j,k=1}^n u_j u_k \langle \mathbf{x}_j, \mathbf{x}_k \rangle = \left\| \sum_{j=1}^n u_j \mathbf{x}_j \right\|^2 \geq 0. \qquad \square$$

2.3.2. *(Toeplitz) Cosine matrices.*

Definition 2.16. A matrix $A = (a_{jk})$ is *Toeplitz* if a_{jk} depends only on $j - k$.

Lemma 2.17. *Let $\theta_1, \dots, \theta_n \in [0, 2\pi]$. Then the matrix $C := (\cos(\theta_j - \theta_k))_{j,k=1}^n$ is positive semidefinite, with rank at most 2. In particular, $\alpha \mathbf{1}_{n\times n} + \beta C$ has rank at most 3 (for scalars α, β), and it is positive semidefinite if $\alpha, \beta \geq 0$.*

Proof. Define the vectors $u, v \in \mathbb{R}^n$ via $u^T = (\cos\theta_1, \dots, \cos\theta_n)$, $v^T = (\sin\theta_1, \dots, \sin\theta_n)$. Then $C = uu^T + vv^T$ via the identity $\cos(a-b) = \cos a \cos b + \sin a \sin b$, and clearly the rank of C is at most 2. (For instance, it can have rank 1 if the θ_j are equal.) As a consequence,

$$\alpha \mathbf{1}_{n\times n} + \beta C = \alpha \mathbf{1}_n \mathbf{1}_n^T + \beta u u^T + \beta v v^T$$

has rank at most 3; the final assertion is straightforward. $\square$

As a special case, if $\theta_1, \dots, \theta_n$ are in arithmetic progression, i.e., $\theta_{j+1} - \theta_j = \theta \ \forall j$ for some θ, then we obtain a positive semidefinite *Toeplitz* matrix

$$C = \begin{pmatrix} 1 & \cos\theta & \cos 2\theta & \dots & & & \\ \cos\theta & 1 & \cos\theta & \cos 2\theta & \dots & & \\ \cos 2\theta & \cos\theta & 1 & \cos\theta & \cos 2\theta & \dots & \\ & \cos 2\theta & \cos\theta & 1 & \cos\theta & \cos 2\theta & \dots \\ \vdots & & \ddots & & \ddots & & \ddots \end{pmatrix}.$$

This family of Toeplitz matrices was used by Rudin in a 1959 paper [306] on entrywise positivity preservers; see Theorem 16.3 for his result.

2.3.3. *Hankel matrices.*

Definition 2.18. A matrix $A = (a_{jk})$ is *Hankel* if a_{jk} depends only on $j + k$.

Example 2.19. $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ is Hankel but not positive semidefinite.

Example 2.20. For each $x \geq 0$, the matrix $\begin{pmatrix} 1 & x & x^2 \\ x & x^2 & x^3 \\ x^2 & x^3 & x^4 \end{pmatrix} = \begin{pmatrix} 1 \\ x \\ x^2 \end{pmatrix} \begin{pmatrix} 1 & x & x^2 \end{pmatrix}$ is Hankel and positive semidefinite of rank 1.

A more general perspective is as follows. Define $H_x := \begin{pmatrix} 1 & x & x^2 & \cdots \\ x & x^2 & x^3 & \cdots \\ x^2 & x^3 & x^4 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$, and let δ_x be the Dirac measure at $x \in \mathbb{R}$. The moments of this measure are given by $s_k(\delta_x) = \int_{\mathbb{R}} y^k \, d\delta_x(y) = x^k$, $k \geq 0$. Thus, H_x is the "moment matrix" of δ_x. More generally, given any non-negative measure μ supported on $\mathbb{R}$, with all moments finite, the corresponding *Hankel moment matrix* is the bi-infinite "matrix" given by

$$H_\mu := \begin{pmatrix} s_0 & s_1 & s_2 & \cdots \\ s_1 & s_2 & s_3 & \cdots \\ s_2 & s_3 & s_4 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \qquad \text{where } s_k = s_k(\mu) := \int_{\mathbb{R}} y^k \, d\mu(y). \tag{2.21}$$

Lemma 2.22. *The matrix H_μ is positive semidefinite. In other words, every finite principal submatrix is positive semidefinite.*

Note, this is equivalent to every leading principal submatrix being positive semidefinite.

Proof. Fix $n \geq 1$ and consider the finite principal (Hankel) submatrix H'_μ with the first n rows and columns. Let H'_{δ_x} be the Hankel matrix defined in a similar manner for the measure $\delta_x, x \in \mathbb{R}$. Now to show that H'_μ is positive semidefinite, we compute for any vector $\mathbf{u} \in \mathbb{R}^n$:

$$\mathbf{u}^T H'_\mu \mathbf{u} = \int_{\mathbb{R}} \mathbf{u}^T H'_{\delta_x} \mathbf{u} \, d\mu(x) = \int_{\mathbb{R}} ((1, x, \dots, x^{n-1})\mathbf{u})^2 \, d\mu(x) \geq 0,$$

where the final equality holds because H'_{δ_x} has rank 1 and factorizes as in Example 2.20. (Note that the first equality holds because we are taking finite linear combinations of the integrals in the entries of H'_μ.) $\square$

Remark 2.23. Lemma 2.22 is (the easier) half of a famous classical result by Hamburger. The harder converse result says that if a semi-infinite Hankel matrix is positive semidefinite, with (j,k)-entry s_{j+k} for $j, k \geq 0$, then there exists a non-negative Borel measure on the real line whose kth moment is s_k for all $k \geq 0$. This theorem will be useful later; it was shown by Hamburger in 1920–21, when he extended the Stieltjes moment problem to the entire real line in the series of papers [162]. These works established the moment problem in its own right, as opposed to being a tool used to determine the convergence or divergence of continued fractions (as previously developed by Stieltjes – see Remark 4.4).

There is also a multivariate version of Lemma 2.22 which is no harder than the lemma, modulo notation:

Lemma 2.24. *Given a measure μ on $\mathbb{R}^d$ for some integer $d \geq 1$, we define its moments for tuples of non-negative integers $\mathbf{n} = (n_1, \dots, n_d)$ via*

$$s_{\mathbf{n}}(\mu) := \int_{\mathbb{R}^d} \mathbf{x}^{\mathbf{n}} \, d\mu(\mathbf{x}) = \int_{\mathbb{R}^d} \prod_{j=1}^d x_j^{n_j} \, d\mu,$$

if these integrals converge. (Here, $\mathbf{x}^{\mathbf{n}} := \prod_j x_j^{n_j}$.) Now suppose $\mu \geq 0$ on $\mathbb{R}^d$ and let $\Psi : (\mathbb{Z}^{\geq 0})^d \to \mathbb{Z}^{\geq 0}$ be any bijection, such that $\Psi(\mathbf{0}) = 0$ (although this restriction is not really required). Define the semi-infinite matrix $H_\mu := (a_{jk})_{j,k=0}^\infty$ via $a_{jk} := s_{\Psi^{-1}(j)+\Psi^{-1}(k)}$, where we assume that all moments of μ exist. Then H_μ is positive semidefinite.

Proof. Given a real vector $\mathbf{u} = (u_0, u_1, \dots)^T$ with finitely many non-zero coordinates, we have

$$\mathbf{u}^T H_\mu \mathbf{u} = \sum_{j,k \geq 0} \int_{\mathbb{R}^d} u_j u_k \mathbf{x}^{\Psi^{-1}(j)+\Psi^{-1}(k)} \, d\mu(\mathbf{x}) = \int_{\mathbb{R}^d} ((1, \mathbf{x}^{\Psi^{-1}(1)}, \mathbf{x}^{\Psi^{-1}(2)}, \dots)\mathbf{u})^2 \, d\mu(\mathbf{x}) \geq 0.$$

□

2.3.4. *Matrices with sparsity.* Another family of positive semidefinite matrices involves matrices with a given zero pattern, i.e., structure of (non)zero entries. Such families are important in applications, as well as in combinatorial linear algebra, spectral graph theory, and graphical models/Markov random fields.

Definition 2.25. A graph $G = (V, E)$ is *simple* if the sets of vertices/nodes V and edges E are finite, and E contains no self-loops (v, v) or multi-edges. In this text, all graphs will be finite and simple. Given such a graph $G = (V, E)$, with node set $V = [n] = \{1, \dots, n\}$, define

$$\mathbb{P}_G := \{A \in \mathbb{P}_n \ : \ a_{jk} = 0 \text{ if } j \neq k \text{ and } (j,k) \notin E\}, \tag{2.26}$$

where $\mathbb{P}_n$ comprises the (real symmetric) positive semidefinite matrices of dimension n.

Also, a subset $C \subset X$ of a real vector space X is *convex* if $\lambda v + (1-\lambda) w \in C$ for all $v, w \in C$ and $\lambda \in [0,1]$. If instead $\alpha C \subset C$ for all $\alpha \in (0, \infty)$, then we say C is a *cone*.

Remark 2.27. The set $\mathbb{P}_G$ is a natural mathematical generalization of the cone $\mathbb{P}_n$ (and shares several of its properties). In fact, two "extreme" special cases are: (i) G is the complete graph, in which case $\mathbb{P}_G$ is the full cone $\mathbb{P}_n$ for $n = |V|$; and (ii) G is the empty graph, in which case $\mathbb{P}_G$ is the cone of $|V| \times |V|$ diagonal matrices with non-negative entries.

Akin to both of these cases, for all graphs G, the set $\mathbb{P}_G$ is in fact a closed convex cone.

Example 2.28. Let $G = \{\{v_1, v_2, v_3\}, \{(v_1, v_3), (v_2, v_3)\}\}$. The adjacency matrix is given by $A_G = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$. This is not in $\mathbb{P}_G$ (but $A_G - \lambda_{\min}(A_G) \operatorname{Id}_{3\times 3} \in \mathbb{P}_G$, see Corollary 2.6).

Example 2.29. For any graph G with node set $[n]$, let D_G be the diagonal matrix with (j,j) entry the degree of node j, i.e., the number of edges adjacent to j. Then the *graph Laplacian*, defined to be $L_G := D_G - A_G$ (where A_G is the adjacency matrix), is in $\mathbb{P}_G$.

Example 2.30. An important class of examples of positive semidefinite matrices arises from the Hessian matrix of (suitably differentiable) functions. In particular, if the Hessian is positive definite at a point, then this is an isolated local minimum.

2.4. **Schur complements.** We mention some more preliminary results here; these may be skipped for now but will get used in Lemma 9.5 below.

Definition 2.31. Given a matrix $M = \begin{pmatrix} P & Q \\ R & S \end{pmatrix}$, where P and S are square and S is non-singular, the *Schur complement* of M with respect to S is given by $M/S := P - QS^{-1}R$.

Schur complements arise naturally in theory and applications. As an important example, suppose $X_1, \dots, X_n$ and $Y_1, \dots, Y_m$ are random variables with covariance matrix $\Sigma = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix}$, with C non-singular. Then the conditional covariance matrix of X given Y is $\operatorname{Cov}(X|Y) := A - BC^{-1}B^T = \Sigma/C$. That such a matrix is also positive semidefinite is implied by the following folklore result by Albert in *SIAM J. Appl. Math* in 1969.

Theorem 2.32. *Given a symmetric matrix* $\Sigma = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix}$, *with* C *positive definite, the matrix* Σ *is positive (semi)definite if and only if the Schur complement* Σ/C *is thus.*

Proof. We first write down a more general matrix identity: for a non-singular matrix C and a square matrix A, one uses a factorization shown by Schur in 1917 in *J. reine angew. Math.*:

$$\begin{pmatrix} A & B \\ B' & C \end{pmatrix} = \begin{pmatrix} \mathrm{Id} & BC^{-1} \\ 0 & \mathrm{Id} \end{pmatrix} \begin{pmatrix} A - BC^{-1}B' & 0 \\ 0 & C \end{pmatrix} \begin{pmatrix} \mathrm{Id} & 0 \\ C^{-1}B' & \mathrm{Id} \end{pmatrix}. \tag{2.33}$$

(Note, the identity matrices on the right have different sizes.) Now set $B' = B^T$; then $\Sigma = X^T Y X$, where $X = \begin{pmatrix} \mathrm{Id} & 0 \\ C^{-1}B^T & \mathrm{Id} \end{pmatrix}$ is non-singular, and $Y = \begin{pmatrix} A - BC^{-1}B^T & 0 \\ 0 & C \end{pmatrix}$ is block diagonal (and real symmetric). The result is not hard to show from here. ☐

Akin to Sylvester's criterion, the above characterization has a variant for when C is positive semidefinite; however, this is not as easy to prove, and requires a more flexible "inverse":

Definition 2.34 (Moore–Penrose inverse)**.** Given any real $m \times n$ matrix A, the *pseudo-inverse* or *Moore–Penrose inverse* of A is an $n \times m$ matrix $A^\dagger$ satisfying: $AA^\dagger A = A$, $A^\dagger A A^\dagger = A^\dagger$, and $(AA^\dagger)_{m\times m}$, $(A^\dagger A)_{n \times n}$ are symmetric.

Lemma 2.35. *For every* $A_{m\times n}$, *the matrix* $A^\dagger$ *exists and is unique.*

Proof. By Theorem 2.4, write $A = P \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix} Q$, with $P_{m\times m}, Q_{n \times n}$ orthogonal, and Σ_r containing the non-zero singular values of A. It is easily verified that $Q^T \begin{pmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{pmatrix}^T P^T$ works as a choice of $A^\dagger$. To show the uniqueness: if $A_1^\dagger, A_2^\dagger$ are both choices of Moore–Penrose inverse for a matrix $A_{m\times n}$, then first compute using the defining properties:

$$AA_1^\dagger = (AA_2^\dagger A)A_1^\dagger = (AA_2^\dagger)^T (AA_1^\dagger)^T = (A_2^\dagger)^T (A^T (A_1^\dagger)^T A^T) = (A_2^\dagger)^T A^T = (AA_2^\dagger)^T = AA_2^\dagger.$$

Similarly, $A_1^\dagger A = A_2^\dagger A$, so $A_1^\dagger = A_1^\dagger (AA_1^\dagger) = A_1^\dagger (AA_2^\dagger) = (A_1^\dagger A) A_2^\dagger = (A_2^\dagger A) A_2^\dagger = A_2^\dagger$. ☐

Example 2.36. Here are some examples of the Moore–Penrose inverse of square matrices.

(1) If $D = \mathrm{diag}(\lambda_1, \dots, \lambda_r, 0, \dots, 0)$, with all $\lambda_j \neq 0$, then $D^\dagger = \mathrm{diag}(\frac{1}{\lambda_1}, \dots, \frac{1}{\lambda_r}, 0, \dots, 0)$.
(2) If A is positive semidefinite, then $A = U^T D U$ where D is a diagonal matrix. It is easy to verify that $A^\dagger = U^T D^\dagger U$.
(3) If A is non-singular then $A^\dagger = A^{-1}$.

We now mention the connection between the positivity of a matrix and its Schur complement with respect to a singular submatrix. First, note that the Schur complement is now defined in the expected way (here S, M are square), i.e., as follows:

$$M = \begin{pmatrix} P & Q \\ R & S \end{pmatrix} \qquad \Longrightarrow \qquad M/S := P - QS^\dagger R, \tag{2.37}$$

Now the proof of the following result can be found in standard textbooks on matrix analysis.

Theorem 2.38. *Given a symmetric matrix* $\Sigma = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix}$, *with* C *not necessarily invertible, the matrix* Σ *is positive semidefinite if and only if the following conditions hold:*

(1) C *is positive semidefinite.*

(2) The Schur complement Σ/C *is positive semidefinite.*
(3) $(\mathrm{Id} - CC^\dagger)B^T = 0$.

3. Schur product theorem. Totally positive/non-negative matrices. Examples.

3.1. **The Schur product.** We now make some straightforward observations about the set $\mathbb{P}_n$. The first is that $\mathbb{P}_n$ is topologically closed, convex, and closed under scaling by positive multiples (a "cone").

Lemma 3.1. *$\mathbb{P}_n$ is a closed, convex cone in $\mathbb{R}^{n\times n}$.*

Proof. All properties are easily verified using the definition of positive semidefiniteness. □

If A and B are positive semidefinite matrices, then we expect the product AB to also be positive semidefinite. This is true if AB is symmetric.

Lemma 3.2. *For $A, B \in \mathbb{P}_n$, if AB is symmetric then $AB \in \mathbb{P}_n$.*

Proof. In fact, $AB = (AB)^T = B^T A^T = BA$, so A and B commute. Writing $A = U^T D_1 U$ and $B = U^T D_2 U$ as per the Spectral theorem 2.3 for commuting matrices, we have

$$x^T(AB)x = x^T(U^T D_1 U \cdot U^T D_2 U)x = x^T U^T (D_1 D_2) U x = \|\sqrt{D_1 D_2} U x\|^2 \geq 0.$$

Hence, $AB \in \mathbb{P}_n$. □

Note, however, that AB need not be symmetric even if A and B are symmetric. In this case, the matrix AB certainly cannot be positive semidefinite; however, it still satisfies one of the equivalent conditions for positive semidefiniteness (shown above for symmetric matrices), namely, having a non-negative spectrum. We prove this with the help of another result, which shows the "tracial" property of the spectrum.

Lemma 3.3. *Given $A_{n\times m}, B_{m\times n}$, the non-zero eigenvalues of AB and BA (and their multiplicities) agree.*

(Here, "tracial" suggests that the expression for AB equals that for BA, as does the trace.)

Proof. Assume without loss of generality that $1 \leq m \leq n$. The result will follow if we can show that $\det(\lambda \operatorname{Id}_{n\times n} - AB) = \lambda^{n-m} \det(\lambda \operatorname{Id}_{m\times m} - BA)$ for all λ. In turn, this follows from the equivalence of characteristic polynomials of AB and BA up to a power of λ, which is why we must take the union of both spectra with zero. (In particular, the sought-for equivalence would also imply that the non-zero eigenvalues of AB and BA are equal up to multiplicity).

The proof finishes by considering the two following block matrix identities

$$\begin{pmatrix} \operatorname{Id}_{n\times n} & -A \\ 0 & \lambda \operatorname{Id}_{m\times m} \end{pmatrix} \begin{pmatrix} \lambda \operatorname{Id}_{n\times n} & A \\ B & \operatorname{Id}_{m\times m} \end{pmatrix} = \begin{pmatrix} \lambda \operatorname{Id}_{n\times n} - AB & 0 \\ \lambda B & \lambda \operatorname{Id}_{m\times m} \end{pmatrix},$$
$$\begin{pmatrix} \operatorname{Id}_{n\times n} & 0 \\ -B & \lambda \operatorname{Id}_{m\times m} \end{pmatrix} \begin{pmatrix} \lambda \operatorname{Id}_{n\times n} & A \\ B & \operatorname{Id}_{m\times m} \end{pmatrix} = \begin{pmatrix} \lambda \operatorname{Id}_{n\times n} & A \\ 0 & \lambda \operatorname{Id}_{m\times m} - BA \end{pmatrix}.$$

Note that the determinants on the two left-hand sides are equal. Now, equating the determinants on the right-hand sides and canceling λ^m shows the desired identity

$$\det(\lambda \operatorname{Id}_{n\times n} - AB) = \lambda^{n-m} \det(\lambda \operatorname{Id}_{m\times m} - BA) \tag{3.4}$$

for $\lambda \neq 0$. But since both sides here are polynomial (hence continuous) functions of λ, taking limits implies the identity for $\lambda = 0$ as well. (Alternately, AB is singular if $n > m$, which shows the identity for $\lambda = 0$.) □

With Lemma 3.3 at hand, we can prove the following:

Proposition 3.5. *For $A, B \in \mathbb{P}_n$, AB has non-negative eigenvalues.*

Proof. Let $X = \sqrt{A}$ and $Y = \sqrt{A}B$, where $A = U^T DU \implies \sqrt{A} = U^T\sqrt{D}U$. Then $XY = AB$ and $YX = \sqrt{A}B\sqrt{A}$. In fact, YX is (symmetric and) positive semidefinite, since

$$x^T YXx = \|\sqrt{B}\sqrt{A}x\|^2, \quad \forall x \in \mathbb{R}^n.$$

It follows that YX has non-negative eigenvalues, so the same holds by Lemma 3.3 for $XY = AB$, even if AB is not symmetric. □

We next introduce a different multiplication operation on matrices (possibly rectangular, including row or column matrices), which features extensively in this text.

Definition 3.6. Given positive integers m, n, the *Schur product* of $A_{m\times n}$ and $B_{m\times n}$ is the matrix $C_{m\times n}$ with $c_{jk} = a_{jk}b_{jk}$ for $1 \leq j \leq m,\ 1 \leq k \leq n$. We denote the Schur product by $\circ$ (to distinguish it from the conventional matrix product).

Lemma 3.7. *Given integers $m, n \geq 1$, $(\mathbb{R}^{m\times n}, +, \circ)$ is a commutative associative algebra.*

Proof. The easy proof is omitted. More formally, $\mathbb{R}^{m\times n}$ under coordinatewise addition and multiplication is the direct sum (or direct product) of copies of $\mathbb{R}$ under these operations. □

Remark 3.8. Schur products occur in a variety of settings in mathematics. These include the theory of Schur multipliers (introduced by Schur himself in 1911 [329], in the same paper where he shows the Schur product theorem 3.12), association schemes in combinatorics, characteristic functions in probability theory, and the weak minimum principle in partial differential equations. Another application is to products of integral equation kernels and the connection to Mercer's theorem (see Theorem 39.9). Yet another, well-known result connects the functional calculus to this entrywise product: the Daletskii–Krein formula (1956) expresses the Fréchet derivative of $f(\cdot)$ (the usual "functional calculus") at a diagonal matrix A as the Schur product/multiplier against the Loewner matrix L_f of f (see Theorem 16.8). More precisely, let $(a,b) \subset \mathbb{R}$ be open, and $f : (a,b) \to \mathbb{R}$ be C^1. Choose scalars $a < x_1 < \cdots < x_k < b$ and let $A := \operatorname{diag}(x_1, \ldots, x_k)$. Then Daletskii and Krein [98] showed:

$$(Df)(A)(C) := \frac{d}{d\lambda} f(A + \lambda C)\bigg|_{\lambda=0} = L_f(x_1, \ldots, x_k) \circ C, \qquad \forall C = C^* \in \mathbb{C}^{k\times k}, \tag{3.9}$$

where L_f has (j,k) entry $\frac{f(x_j)-f(x_k)}{x_j - x_k}$ if $j \neq k$, and $f'(x_j)$ otherwise. A final appearance of the Schur product that we mention here is to trigonometric moments. Suppose $f_1, f_2 : \mathbb{R} \to \mathbb{R}$ are continuous and 2π-periodic, with Fourier coefficients / trigonometric moments

$$a_j^{(k)} := \int_0^{2\pi} e^{-ik\theta}\ f_j(\theta)\ d\theta, \qquad j = 1, 2,\ k \in \mathbb{Z}.$$

If one defines the convolution product of f_1, f_2, via

$$(f_1 * f_2)(\theta) := \int_0^{2\pi} f_1(t) f_2(\theta - t)\ dt,$$

then this function has corresponding kth Fourier coefficient $a_1^{(k)} a_2^{(k)}$. Thus, the bi-infinite Toeplitz matrix of Fourier coefficients for $f_1 * f_2$ equals the Schur product of the Toeplitz matrices $(a_1^{(p-q)})_{p,q\in\mathbb{Z}}$ and $(a_2^{(p-q)})_{p,q\in\mathbb{Z}}$.

Remark 3.10. The Schur product is also called the *entrywise product* or the *Hadamard product* in the literature; the latter is likely owing to the famous paper by Hadamard [160] in *Acta Math.* (1899), in which he shows (among other things) the Hadamard multiplication theorem. This relates the radii of convergence and singularities of two power series $\sum_{j\geq 0} a_j z^j$

and $\sum_{j\geq 0} b_j z^j$ with those of $\sum_{j\geq 0} a_j b_j z^j$. This "coefficientwise product" will recur later in this text, when we discuss Maló's theorem 33.8(3), in the context of Pólya–Schur multipliers.

Before proceeding further, we define another product on matrices of any dimensions.

Definition 3.11. Given matrices $A_{m\times n}, B_{p\times q}$, the *Kronecker product* of A and B, denoted $A \otimes B$ is the $mp \times nq$ block matrix, defined as:

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{pmatrix}$$

While the Kronecker product (as defined) is asymmetric in its arguments, it is easily seen that the analogous matrix $B \otimes A$ is obtained from $A \otimes B$ by permuting its rows and columns.

The next result, by Schur in *J. reine angew. Math.* (1911), is important later in this text. We provide four proofs.

Theorem 3.12 (Schur product theorem)**.** *$\mathbb{P}_n$ is closed under $\circ$.*

Proof. Suppose $A, B \in \mathbb{P}_n$; we present four proofs that $A \circ B \in \mathbb{P}_n$.

(1) Let $A, B \in \mathbb{P}_n$ have eigenbases (λ_j, v_j) and (μ_k, w_k), respectively. Then,
$$(A \otimes B)(v_j \otimes w_k) = \lambda_j \mu_k (v_j \otimes w_k), \qquad \forall 1 \leq j, k \leq n. \tag{3.13}$$
It follows that the Kronecker product has spectrum $\{\lambda_j \mu_k\}$, and hence is positive (semi)definite if A, B are positive (semi)definite. Hence, every principal submatrix is also positive (semi)definite by Lemma 2.9. But now observe that the principal submatrix of $A \otimes B$ with entries $a_{jk} b_{jk}$ is precisely the Schur product $A \circ B$.

(2) By the spectral theorem and the bilinearity of the Schur product,
$$A = \sum_{j=1}^n \lambda_j v_j v_j^T,\ B = \sum_{k=1}^n \mu_k w_k w_k^T \quad \Longrightarrow \quad A \circ B = \sum_{j,k=1}^n \lambda_j \mu_k (v_j \circ w_k)(v_j \circ w_k)^T.$$
This is a non-negative linear combination of rank-1 positive semidefinite matrices, hence lies in $\mathbb{P}_n$ by Lemma 3.1.

(3) This proof uses a clever computation. Given any commutative ring R, square matrices $A, B \in R^{n\times n}$ and vectors $u, v \in R^n$, we have
$$u^T (A \circ B) v = \operatorname{tr}(B^T D_u A D_v), \tag{3.14}$$
where D_v denotes the diagonal matrix with diagonal entries the coordinates of v (in the same order). Now if $R = \mathbb{R}$ and $A, B \in \mathbb{P}_n(\mathbb{R})$, then
$$v^T (A \circ B) v = \operatorname{tr}(A D_v B^T D_v) = \operatorname{tr}(A^{1/2} D_v B^T D_v A^{1/2}).$$
But $A^{1/2} D_v B D_v A^{1/2}$ is positive semidefinite, so its trace is non-negative, as desired.

(4) Given $t > 0$, let X, Y be independent multivariate normal vectors centered at 0 and with covariance matrices $A + t\,\mathrm{Id}_{n\times n}, B + t\,\mathrm{Id}_{n\times n}$ respectively. (Note that these always exist.) The Schur product of X and Y is then a random vector with mean zero and covariance matrix $(A + t\,\mathrm{Id}_{n\times n}) \circ (B + t\,\mathrm{Id}_{n\times n})$. Now the result follows from the fact that covariance matrices are positive semidefinite, by letting $t \to 0^+$. $\square$

Remark 3.15. The first of the above proofs also shows the *Schur product theorem for (complex) positive definite matrices:* If $A, B \in \mathbb{P}_n(\mathbb{C})$ are positive definite, then so is their Kronecker product $A \otimes B$, hence also its principal submatrix $A \circ B$, the Schur product.

The Schur product theorem is “qualitative” – it says $M \circ N \geq 0$ if $M, N \geq 0$. In the century after its formulation, tight quantitative non-zero lower bounds have been discovered. For instance, in *Linear Algebra Appl.*, Fiedler–Markham (1995) and Reams (1999) showed

$$\begin{aligned} M \circ N &\geq \lambda_{\min}(N)(M \circ \mathrm{Id}_{n\times n}), \\ M \circ N &\geq \frac{1}{\mathbf{1}^T N^{-1} \mathbf{1}} M, \text{ if } \det(N) > 0. \end{aligned} \tag{3.16}$$

We now give a lower bound by Khare, first obtained in a weaker form by Vybíral (2020):

Theorem 3.17. *Fix integers $a, n \geq 1$ and non-zero matrices $A, B \in \mathbb{C}^{n\times a}$. Then we have the (rank ≤ 1) lower bound*

$$AA^* \circ BB^* \geq \frac{1}{\min(\operatorname{rk} AA^*, \operatorname{rk} BB^*)} d_{AB^T} d_{AB^T}^*,$$

where given a square matrix $M = (m_{jk})$, the column vector $d_M := (m_{jj})$. Moreover, the coefficient $1/\min(\cdot,\cdot)$ is best possible.

We make several remarks here. First, if A, B are rank 1, then a stronger result holds: we get equality above (say unlike in (3.16), for instance). Second, reformulating the result in terms of $M = AA^*, N = BB^*$ says that $M \circ N$ is bounded below by many possible rank-1 submatrices, one for every (square) matrix decomposition $M = AA^*, N = BB^*$. Third, the result extends to Hilbert–Schmidt operators, in which case it again provides non-zero positive lower bounds – and in a more general form even on $\mathbb{P}_n(\mathbb{C})$; see a 2021 paper in *Proc. Amer. Math. Soc.* by Khare (following its special case in 2020 in *Adv. Math.* by Vybíral). Finally, specializing to the positive semidefinite square roots $A = \sqrt{M}$ and $B = \sqrt{N}$ yields a novel connection between the matrix functional calculus and entrywise operations on matrices:

$$M \circ N \geq \frac{1}{\min(\operatorname{rk}(M), \operatorname{rk}(N))} d_{\sqrt{M}\sqrt{N}} d^*_{\sqrt{M}\sqrt{N}}, \qquad \forall M, N \in \mathbb{P}_n(\mathbb{C}),\ n \geq 1.$$

Proof. We write down the proof as it serves to illustrate another important tool in matrix analysis: the *trace form* on matrix space $\mathbb{C}^{a\times a}$. Compute using (3.14):

$$u^*(AA^* \circ BB^*)u = \operatorname{tr}(\overline{B}B^T D_{\overline{u}} AA^* D_u) = \operatorname{tr}(T^*T), \quad \text{where} \quad T := A^* D_u \overline{B}.$$

Use the inner product on $\mathbb{C}^{a\times a}$, given by $\langle X, Y \rangle := \operatorname{tr}(X^*Y)$. Define the projection operator

$$P := \operatorname{proj}_{(\ker A)^\perp}|_{\operatorname{im}(B^T)};$$

thus, $P \in \mathbb{C}^{a\times a}$. Now compute

$$\langle P, P \rangle \leq \min(\dim(\ker A)^\perp, \dim \operatorname{im}(B^T)) = \min(\operatorname{rk}(A^*), \operatorname{rk}(B^*)) = \min(\operatorname{rk}(AA^*), \operatorname{rk}(BB^*)).$$

Here we use that AA^* and A^* have the same null space, since

$$A^*x = 0 \implies AA^*x = 0 \implies \|A^*x\|^2 = 0 \implies A^*x = 0,$$

and hence the same rank. Now using the Cauchy–Schwarz inequality (for this tracial inner product) and the above computations, we have

$$\begin{aligned} u^*(AA^* \circ BB^*)u = \langle T, T \rangle &\geq \frac{|\langle T, P \rangle|^2}{\langle P, P \rangle} = \frac{|\operatorname{tr}(APB^T D_{\overline{u}})|^2}{\langle P, P \rangle} = \frac{|u^* d_{APB^T}|^2}{\langle P, P \rangle} \\ &\geq \frac{1}{\min(\operatorname{rk}(AA^*), \operatorname{rk}(BB^*))} u^* d_{APB^T} d^*_{APB^T} u. \end{aligned}$$

As this holds for all vectors u, the result follows because by the choice of P, we have $APB^T = AB^T$.

Finally, to see the tightness of the lower bound, choose integers $1 \leq r, s \leq a$ with $r, s \leq n$, and complex block diagonal matrices

$$A_{n\times a} := \begin{pmatrix} D_{r\times r} & 0 \\ 0 & 0 \end{pmatrix}, \quad B_{n\times a} := \begin{pmatrix} D'_{s\times s} & 0 \\ 0 & 0 \end{pmatrix},$$

with both D, D' non-singular. Now $P = \mathrm{Id}_{\min(r,s)} \oplus \mathbf{0}_{a-\min(r,s)}$, and the inequality is indeed tight, as can be verified using the Cauchy–Schwarz identity. $\square$

Remark 3.18. Vybíral also provided a simpler lower bound for Schur products for square matrix decompositions: if $A, B \in \mathbb{C}^{n\times n}$, then

$$AA^* \circ BB^* \geq (A \circ B)(A \circ B)^*,$$

where both sides are positive semidefinite. Indeed, if v_j, w_k denote the columns of A, B respectively, then $AA^* = \sum_j v_j v_j^*$ and $BB^* = \sum_k w_k w_k^*$, so

$$AA^* \circ BB^* = \sum_{j,k=1}^{n} (v_j v_j^*) \circ (w_k w_k^*) \geq \sum_{j=1}^{n} (v_j \circ w_j)(v_j^* \circ w_j^*) = (A \circ B)(A \circ B)^*.$$

3.2. **Totally Positive (TP) and Totally Non-negative (TN) matrices.**

Definition 3.19. Given an integer $p \geq 1$, we say a matrix is *totally positive (totally non-negative) of order* p, denoted TP_p (TN_p), if all its $1 \times 1, 2 \times 2, \dots,$ and $p \times p$ minors are positive (non-negative). We will also abuse notation and write $A \in TP_p$ ($A \in TN_p$) if A is TP_p (TN_p). A matrix is *totally positive (TP)* (respectively, *totally non-negative (TN)*) if A is TP_p (respectively TN_p) for all $p \geq 1$.

Remark 3.20. In classical works, as well as the books by Karlin and Pinkus, totally non-negative and totally positive matrices were referred to, respectively, as *totally positive* and *strictly totally positive matrices.*

Here are some distinctions between TP/TN matrices and positive (semi)definite ones:

- For TP/TN matrices we consider all minors, not just the principal ones.
- As a consequence of considering the 1×1 minors, it follows that the entries of TP (TN) matrices are all positive (non-negative).
- TP/TN matrices need not be symmetric, unlike positive semidefinite matrices.

Example 3.21. The matrix $\begin{pmatrix} 1 & 2 \\ 3 & 16 \end{pmatrix}$ is totally positive, while the matrix $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ is totally non-negative but not totally positive.

Totally positive matrices and kernels have featured in the mathematics literature in a variety of classical and modern topics. A few of these topics are now listed, as well as a few of the experts who have worked/written on them:

- Interacting particle systems, mechanics, and physics (Gantmacher–Krein), see [137, 138] and follow-up papers, as well as [74] and the references therein.
- Analysis (Aissen, Edrei, Pólya, Schoenberg, Whitney, Hirschman and Widder), see Part 4 for details. (Also Loewner [241] and the author, including with Belton–Guillot–Putinar [32, 29] and [215].)
- Differential equations and applications – see, e.g., Loewner [241], Schwarz [332], the book of Karlin [200], and numerous follow-ups, e.g. [254].
- Probability and statistics (Efron, Karlin, McGregor, Pitman, Rinott), see, e.g., [109, 199, 200, 201, 202, 204, 205, 218, 283] and numerous follow-ups.

- Matrix theory and applications (Ando, Cryer, Fallat, Garloff, Johnson, Pinkus, Sokal, Wagner), see, e.g., [12, 93, 94, 112, 113, 139, 140, 141, 282, 357].
- Gabor analysis (Gröchenig, Romero, Stöckler), see [148, 149] and related references.
- Interpolation theory and splines – see, e.g., numerous works by Schoenberg (e.g., with Curry, Whitney), de Boor, Karlin, Micchelli (e.g., [59, 95, 96, 206, 262, 319, 327]).
- Combinatorics (Brenti, Gessel–Viennot, Karlin–McGregor, Lindström, Skandera, Sturmfels), see, e.g., [69, 70, 142, 202, 238, 339, 349] and follow-up works, for example, [362].
- Representation theory, flag varieties, and canonical bases (Goodearl–Launois–Lenagan, Lusztig, Postnikov, Rietsch), e.g., [146, 245, 246, 292, 299, 300] and follow-up papers.
- Cluster algebras (Berenstein, Fomin, Zelevinsky), see, for example, [35, 36, 37, 125, 126, 127] and numerous follow-up papers.
- Quadratic algebras, Witt vector theory (Borger, Davydov, Grinberg, Hô Hai, Skryabin), see [66, 67, 99, 177, 340].
- Integrable systems (Kodama, Williams), see, e.g., [221, 222].

A very important, and widely used, property of TN_p matrices and kernels K is their *variation diminishing property*. Roughly speaking, if a vector u (or a function f on an interval J) has finitely many sign changes in its values, say s, then the vector Ku (or the function $\int_J K(u,x)\,dx$) has at most $\min(p,s)$ sign changes. In *Math. Z.* in 1930, Schoenberg showed that if K is a TN matrix, then $S^-(Kx) \leq S^-(x)$ $\forall x \in \mathbb{R}^n$. (See Section 29.1 for a proof.) A characterization of such matrices was then shown by Motzkin in his thesis:

Theorem 3.22 (Motzkin, 1936, [263])**.** *The following are equivalent for a matrix $K \in \mathbb{R}^{m\times n}$:*

(1) K is variation diminishing: $S^-(Kx) \leq S^-(x)$ $\forall x \in \mathbb{R}^n$. Here, $S^-(x)$ for a vector x denotes the number of changes in sign, after removing all zero entries in x.

(2) Let K have rank r. Then K should not have two minors of equal size $< r$ but opposite signs; and K should not have two minors of equal size $= r$ but opposite signs if these minors come from the same rows or columns of K.

In this section and the next two, we discuss examples of TN matrices. We begin by showing that the (positive semidefinite) Toeplitz cosine matrices and Hankel moment matrices considered above are in fact totally non-negative.

Example 3.23 (Toeplitz cosine matrices)**.** We claim that the following matrices are TN:

$$C(\theta) := (\cos(j-k)\theta)_{j,k=1}^n, \quad \text{where } \theta \in [0, \tfrac{\pi}{2(n-1)}].$$

Indeed, all 1×1 minors are non-negative, and as discussed above, $C(\theta)$ has rank at most 2, and so all 3×3 and larger minors vanish. It remains to consider all 2×2 minors. Now a 2×2 submatrix of $C(\theta)$ is of the form $C' = \begin{pmatrix} C_{ab} & C_{ac} \\ C_{db} & C_{dc} \end{pmatrix}$, where $1 \leq a < d \leq n$, $1 \leq b < c \leq n$, and C_{ab} denotes the matrix entry $\cos((a-b)\theta)$. Writing a,b,c,d in place of $a\theta, b\theta, c\theta, d\theta$ in the next computation for ease of exposition, the corresponding minor is

$$\begin{aligned}
\det C' &= \frac{1}{2}\big\{2\cos(a-b)\cos(d-c) - 2\cos(a-c)\cos(d-b)\big\} \\
&= \frac{1}{2}\big\{\cos(a-b+d-c) + \cos(a-b-d+c) - \cos(a-c+d-b) - \cos(a-c-d+b)\big\} \\
&= \frac{1}{2}\big\{\cos(a-b-d+c) - \cos(a-c-d+b)\big\} = \frac{1}{2}(-2)\sin(a-d)\sin(c-b).
\end{aligned}$$

Thus, $\det C' = \sin(d\theta - a\theta)\sin(c\theta - b\theta)$, which is non-negative because $a < d, b < c$, and $\theta \in [0, \frac{\pi}{2(n-1)}]$. This shows that $C(\theta)$ is totally non-negative.

4. Fekete's result. Hankel moment matrices are TN. Hankel positive semidefinite and TN matrices.

Continuing from the previous section, we show that the Hankel moment matrices H_μ studied above are not only positive semidefinite, but more strongly, totally non-negative (TN). Akin to the Toeplitz cosine matrices (where the angle θ is restricted to ensure the entries are non-negative), we restrict the support of the measure to $[0, \infty)$, which guarantees that the entries are non-negative.

To achieve these goals, we prove the following result, which is crucial in relating positive semidefinite matrices (and kernels) and their preservers, to Hankel TN matrices (and kernels) and their preservers, (later in this text):

Theorem 4.1. *If $1 \leq p \leq n$ are integers, and $A_{n\times n}$ is a real Hankel matrix, then A is TP_p (TN_p) if and only if all contiguous principal submatrices of both A and $A^{(1)}$, of order $\leq p$, are positive (semi)definite, Here $A^{(1)}$ is obtained from A by removing the first row and last column; and by a contiguous submatrix (or minor) we mean (the determinant of) a square submatrix corresponding to successive rows and to successive columns.*

In particular, A is TP (TN) if and only if A and $A^{(1)}$ are positive (semi)definite.

From this theorem, we derive the following two consequences, both of which are useful later. The first follows from the fact that all contiguous submatrices of a Hankel matrix are Hankel, hence symmetric:

Corollary 4.2. *For all integers $1 \leq p \leq n$, the set of Hankel TN_p $n \times n$ matrices is a closed, convex cone, further closed under taking Schur products.*

The second corollary of Theorem 4.1 provides a large class of examples of such Hankel TN matrices:

Corollary 4.3. *Suppose μ is a non-negative measure supported in $[0, \infty)$, with all moments finite. Then H_μ is TN.*

The proofs are left as easy exercises; the second proof uses Lemma 2.22.

Remark 4.4. Akin to Lemma 2.22 and the remark following its proof, Corollary 4.3 is also the easy half of a well-known classical result on moment problems – this time, by Stieltjes. The harder converse of Stieltjes' result says (in particular) that if a semi-infinite Hankel matrix H is TN, with (j, k)-entry $s_{j+k} \geq 0$ for $j, k \geq 0$, then there exists a non-negative Borel measure μ on $\mathbb{R}$ with support in $[0, \infty)$, whose kth moment is s_k for all $k \geq 0$. By Theorem 4.1, this is equivalent to both H as well as $H^{(1)}$ being positive semidefinite, where $H^{(1)}$ is obtained by truncating the first row (or the first column) of H.

In the 1890s, Stieltjes was working on continued fractions and divergent series, following Euler, Laguerre, Hermite, and others. One result that is relevant here is that Stieltjes produced a non-zero function $\varphi : [0, \infty) \to \mathbb{R}$, such that $\int_0^\infty x^k \varphi(x)\, dx = 0$ for all $k = 0, 1, \dots$ – an indeterminate moment problem. (The work in this setting also led him to develop the Stieltjes integral; see, e.g., [220] for a detailed historical account.) This marks the beginning of his exploration of the moment problem, which he resolved in his well-known memoir [347].

A curious follow-up, by Boas in *Bull. Amer. Math. Soc.* in 1939, is that if one replaces the non-negativity of the Borel measure μ by the hypothesis of being of the form $d\alpha(t)$ on $[0, \infty)$ with α of bounded variation, then this recovers *all* real sequences! See [53].

The remainder of this section is devoted to showing Theorem 4.1. The proof uses a sequence of lemmas shown by Gantmacher–Krein, Fekete, Schoenberg, and others. The first of these lemmas may be (morally) attributed to Laplace.

Lemma 4.5. *Let $r \geq 1$ be an integer and $U = (u_{jk})$ an $(r+2) \times (r+1)$ matrix. Given subsets $[a,b],[c,d] \subset (0,\infty)$, let $U_{[a,b]\times[c,d]}$ denote the submatrix of U with entries u_{jk}, such that j,k are integers and $a \leq j \leq b$, $c \leq k \leq d$. Then*

$$\begin{aligned} & \det U_{[1,r]\cup\{r+2\}\times[1,r+1]} \cdot \det U_{[2,r+1]\times[1,r]} \\ = \; & \det U_{[2,r+2]\times[1,r+1]} \cdot \det U_{[1,r]\times[1,r]} \\ & + \det U_{[1,r+1]\times[1,r+1]} \cdot \det U_{[2,r]\cup\{r+2\}\times[1,r]}. \end{aligned} \tag{4.6}$$

Note that in each of the three products of determinants, the second factor in the subscript for the first (respectively second) determinant terms is the same: $[1,r+1]$ (respectively $[1,r]$).

To give a feel for the result, the special case of $r=1$ asserts that

$$u_{11}\begin{vmatrix} u_{21} & u_{22} \\ u_{31} & u_{32} \end{vmatrix} - u_{21}\begin{vmatrix} u_{11} & u_{12} \\ u_{31} & u_{32} \end{vmatrix} + u_{31}\begin{vmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{vmatrix} = 0.$$

But this is precisely the Laplace expansion along the third column of the singular matrix

$$\det \begin{pmatrix} u_{11} & u_{12} & u_{11} \\ u_{21} & u_{22} & u_{21} \\ u_{31} & u_{32} & u_{31} \end{pmatrix} = 0.$$

Proof. Consider the $(2r+1) \times (2r+1)$ block matrix of the form

$$M = \left(\begin{array}{cc|c} \mathbf{b}^T & u_{1,r+1} & \mathbf{b}^T \\ A & \mathbf{a} & A \\ \mathbf{c}^T & u_{r+1,r+1} & \mathbf{c}^T \\ \mathbf{d}^T & u_{r+2,r+1} & \mathbf{d}^T \\ \hline A & \mathbf{a} & 0_{(r-1)\times r} \end{array}\right) = \begin{pmatrix} (u_{jk})_{j\in[1,r+2],\, k\in[1,r+1]} & (u_{jk})_{j\in[1,r+2],\, k\in[1,r]} \\ (u_{jk})_{j\in[2,r],\, k\in[1,r+1]} & 0_{(r-1)\times r} \end{pmatrix};$$

that is, where

$$\begin{aligned} \mathbf{a} = (u_{2,r+1},\dots,u_{r,r+1})^T, \quad \mathbf{b} =& (u_{1,1},\dots,u_{1,r})^T, \quad \mathbf{c} = (u_{r+1,1},\dots,u_{r+1,r})^T, \\ \mathbf{d} = (u_{r+2,1},\dots,u_{r+2,r})^T, &\quad A = (u_{jk})_{2\leq j\leq r,\, 1\leq k\leq r}. \end{aligned}$$

Notice that M is a square matrix whose first $r+2$ rows have column space of dimension at most $r+1$; hence $\det M = 0$. Now we compute $\det M$ using the (generalized) Laplace expansion by complementary minors: choose all possible $(r+1)$-tuples of rows from the first $r+1$ columns to obtain a submatrix $M'_{(r+1)}$, and deleting these rows and columns from M yields the complementary $r \times r$ submatrix $M''_{(r)}$ from the final r columns. The generalized Laplace expansion says that if one multiplies $\det M'_{(r+1)} \cdot \det M''_{(r)}$ by $(-1)^{\Sigma}$, with Σ the sum of the row numbers in $M'_{(r+1)}$, then summing over all such products (running over subsets of rows) yields $\det M$ – which vanishes for this particular matrix M.

Now in the given matrix, to avoid obtaining zero terms, the rows in $M'_{(r+1)}$ must include all entries from the final $r-1$ rows (and the first $r+1$ columns). But then it, moreover, cannot include entries from the rows of M labeled $2,\dots,r$; and it must include two of the remaining three rows (and entries from only the first $r+1$ columns).

Thus, we obtain three product terms that sum to: $\det M = 0$. Upon carefully examining the terms and computing the companion signs (by row permutations), we obtain (4.6). $\square$

The next two results are classical facts about totally positive matrices, first shown by Fekete in his correspondence [118] with Pólya, published in *Rend. Circ. Mat. Palermo.* (As an aside, that correspondence was published in 1912 under the title "Über ein Problem von Laguerre". This problem – from Laguerre's 1883 paper – and its connection to TN matrices and to the variation diminishing property alluded to in the preceding section, are discussed in detail in Section 29.3.)

Lemma 4.7. *Given integers $m \geq n \geq 1$ and a real matrix $A_{m \times n}$, such that*
(a) all $(n-1) \times (n-1)$ minors $\det A_{J \times [1,n-1]} > 0$ for $J \subset [1,m]$ of size $n-1$, and
(b) all $n \times n$ minors $\det A_{[j+1,j+n] \times [1,n]} > 0$ for $0 \leq j \leq m-n$,
we have that all $n \times n$ minors of A are positive.

Proof. Define the *gap*, or "index" of a subset of integers $J = \{j_1 < j_2 < \cdots < j_n\} \subset [1,m]$, to be $g_J := j_n - j_1 - (n-1)$. Thus, the gap is zero if and only if J consists of successive integers, and in general it counts precisely the number of integers between j_1 and j_n that are missing from J.

We claim that $\det A_{J \times [1,n]} > 0$ for $|J| = n$, by induction on the gap $g_J \geq 0$; the base case $g_J = 0$ is given as hypothesis. For the induction step, suppose j^0 is a missing index (or row number) in $J = \{j_1 < \cdots < j_n\}$. By suitably specializing the identity (4.6), we obtain

$$\begin{aligned}
& \det A_{(j_1,\dots,j_n) \times [1,n]} \cdot \det A_{(j_2,\dots,j_{n-1},j^0) \times [1,n-1]} \\
= \ & \det A_{(j_1,\dots,j_{n-1},j^0) \times [1,n]} \cdot \det A_{(j_2,\dots,j_n) \times [1,n-1]} \\
& - \det A_{(j_2,\dots,j_n,j^0) \times [1,n]} \cdot \det A_{(j_1,\dots,j_{n-1}) \times [1,n-1]}.
\end{aligned} \tag{4.8}$$

Consider the six factors in serial order. The first, fourth, and sixth factors have indices listed in increasing order, while the other three factors have j^0 listed at the end, so their indices are not listed in increasing order. For each of the six factors, the number of "bubble sorts" required to rearrange indices in increasing order (by moving j^0 down the list) equals the number of row exchanges in the corresponding determinants; label these numbers $n_1, \dots, n_6$. Thus, $n_1 = n_4 = n_6 = 0$ as above, while $n_2 = n_3$ (since $j_1 < j^0 < j_n$), and $|n_2 - n_5| = 1$. Now multiply Equation (4.8) by $(-1)^{n_2}$, and divide both sides by

$$c_0 := (-1)^{n_2} \det A_{(j_2,\dots,j_{n-1},j^0) \times [1,n-1]} > 0.$$

Using the given hypotheses as well as the induction step (since all terms involving j^0 have a gap equal to $g_J - 1$), it follows that

$$\begin{aligned}
& \det A_{(j_1,\dots,j_n) \times [1,n]} \\
= \ & c_0^{-1} \big((-1)^{n_2} \det A_{(j_1,\dots,j_{n-1},j^0) \times [1,n]} \cdot \det A_{(j_2,\dots,j_n) \times [1,n-1]} \\
& \qquad + (-1)^{n_2+1} \det A_{(j_2,\dots,j_n,j^0) \times [1,n]} \cdot \det A_{(j_1,\dots,j_{n-1}) \times [1,n-1]} \big) \\
> \ & 0.
\end{aligned}$$

This completes the induction step, and with it, the proof. □

We can now state and prove another 1912 result by Fekete for TP matrices – extended to TP_p matrices by Schoenberg in *Ann. of Math.* 1955:

Lemma 4.9 (Fekete–Schoenberg lemma)**.** *Suppose $m, n \geq p \geq 1$ are integers, and $A \in \mathbb{R}^{m \times n}$ is a matrix, all of whose contiguous minors of order at most p are positive. Then A is TP_p.*

Notice that the analogous statement for TN_p is false, e.g., $p = 2$ and $A = \begin{pmatrix} 1 & 0 & 2 \\ 1 & 0 & 1 \end{pmatrix}$.

Proof. We show that for any integer $s \in [1, p]$, every $s \times s$ minor of A is positive. The proof is by induction on s (and running over all real matrices satisfying the hypotheses); note that the base case of $s = 1$ is immediate from the assumptions. For the induction step, suppose

$$2 \leq s = |J| = |K| \leq p, \qquad J \subset \mathbb{Z} \cap [1, m], \ K \subset \mathbb{Z} \cap [1, n].$$

First fix a subset K that consists of consecutive rows, i.e., has gap $g_K = 0$ (as in the proof of Lemma 4.7). Let B denote the submatrix $A_{[1,m] \times K}$. Then all $s \times s$ minors of B are positive, by Lemma 4.7. In particular, it follows for all J that all $s \times s$ minors $\det A_{J \times K'}$ are positive, whenever $K' \subset [1, n]$ has size s and gap $g_{K'} = 0$. Now apply Lemma 4.7 to the matrix $B := (A_{J \times [1,n]})^T$ to obtain: $\det(A_{J,K})^T > 0$ for (possibly non-consecutive subsets) K. This concludes the proof. $\square$

The final ingredient required to prove Theorem 4.1 is the following result:

Lemma 4.10. *If $A_{n \times n}$ is a Hankel matrix, then every contiguous minor of A (see Lemma 4.9) is a contiguous principal minor of A or of $A^{(1)}$.*

Recall that $A^{(1)}$ was defined in Theorem 4.1.

Proof. Let the first row (respectively, last column) of A contain the entries $s_0, s_1, \dots, s_{n-1}$ (respectively, $s_{n-1}, s_n, \dots, s_{2n-2}$). Then every contiguous minor is the determinant of a submatrix of the form

$$M = \begin{pmatrix} s_j & \cdots & s_{j+m} \\ \vdots & \ddots & \vdots \\ s_{j+m} & \cdots & s_{j+2m} \end{pmatrix}, \qquad 0 \leq j \leq j + m \leq n - 1.$$

It is now immediate that if j is even (respectively odd), then M is a contiguous principal submatrix of A (respectively $A^{(1)}$). $\square$

With these results in hand, we conclude by proving the above theorem.

Proof of Theorem 4.1. If the Hankel matrix A is TN_p (TP_p), then all contiguous minors of A of order $\leq p$ are non-negative (positive), proving one implication. Conversely, suppose all contiguous principal minors of A and $A^{(1)}$ of order $\leq p$ are positive. By Lemma 4.10, this implies every contiguous minor of A of order $\leq p$ is positive. By the Fekete–Schoenberg Lemma 4.9, A is TP_p as desired.

Finally, suppose all contiguous principal minors of $A, A^{(1)}$ or size $\leq p$ are non-negative. It follows by Lemma 4.10 that every contiguous square submatrix of A of order $\leq p$ is positive semidefinite. Also choose and fix an $n \times n$ Hankel TP matrix B (note by (5.9) or Lemma 6.9 below that such matrices exist for all $n \geq 1$). Applying Lemma 4.10, $B, B^{(1)}$ are positive definite, hence so is every contiguous square submatrix of B. Now for $\epsilon > 0$, it follows (by Sylvester's criterion, Theorem 2.8) that every contiguous principal minor of $A + \epsilon B$ of size $\leq p$ is positive. Again applying Lemma 4.9, the Hankel matrix $A + \epsilon B$ is necessarily TP_p, and taking $\epsilon \to 0^+$ finishes the proof.

The final statement is the special case $p = n$. $\square$

5. Generalized Vandermonde matrices. Cauchy–Binet formula and its generalization.

The previous sections discussed examples (Toeplitz, Hankel) of totally non-negative (TN) matrices. These examples consisted of symmetric matrices.

- We will now look at some examples of non-symmetric matrices that are totally positive (TP). We then prove the Cauchy–Binet formula, which will lead to the construction of additional examples of symmetric TP matrices.
- Let $H_\mu := (s_{j+k}(\mu))_{j,k\geq 0}$ denote the moment matrix associated to a non-negative measure μ supported on $[0,\infty)$. We have already seen that this matrix is Hankel and positive semidefinite – in fact, TN. We will show in this section and the next that H_μ is in fact TP in many cases. The proof will use a continuous generalization of the Cauchy–Binet formula.

5.1. **Generalized Vandermonde Matrices.** A *generalized Vandermonde matrix* is a matrix $(x_j^{\alpha_k})_{j,k=1}^n$, where $x_j > 0$ and $\alpha_j \in \mathbb{R}$ for all j. If the x_j are pairwise distinct, as are the α_k, then the corresponding generalized Vandermonde matrix is non-singular. In fact, a stronger result holds:

Theorem 5.1. *If $0 < x_1 < \cdots < x_m$ and $\alpha_1 < \cdots < \alpha_n$ are real numbers, then the generalized Vandermonde matrix $V_{m\times n} := (x_j^{\alpha_k})$ is totally positive.*

As an illustration, consider the special case $m = n$ and $\alpha_k = k-1$, which recovers the usual Vandermonde matrix

$$V = \begin{pmatrix} 1 & x_1 & \cdots & x_1^{n-1} \\ 1 & x_2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^{n-1} \end{pmatrix}.$$

This matrix has determinant $\displaystyle\prod_{1\leq j<k\leq n}(x_k - x_j) > 0$. Thus, if $0 < x_1 < x_2 < \cdots < x_n$ then $\det V > 0$. However, note this is not enough to prove that the matrix is totally positive. Thus, more work is required to prove total positivity, even for usual Vandermonde matrices. The following 1637 result by Descartes will help in the proof. (Curiously, this preliminary result is also the beginning of a mathematical journey that led to both total positivity and to the variation diminishing property! See Section 29.3 for the details.)

Lemma 5.2 (Descartes' rule of signs, weaker version)**.** *Fix pairwise distinct real numbers $\alpha_1, \alpha_2, \dots, \alpha_n \in \mathbb{R}$ and n scalars $c_1, c_2, \dots, c_n \in \mathbb{R}$, such that not all scalars are 0. Then the function $f(x) := \sum_{k=1}^n c_k x^{\alpha_k}$ can have at most $(n-1)$ distinct positive roots.*

The following proof is due to Laguerre (1883); that said, the trick of multiplying by a faster-decaying function and applying Rolle's theorem was previously employed by Poulain in 1867. See Theorem 33.3.

Proof. By induction on n. For $n = 1$, clearly $f(x)$ has no positive root. For the induction step, without loss of generality we may assume $\alpha_1 < \alpha_2 < \cdots < \alpha_n$ and that all c_j are non-zero. If f has n distinct positive roots, then so does the function

$$g(x) := x^{-\alpha_1} f(x) = c_1 + \sum_{k=2}^n c_k x^{\alpha_k - \alpha_1}.$$

But then Rolle's theorem implies that $g'(x) = \sum_{k=2}^n c_k(\alpha_k - \alpha_1)x^{\alpha_k - \alpha_1 - 1}$ has $(n-1)$ distinct positive roots. This contradicts the induction hypothesis, completing the proof. $\square$

With this result in hand, we can prove that generalized Vandermonde matrices are TP.

Proof of Theorem 5.1. As any submatrix of V is also a generalized Vandermonde matrix, it suffices to show that the determinant of V is positive when $m = n$.

We first claim that $\det V \neq 0$. Indeed, suppose for contradiction that V is singular. Then there is a non-zero vector $\mathbf{c} = (c_1, c_2, \ldots, c_n)^T$, such that $V\mathbf{c} = 0$. But then there exist n distinct positive numbers $x_1, x_2, \ldots, x_n$, such that $\sum_{k=1}^n c_k x_j^{\alpha_k} = 0$, which contradicts Lemma 5.2 for $f(x) = \sum_{k=1}^n c_k x^{\alpha_k}$. Thus, the claim follows.

We now prove the theorem via a homotopy argument. Consider a (continuous) path $\gamma : [0,1] \to \mathbb{R}^n$ going from $\gamma(0) = (0, 1, \ldots, n-1)$ to $\gamma(1) = (\alpha_1, \alpha_2, \ldots, \alpha_n)$, such that at each timepoint $t \in [0,1]$, the coordinates of $\gamma(t)$ are in increasing order. It is possible to choose such a path; indeed, the straight line geodesic path is one such path.

Now let $W(t) := \det\, (x_j^{\gamma_k(t)})_{j,k=1}^n$. Then $W : [0,1] \to \mathbb{R}$ is a continuous map that never vanishes. Since $[0,1]$ is connected and $W(0) > 0$ (see remarks above), it follows that $W(1) = \det V > 0$. $\square$

Remark 5.3. If we have $0 < x_n < x_{n-1} < \cdots < x_1$ and $\alpha_n < \alpha_{n-1} < \cdots < \alpha_1$, then observe that the corresponding generalized Vandermonde matrix $V' := (x_j^{\alpha_k})_{j,k=1}^n$ is also TP. Indeed, once again we only need to show $\det V' > 0$, and this follows from applying the same permutation to the rows and to the columns of V' to reduce it to the situation in Theorem 5.1 (since then the determinant does not change in sign).

5.2. **The Cauchy–Binet formula.**

The following is a recipe to construct new examples of TP/TN matrices from known ones.

Proposition 5.4. *If $A_{m\times n}, B_{n\times k}$ are both TN, then so is the matrix $(AB)_{m\times k}$. This assertion is also valid upon replacing "TN" by "TP", provided $n \geq \min\{m,k\}$.*

To prove this proposition, we require the following important result.

Theorem 5.5 (Cauchy–Binet formula)**.** *Given matrices $A_{m\times n}$ and $B_{n\times m}$, we have*

$$\det(AB)_{m\times m} = \sum_{J\subset[n] \text{ of size } m} \det(A_{[m]\times J^{\uparrow}}) \det(B_{J^{\uparrow}\times[m]}), \tag{5.6}$$

where $J^{\uparrow}$ reiterates the fact that the elements of J are arranged in increasing order.

For example, if $m = n$, this theorem just reiterates the fact that the determinant map is multiplicative on square matrices. If $m > n$, the theorem says that determinants of singular matrices are zero. If $m = 1$, we obtain the inner product of a row and column vector.

Proof. Notice that

$$\det(AB) = \det \begin{pmatrix} \displaystyle\sum_{j_1=1}^n a_{1j_1}b_{j_1 1} & \cdots & \displaystyle\sum_{j_m=1}^n a_{1j_m}b_{j_m m} \\ \vdots & \ddots & \vdots \\ \displaystyle\sum_{j_1=1}^n a_{mj_1}b_{j_1 1} & \cdots & \displaystyle\sum_{j_m=1}^n a_{mj_m}b_{j_m m} \end{pmatrix}_{m\times m}.$$

By the multilinearity of the determinant, expanding $\det(AB)$ as a sum over all j_l yields

$$\det(AB) = \sum_{(j_1,j_2,\dots,j_m)\in[n]^m} b_{j_1 1} b_{j_2 2}\cdots b_{j_m m}\cdot \det\begin{pmatrix} a_{1j_1} & \cdots & a_{1j_m} \\ \vdots & \ddots & \vdots \\ a_{mj_1} & \cdots & a_{mj_m}\end{pmatrix}.$$

The determinant in the summand vanishes if $j_k = j_m$ for any $k \neq m$. Therefore,

$$\begin{aligned}\det(AB) &= \sum_{\substack{(j_1,j_2,\dots,j_m)\in[n]^m,\\ \text{all } j_l \text{ are distinct}}} b_{j_1 1} b_{j_2 2}\cdots b_{j_m m}\cdot \det\begin{pmatrix} a_{1j_1} & \cdots & a_{1j_m} \\ \vdots & \ddots & \vdots \\ a_{mj_1} & \cdots & a_{mj_m}\end{pmatrix}\\ &= \sum_{\substack{(j_1,j_2,\dots,j_m)\in[n]^m,\\ \text{all } j_l \text{ are distinct}}} b_{j_1 1} b_{j_2 2}\cdots b_{j_m m}\cdot \det A_{[m]\times(j_1,j_2,\dots,j_m)}.\end{aligned}$$

We split this sum into two subsummations. One part runs over all collections of indices, while the other runs over all possible orderings – that is, permutations – of each fixed collection of indices. Thus, for each ordering $\mathbf{j} = (j_1,\dots,j_m)$ of $J = \{j_1,\dots,j_m\}$, there exists a unique permutation $\sigma_{\mathbf{j}} \in S_m$, such that $(j_1,\dots,j_m) = \sigma_{\mathbf{j}}(J^{\uparrow})$. Now,

$$\begin{aligned}\det(AB) &= \sum_{J=\{j_1,j_2,\dots,j_m\}\subset[n],\ \text{all } j_l \text{ distinct}}\ \sum_{\sigma=\sigma_{\mathbf{j}}\in S_m} b_{j_1 1} b_{j_2 2}\cdots b_{j_m m}(-1)^{\sigma_{\mathbf{j}}} \det A_{[m]\times J^{\uparrow}}\\ &= \sum_{J\subset[n] \text{ of size } m} \det(A_{[m]\times J^{\uparrow}})\det(B_{J^{\uparrow}\times[m]}).\end{aligned}$$ $\square$

Proof of Proposition 5.4. Suppose two matrices $A_{m\times n}$ and $B_{n\times k}$ are both TN. Let $I \subset [m]$ and $K \subset [k]$ be index subsets of the same size; we are to show $\det(AB)_{I\times K}$ is non-negative. Define matrices, $A' := A_{I\times[n]}$ and $B' := B_{[n]\times K}$. Now it is easy to show that $(AB)_{I\times K} = A'B'$. In particular, $\det(AB)_{I\times K} = \det(A'B')$. Hence, the Cauchy–Binet theorem implies

$$\det(AB)_{I\times K} = \sum_{\substack{J\subset[n],\\ |J|=|K|=|I|}} \det A'_{I\times J^{\uparrow}} \det B'_{J^{\uparrow}\times K} = \sum_{\substack{J\subset[n],\\ |J|=|K|=|I|}} \det A_{I\times J^{\uparrow}} \det B_{J^{\uparrow}\times K} \geq 0. \quad (5.7)$$

It follows that AB is TN if both A and B are TN. For the corresponding TP-version, the above proof works as long as the sums in the preceding equation are always over non-empty sets; but this happens whenever $n \geq \min\{m,k\}$. $\square$

Remark 5.8. This proof shows that Proposition 5.4 holds upon replacing TN/TP by TP_p for any $1 \leq p \leq n$. (E.g. the condition TP_n coincides with TP_{n-1} if $\min\{m,k\} < n$.)

5.3. **Generalized Cauchy–Binet formula.** We showed in Section 5.1 that generalized Vandermonde matrices are examples of totally positive but non-symmetric matrices. Using these, we can construct additional examples of totally positive symmetric matrices: let $V = (x_j^{\alpha_k})_{j,k=1}^n$ be a be a generalized Vandermonde matrix with $0 < x_1 < x_2 < \cdots < x_n$ and $\alpha_1 < \alpha_2 < \cdots < \alpha_n$. Then Proposition 5.4 implies that the symmetric matrices V^TV and VV^T are totally positive.

For instance, if we take $n = 3$ and $\alpha_k = k-1$, then

$$V = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \end{pmatrix}, \qquad V^TV = \begin{pmatrix} 3 & \sum_{j=1}^3 x_j & \sum_{j=1}^3 x_j^2 \\ \sum_{j=1}^3 x_j & \sum_{j=1}^3 x_j^2 & \sum_{j=1}^3 x_j^3 \\ \sum_{j=1}^3 x_j^2 & \sum_{j=1}^3 x_j^3 & \sum_{j=1}^3 x_j^4 \end{pmatrix} \quad (5.9)$$

This is clearly the Hankel moment matrix H_μ for the counting measure on the set $\{x_1, x_2, x_3\}$. Moreover, $V^T V$ is (symmetric and) totally positive by the Cauchy–Binet formula. More generally, for all increasing α_k which are in arithmetic progression, the matrix $V^T V$ (defined similarly as above) is a totally positive Hankel moment matrix for some non-negative measure on $[0,\infty)$ – more precisely, supported on $\{x_1^{\alpha_2-\alpha_1}, x_2^{\alpha_2-\alpha_1}, \dots, x_n^{\alpha_2-\alpha_1}\}$.

The following discussion aims to show (among other things) that the moment matrices H_μ defined in (2.21) are totally positive for "most" non-negative measures μ supported in $[0,\infty)$. We begin with by studying functions that are TP or TN.

Definition 5.10. Let $X, Y \subset \mathbb{R}$ and $K : X \times Y \to \mathbb{R}$ be a function. Given $p \in \mathbb{N}$, we say $K(x,y)$ is a *totally non-negative/totally positive kernel of order* p (denoted TN_p or TP_p) if for any integer $1 \leq n \leq p$ and elements $x_1 < x_2 < \cdots < x_n \in X$ and $y_1 < y_2 < \cdots < y_n \in Y$, we have $\det K(x_j, y_k)_{j,k=1}^n$ is non-negative (positive). Similarly, we say that the kernel $K : X \times Y \to \mathbb{R}$ is *totally non-negative/totally positive* if K is TN_p (or TP_p) for all $p \geq 1$.

Example 5.11. The kernel $K(x,y) = e^{xy}$ is totally positive, with $X = Y = \mathbb{R}$. Indeed, choosing real numbers $x_1 < x_2 < \cdots < x_n$ and $y_1 < y_2 < \cdots < y_n$, the matrix $(e^{x_j y_k})_{j,k=1}^n = ((e^{x_j})^{y_k})_{j,k=1}^n$ is a generalized Vandermonde matrix, hence TP, so its determinant is positive.

We next generalize the Cauchy–Binet formula to TP/TN *kernels*. Let $X, Y, Z \subset \mathbb{R}$ and μ be a non-negative Borel measure on Y. Let $K(x,y)$ and $L(y,z)$ be "nice" functions (i.e., Borel measurable with respect to Y), and assume the following function is well defined:

$$M : X \times Z \to \mathbb{R}, \qquad M(x,z) := \int_Y K(x,y) L(y,z)\, d\mu(y). \tag{5.12}$$

For example, consider $K(x,y) = e^{xy}$ and $L(y,z) = e^{yz}$. Take $X = Z = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$ and $Y = \{\log(x_1), \log(x_2), \dots, \log(x_n)\}$, such that $0 < x_1 < x_2 < \cdots < x_n$ and $\alpha_1 < \alpha_2 < \cdots < \alpha_n$. Finally, let μ denote the counting measure on Y. Then $M(\alpha_i, \alpha_k) = \sum_{j=1}^n x_j^{\alpha_i} x_j^{\alpha_k}$. So $(M(\alpha_i, \alpha_k))_{i,k=1}^n = V^T V$, where $V = (x_j^{\alpha_k})_{j,k=1}^n$ is a generalized Vandermonde matrix.

In this "discrete" example (i.e., where the support of μ is a discrete set), $\det M$ is shown to be positive using the total positivity of V, V^T and the Cauchy–Binet formula. In fact, this phenomenon extends to the more general setting above as follows:

Exercise 5.13 (Pólya–Szegő, Basic Composition formula, or Generalized Cauchy–Binet formula)**.** Suppose $X, Y, Z \subset \mathbb{R}$ and $K(x,y), L(y,z), M(x,z)$ are as above. Then using an argument similar to the above proof of the Cauchy–Binet formula, show that

$$\begin{aligned} &\det \begin{pmatrix} M(x_1, z_1) & \dots & M(x_1, z_m) \\ \vdots & \ddots & \vdots \\ M(x_m, z_1) & \dots & M(x_m, z_m) \end{pmatrix} \\ &= \int \cdots \int_{y_1 < y_2 < \cdots < y_m \text{ in } Y} \det(K(x_i, y_j))_{i,j=1}^m \cdot \det(L(y_j, z_k))_{j,k=1}^m \prod_{j=1}^m d\mu(y_j). \end{aligned} \tag{5.14}$$

Remark 5.15. In the right-hand side of Equation (5.14), we may also integrate over the region $y_1 \leq \cdots \leq y_m$ in Y, since matrices with equal rows or columns are singular.

6. Hankel moment matrices are TP. Andréief's identity. Density of TP in TN.

6.1. **Total positivity of H_μ for 'most' measures; Andréief's identity.** Continuing from the previous section with the generalized Cauchy–Binet formula of Pólya–Szegő, from (5.14) and Remark 5.8 we obtain the following consequence:

Corollary 6.1. *(Notation as in the previous section.) If the kernels K and L are both TN_p for some integer $p > 0$ (or even TN), then so is M, where M was defined in* (5.12)*. If instead, K and L are TP_p kernels, where $p \leq |\mathrm{supp}(\mu)|$, then so is M.*

We will apply this result to the moment matrices H_μ defined in (2.21). We begin more generally: suppose $Y \subset \mathbb{R}$ and $u : Y \to (0,\infty)$ is a positive and strictly increasing function, all of whose moments exist with respect to some non-negative measure μ:

$$\int_Y u(y)^n \ d\mu(y) < \infty, \qquad \forall n \geq 0.$$

Then we claim:

Proposition 6.2. *The kernel $M : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, given by*

$$M(n,m) := \int_Y u(y)^{n+m} \ d\mu(y)$$

is TN as well as $TP_{|Y_+|}$, where $Y_+ := \mathrm{supp}(\mu) \subset Y$ is finite or infinite.

Proof. To show M is $TP_{|Y_+|}$, the first claim is that the kernel $K(n,y) := u(y)^n$ is TP on $\mathbb{R}\times Y$. Indeed, we can rewrite $K(n,y) = e^{n\log(u(y))}$. Now given increasing tuples of elements $n_j \in \mathbb{R}$ and $y_k \in Y$, the matrix $K(n_j, y_k)$ is TP, by the total positivity of e^{xy} (see Example 5.11).

Similarly, $L(y,m) := u(y)^m$ is also TP on $Y \times \mathbb{R}$. The result now follows by Corollary 6.1. That M is TN follows from the same arguments, via Remark 5.15. $\square$

This result implies the total positivity of the Hankel moment matrices (2.21). Indeed, setting $u(y) = y$ on domains $Y \subset [0,\infty)$, we obtain:

Corollary 6.3. *Suppose $Y \subset [0,\infty)$ and μ is a non-negative measure on Y with infinite support. Then the moment matrix H_μ is totally positive (of all orders).*

We now show a result that will be used to provide another proof of the preceding corollary.

Theorem 6.4 (Andréief's identity, 1883)**.** *Suppose $Y \subset \mathbb{R}$ is a bounded interval, $n > 0$ is an integer, and $f_1, f_2, \dots, f_n; g_1, g_2, \dots, g_n : Y \to \mathbb{R}$ are integrable functions with respect to a positive measure μ on Y. Define $\mathbf{y} := (y_1, \dots, y_n)$, and*

$$K(\mathbf{y}) := (f_i(y_j))_{i,j=1}^n, \qquad L(\mathbf{y}) := (g_k(y_j))_{j,k=1}^n, \qquad M' := \left(\int_Y f_i(y) g_k(y) \ d\mu(y) \right)_{i,k=1}^n.$$

Then,

$$\det M' = \frac{1}{n!} \int \cdots \int_{Y^n} \det(K(\mathbf{y})) \det(L(\mathbf{y})) \ \prod_{j=1}^n d\mu(y_j). \tag{6.5}$$

Proof. We compute, beginning with the right-hand side:

$$\int \cdots \int_{Y^n} \det(K(\mathbf{y})) \det(L(\mathbf{y})) \prod_{j=1}^{n} d\mu(y_j)$$

$$= \sum_{\sigma,\tau \in S_n} \operatorname{sgn}(\sigma) \operatorname{sgn}(\tau) \prod_{j=1}^{n} \int_Y f_{\sigma(j)}(y_j) g_{\tau(j)}(y_j)\ d\mu(y_j).$$

Let $\beta = \sigma\tau^{-1}$. Then a change of variables shows that this expression equals

$$\begin{aligned} &= \sum_{\tau,\beta \in S_n} \operatorname{sgn}(\beta) \prod_{j=1}^{n} \int_Y g_{\tau(j)}(y) f_{\beta\tau(j)}(y)\ d\mu(y) \\ &= \sum_{\tau \in S_n} \det \left(\int_Y f_{\tau(i)}(y) g_{\tau(k)}(y)\ d\mu(y) \right)_{i,k=1}^{n} \\ &= n! \det M'. \end{aligned}$$

□

As a special case, let $u : Y \to \mathbb{R}$ be positive and strictly increasing, and set $f_i(y) = u(y)^{n_i}, g_k(y) = u(y)^{m_k}$ for all $1 \leq i, k \leq n$ and increasing sequences of integers $n_1 < n_2 < \cdots$ and $m_1 < m_2 < \cdots$. Then the matrix M' has (i, k) entry $\int_Y u(y)^{n_i + m_k}\ d\mu(y)$. Now using Andréief's identity – and the analysis from earlier in this section – we obtain a second proof of Proposition 6.2.

In particular, specializing to $u(y) = y$ and $Y \subset [0, \infty)$ reproves the total positivity of moment matrices H_μ for measures $\mu \geq 0$ with infinite support in Y. In this case we have $n_j = m_j = j - 1$ for $j = 1, \ldots, n$. The advantage of this proof (over using the generalized Cauchy–Binet formula) is that we can compute $\det M'$ 'explicitly' using Andréief's identity:

$$\begin{aligned} M' &= (s_{i+k-2}(\mu))_{i,k=1}^{n}, \quad s_{i+k-2}(\mu) = \int_Y y^{i-1} y^{k-1}\ d\mu(y), \\ \det M' &= \frac{1}{n!} \int_{Y^n} \prod_{1 \leq r < s \leq n} (y_s - y_r)^2\ d\mu(y_1) d\mu(y_2) \cdots d\mu(y_n). \end{aligned} \tag{6.6}$$

(This uses the Vandermonde determinant identity $\det K(\mathbf{y}) = \det L(\mathbf{y}) = \prod_{1 \leq r < s \leq n} (y_s - y_r)$.)

6.2. **Density of TP matrices in TN matrices.** We will now prove an important *density result* due to Whitney in *J. d'Analyse Math.* (1952). Standard/well-known examples of such results are:

1. Every square real matrix can be approximated by non-singular real matrices.
2. Symmetric non-singular real matrices are dense in symmetric real matrices.
3. $n \times n$ positive definite matrices are dense in $\mathbb{P}_n$.

The goal of this section is to prove the following

Theorem 6.7 (Whitney density). *Given positive integers $m, n \geq p$, the set of TP_p $m \times n$ matrices is (entrywise) dense in the set of TN_p $m \times n$ matrices.*

In order to prove this theorem, we first prove a lemma by Pólya.

Lemma 6.8 (Pólya). *For all $\sigma > 0$, the Gaussian kernel $T_{G_\sigma} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ given by*

$$T_{G_\sigma}(x, y) := e^{-\sigma(x-y)^2}$$

is totally positive.

Note that the function $f(x) = e^{-\sigma x^2}$ is such that the TP kernel $T_{G_\sigma}(x,y)$ can be rewritten as $f(x-y)$. Such (integrable) functions on the real line are known as *Pólya frequency (PF) functions*, and we will study these functions – and their preservers – in detail in Sections 37 and 38.

Proof. Given real numbers $x_1 < x_2 < \cdots < x_n$ and $y_1 < y_2 < \cdots < y_n$, we have:

$$\begin{aligned}(T_{G_\sigma}(x_j,y_k))_{j,k=1}^n &= (e^{-\sigma x_j{}^2} e^{2\sigma x_j y_k} e^{-\sigma y_k{}^2})_{j,k=1}^n \\ &= \operatorname{diag}(e^{-\sigma x_j{}^2})_{j=1}^n \begin{pmatrix} e^{2\sigma x_1 y_1} & \dots & e^{2\sigma x_1 y_n} \\ \vdots & \ddots & \vdots \\ e^{2\sigma x_n y_1} & \dots & e^{2\sigma x_n y_n} \end{pmatrix} \operatorname{diag}(e^{-\sigma y_k{}^2})_{k=1}^n,\end{aligned}$$

and this has positive determinant by the previous section (see Example 5.11). $\square$

In a similar vein, we have the following:

Lemma 6.9. *For all $\sigma > 0$, the kernel $H'_\sigma : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, given by*

$$H'_\sigma(x,y) := e^{\sigma(x+y)^2}$$

is totally positive. In particular, the kernels T_{G_σ} (from Lemma 6.8) and H'_σ provide examples of TP Toeplitz and Hankel matrices, respectively.

Proof. The proof of the total positivity of H'_σ is similar to that of T_{G_σ} above, and hence left as an exercise. To obtain TP Toeplitz and Hankel matrices, akin to Example 3.23, we choose any arithmetic progression $x_1, \dots, x_n$ of finite length, and consider the matrices with (j,k)th entry $T_{G_\sigma}(x_j,x_k)$ and $H'_\sigma(x_j,x_k)$, respectively. $\square$

Now we come to the main proof of this section.

Proof of Theorem 6.7. Let $A_{m\times n}$ be TN_p of rank r. Define for each integer $m > 0$ the matrix

$$\begin{aligned}(G_{\sigma,m})_{m\times m} &= (e^{-\sigma(j-k)^2})_{j,k=1}^m, \\ A(\sigma) &= G_{\sigma,m} A G_{\sigma,n}.\end{aligned} \tag{6.10}$$

Note that $G_{\sigma,m}$ is TP by Lemma 6.8, and $G_{\sigma,m} \to \mathrm{Id}_{m\times m}$ as $\sigma \to \infty$. Now as the product of totally non-negative matrices is totally non-negative (Proposition 5.4), and $G_{\sigma,m}$ is non-singular for all m, we have that $A(\sigma)$ is TN_p of rank r.

Claim 6.11. $A(\sigma)$ is $TP_{\min(r,p)}$.

Proof. For any $s \leq \min(r,p)$, let $J \subset [m]$, $K \subset [n]$ of size s. Using the Cauchy–Binet Formula, we compute:

$$\det A(\sigma)_{J\times K} = \sum_{\substack{L\subset[m],\ M\subset[n] \\ |L|=|M|=s}} \det(G_{\sigma,m})_{J\times L} \det A_{L\times M} \det(G_{\sigma,n})_{M\times K}.$$

Now note that for all $1 \leq k \leq \min(r,p)$, at least one $k\times k$ minor of A is positive, and all other $k\times k$ minors are non-negative. Combined with the total positivity of $G_{\sigma,m}$ and $G_{\sigma,n}$, this shows that $\det A(\sigma)_{J\times K} > 0$. This concludes the proof. $\square$

Returning to the proof of the theorem, if $r \geq p$ then the TP matrices $A(\sigma)$ approximate A as $\sigma \to \infty$; thus the proof is complete.

For the remainder of the proof, assume that A and $A(\sigma)$ both have rank $r < p$. Define

$$A^{(1)} := A(\sigma) + \frac{1}{\sigma} E_{11},$$

where E_{11} is the elementary $m \times n$ matrix with $(1,1)$ entry 1, and all other entries 0.

Claim 6.12. $A^{(1)}$ is TN_p of rank $r+1$.

Proof. Fix an integer $1 \leq s \leq p$ and subsets $J \subset [m], K \subset [n]$ of size s. Now consider the $s \times s$ submatrix $A^{(1)}_{J \times K}$. If $1 \notin J$ or $1 \notin K$, then we have: $\det A^{(1)}_{J \times K} = \det A(\sigma)_{J \times K} \geq 0$, whereas if $1 \in J \cap K$, then expanding along the first row or column shows that $A^{(1)}$ is TN_p:

$$\det A^{(1)}_{J \times K} = \det A(\sigma)_{J \times K} + \frac{1}{\sigma} \det(A(\sigma))_{J \setminus \{1\} \times K \setminus \{1\}} \geq 0.$$

As $A, A(\sigma)$ have rank r, and we are changing only one entry, all the $(r+2) \times (r+2)$ minors of $A^{(1)}$ have determinant 0. Now an easy computation yields:

$$\det A^{(1)}_{[r+1] \times [r+1]} = \det A(\sigma)_{[r+1] \times [r+1]} + \frac{1}{\sigma} \det A(\sigma)_{[r+1] \setminus \{1\} \times [r+1] \setminus \{1\}} > 0,$$

where the last inequality occurs because $A(\sigma)$ is TP_r (shown above) and $\det A(\sigma)_{[r+1] \times [r+1]} = 0$. Thus the rank of $A^{(1)}$ is $r+1$. $\square$

Returning to the proof: note that $A^{(1)}$ also converges to A as $\sigma \to \infty$. Inductively repeating this procedure, after $(p-r)$ iterations we obtain a matrix $A^{(p-r)}$, via the procedure

$$A^{(k)}(\sigma) := G_{\sigma,m} A^{(k)} G_{\sigma,n}, \qquad A^{(k+1)} := A^{(k)}(\sigma) + \frac{1}{\sigma} E_{11}. \tag{6.13}$$

Moreover, $A^{(p-r)}$ is a TN_p matrix with rank p. As $\min(r,p) = p$ for this matrix, it follows that $A^{(p-r)}(\sigma)$ is TP_p with $A^{(p-r)}(\sigma) \to A^{(p-r-1)}(\sigma) \to \cdots \to A$ as $\sigma \to \infty$. Thus, A can be approximated by TP_p matrices, and the proof is complete. $\square$

We now make some observations that further Whitney's theorem. First, this density phenomenon also holds upon restricting to symmetric matrices:

Proposition 6.14. *Given positive integers $n \geq p$, the set of symmetric TP_p $n \times n$ matrices is (entrywise) dense in the set of symmetric TN_p $n \times n$ matrices.*

Proof. The proof of Theorem 6.7 goes through verbatim; at each step, the resulting matrix is symmetric. $\square$

Second, a careful analysis of the above proof further shows that

$$A(\sigma)_{jk} = \sum_{l,m} e^{-\sigma(j-l)^2} a_{lm} e^{-\sigma(m-k)^2} \geq a_{jk}.$$

Thus, given $A_{m \times n}$ that is TN_p (possibly symmetric) and $\epsilon > 0$, there exists $B_{m \times n}$ that is TP_p (possibly symmetric) such that $0 \leq b_{jk} - a_{jk} < \epsilon$ for all j,k.

Remark 6.15. In fact we can further refine this: by working with $B(\sigma)$ for such B, we can insist on $0 < b_{jk} - a_{jk} < \epsilon$. From this it follows that given a (possibly symmetric) TN_p matrix $A_{m \times n}$, there exists a sequence B_l of $m \times n$ matrices, all of them TP_p (and symmetric if A is), such that $B_l \to A$ entrywise as $l \to \infty$, and moreover for all j,k,

$$(B_1)_{jk} > (B_2)_{jk} > \cdots > (B_l)_{jk} > \cdots > a_{jk}.$$

7. (Non-)Symmetric TP completion problems.

The main question in matrix completion problems is as follows. Given a partially filled matrix (that is, a partial matrix), do there exist choices for the 'missing' entries such that the resulting matrix has specified properties?

For example: can $\begin{pmatrix} 1 & 0 & ? \\ 2 & ? & ? \\ ? & ? & ? \end{pmatrix}$ be completed to a Toeplitz matrix? Yes: $\begin{pmatrix} 1 & 0 & a \\ 2 & 1 & 0 \\ b & 2 & 1 \end{pmatrix}$, for arbitrary a, b. Similarly, the above partial matrix can be completed to a non-singular, singular, or totally non-negative matrix. However, it cannot be completed to a Hankel or a symmetric Toeplitz matrix, nor to a positive (semi)definite or totally positive matrix. These are examples of some matrix completion problems. Similarly, one can ask if matrices with specified entries extend to kernels on more general domains.

In this section, we discuss three *TP completion problems*. The first is to understand which 2×2 matrices can 'embed in' (or 'extend to') TP matrices – or even TP kernels:

Theorem 7.1 (2×2 TP kernel completions)**.** *Suppose $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbb{R}^{2\times 2}$. The following are equivalent:*

1. *For any $m, n \geq 2$ and specified pairs of rows and columns $J \subset [m], K \subset [n]$ respectively, the matrix A can be completed to a multiple of a generalized Vandermonde (and hence TP) matrix $\widetilde{A}_{m\times n}$ such that $(\widetilde{A})_{J\times K} = A$.*
2. *For any totally ordered sets X, Y with sizes $|X|, |Y| \geq 2$ which admit a TP kernel on $X \times Y$, and pairs of indices $(x_1 < x_2)$ in X and $(y_1 < y_2)$ in Y, the matrix A can be completed to a TP kernel K on $X \times Y$, such that $K[(x_1, x_2); (y_1, y_2)] = A$.*
3. *A is TP.*

We will also show in Theorem 7.4 below, a 'symmetric' variant of this equivalence. To show these results, we first need to understand for which totally ordered sets X, Y do there exist TP kernels on $X \times Y$. Note, this is not possible for all X, Y. For instance, if $|Y| > |\mathbb{R}|$ (e.g. Y is the power set of $\mathbb{R}$) and $|X| \geq 2$, then fix $x_1 < x_2$ in X. Now any real kernel on $X \times Y$ cannot be TP or even TP_2, since when restricted to $\{x_1, x_2\} \times Y$, it contains two equal columns by the pigeonhole principle as $|Y| > |\mathbb{R}^2|$.

At the same time, TP_1 kernels can exist on $X \times Y$ for any totally ordered sets X, Y – e.g., the constant kernel $\mathbf{1}_{X\times Y}$.

Thus, we begin by classifying all domains on which TP_2 kernels exist. Interestingly, they always embed in the positive semi-axis:

Lemma 7.2. *Given non-empty totally ordered sets X, Y, the following are equivalent:*

1. *There exists a TP kernel on $X \times Y$.*
2. *There exists a TP_2 kernel on $X \times Y$.*
3. *At least one of X, Y is a singleton, or there exist order-preserving maps from X, Y into $(0, \infty)$.*

The same equivalence holds if $Y = X$ and K is symmetric: $K(x, y) = K(y, x)$ for all $x, y \in X$.

Proof. We prove a chain of cyclic implications. Clearly (1) implies (2). Now suppose (3) holds. If X or Y is a singleton then the kernel $K \equiv \mathbf{1}_{X\times Y}$ proves (1). Otherwise, we identify X, Y with subsets of $(0, \infty)$ as given; now (1) follows by considering the kernels $K_{\pm}(x, y) := \exp \pm (x \pm y)^2$ as in Lemmas 6.8 and 6.9.

Finally, we assume (2) and show (3). Suppose $|X|, |Y| \geq 2$. Fix $x_1 < x_2$ in X; since K is TP_2, it is an easy exercise to show that the 'ratio function'

$$\psi(y) := K(x_2, y)/K(x_1, y), \qquad y \in Y$$

is a strictly increasing function of y. This yields the desired order-preserving injection $\psi : Y \hookrightarrow (0, \infty)$; the same argument works for X, implying (3).

This proof applies verbatim if $Y = X$ and we consider symmetric kernels K. $\square$

Proof of Theorem 7.1. Clearly, (1) or (2) both imply (3); and (1) is a special case of (2), so we will show (3) $\implies$ (2). (In fact with Lemma 7.2 at hand, we first embed X, Y into $\mathbb{R}$ and then extend A to a generalized Vandermonde kernel, thereby proving a stronger statement than (2), which now clearly implies/specializes to (1).) First use Lemma 7.2 to identify X, Y with subsets of $(0, \infty)$; then work with the TP matrix $\begin{pmatrix} \beta a & \beta b \\ \beta c & \beta d \end{pmatrix}$ for some scalar $\beta > 0$. We **claim** there exists $\beta > 0$ such that this matrix is of the form $(x_j^{\alpha_k})_{j,k=1}^2 = (\exp(\alpha_k \cdot \log x_j))_{j,k=1}^2$ – i.e., a generalized Vandermonde matrix – where either $x_1 < x_2, \alpha_1 < \alpha_2$ or $x_1 > x_2, \alpha_1 > \alpha_2$. But the latter case reduces to the former, by using $1/x_j$ and $-\alpha_k$ instead.

Thus, if the claim holds, then we may suppose A embeds in / can be completed to the rescaled Vandermonde kernel $K(x, y) = \beta^{-1} e^{xy}$ (noting that $x^y = e^{y \log x}$). Now consider the two unique (increasing) linear maps $\varphi_X, \varphi_Y : \mathbb{R} \to \mathbb{R}$, which change the 'position' of the chosen rows and columns to the specified positions, either to draw from K an $m \times n$ TP matrix as in (1), or a TP kernel K on $X \times Y$ as in (2): $K(x, y) := \beta^{-1} \exp(\varphi_X(x)\varphi_Y(y))$.

Thus, it remains to show the above claim. For this, we repeatedly appeal to the total positivity of generalized Vandermonde matrices $(x_j^{\alpha_k})$ with x_j and α_k either both increasing or both decreasing. See Theorem 5.1 and Remark 5.3.

Case 1: Suppose three of the four entries a, b, c, d are equal (note that all four cannot be equal). Then up to rescaling, the possible matrices are

$$A_1 = \begin{pmatrix} \lambda & 1 \\ 1 & 1 \end{pmatrix}, \qquad A_2 = \begin{pmatrix} 1 & 1 \\ \mu & 1 \end{pmatrix}, \qquad A_3 = \begin{pmatrix} 1 & \mu \\ 1 & 1 \end{pmatrix}, \qquad A_4 = \begin{pmatrix} 1 & 1 \\ 1 & \lambda \end{pmatrix},$$

where $\lambda > 1 > \mu > 0$. Now $A_1, \dots, A_4$ are of the form $(x_j^{\alpha_k})$ with (respectively)

$$(x_1, x_2, \alpha_1, \alpha_2) = \quad (\lambda, 1, 1, 0), \quad (1, \mu, 1, 0), \quad (\mu, 1, 0, 1), \quad (1, \lambda, 0, 1).$$

For A_1, A_2, choosing any $x_2 > x_3 > \cdots > x_m > 0$ and $0 > \alpha_3 > \cdots > \alpha_n$, we are done. The other two cases are treated similarly.

Case 2: Suppose two entries in a row or column are equal (but three entries are not). Up to rescaling, the possible matrices are

$$A'_1 = \begin{pmatrix} 1 & 1 \\ \gamma & \delta \end{pmatrix}, \qquad A'_2 = \begin{pmatrix} \delta & \gamma \\ 1 & 1 \end{pmatrix}, \qquad A'_3 = \begin{pmatrix} \delta & 1 \\ \gamma & 1 \end{pmatrix}, \qquad A'_4 = \begin{pmatrix} 1 & \gamma \\ 1 & \delta \end{pmatrix}, \qquad \gamma, \delta \neq 1,$$

and $0 < \gamma < \delta$. Now $A'_1, \dots, A'_4$ are generalized Vandermonde matrices $(x_j^{\alpha_k})$ with

$$(x_1, x_2, \alpha_1, \alpha_2) = \quad (1, e, \log\gamma, \log\delta), \quad (e, 1, \log\delta, \log\gamma), \quad (\delta, \gamma, 1, 0), \quad (\gamma, \delta, 0, 1),$$

respectively. The result follows as in the previous case.

Case 3: In all remaining cases, $\{a, d\}$ is disjoint from $\{b, c\}$. Set $\alpha_1 = 1$, and claim that there exist scalars $\beta, x_1, x_2 > 0$ and $\alpha_2 \in \mathbb{R}$ such that

$$\beta \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} x_1 & x_1^{\alpha_2} \\ x_2 & x_2^{\alpha_2} \end{pmatrix}. \tag{7.3}$$

To see why, denote $L := \log(\beta)$ for $\beta > 0$, as well as $A = \log(a)$ (this is used locally only in the claim of this proof), $B = \log(b)$ etc. Now applying log entrywise to both sides of (7.3),

$$\begin{pmatrix} L+A & L+B \\ L+C & L+D \end{pmatrix} = \begin{pmatrix} \log x_1 & \alpha_2 \log x_1 \\ \log x_2 & \alpha_2 \log x_2 \end{pmatrix}.$$

Taking determinants, we obtain:

$$(L+A)(L+D) - (L+B)(L+C) = 0 \quad \implies \quad L = \frac{BC - AD}{(A+D)-(B+C)},$$

where $A + D > B + C$ since $ad > bc$. Now check that $x_1 = e^L a, x_2 = e^L c, \alpha_2 = \frac{L+B}{L+A}$ satisfies the conditions in (7.3). (Note here that by the assumptions on a, b, c, d, the sum $L + A$ is non-zero, as are $L + B, L + C, L + D$ also.) This shows the claim.

To complete the proof, we need to check that $\beta \begin{pmatrix} x_1 & x_1^{\alpha_2} \\ x_2 & x_2^{\alpha_2} \end{pmatrix}$ is a generalized Vandermonde matrix. Since $x_1 \neq x_2$ by choice of a, c, there are two cases. If $x_1 < x_2$, then $a < c$, so

$$(x_1/x_2)^{\alpha_2} = b/d < a/c = x_1/x_2 < 1 \qquad \implies \qquad \alpha_2 > 1.$$

Hence we indeed get a generalized Vandermonde matrix $(x_j^{\alpha_k})_{2\times 2}$ with increasing x_j and increasing α_k. The case when $x_1 > x_2$ is similarly verified. $\square$

The second TP completion result embeds symmetric 2×2 TP matrices into symmetric TP matrices or kernels, again in 'any position'. This result, like Theorem 7.1, is used in a later part of this text to classify total-positivity preservers of kernels on general domains.

Theorem 7.4 (Symmetric 2×2 kernel completions)**.** *Suppose $A = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \in \mathbb{R}^{2\times 2}$. The following are equivalent:*

(1) For any $n \geq 2$ and specified pair $J \subset [n]$ of rows and columns, the matrix A can be completed to a symmetric (in fact Hankel) TP matrix $\widetilde{A}_{n\times n}$ such that $(\widetilde{A})_{J\times J} = A$.

(2) For any totally ordered set X of size at least 2 which admits a TP kernel on $X \times X$, and any pair of indices $(x_1 < x_2)$ in X, the matrix A can be completed to a symmetric TP kernel K on $X \times X$, such that $K[(x_1, x_2); (x_1, x_2)] = A$.

(3) A is TP.

As a special case, consider the assertion (3) $\implies$ (1), where we want to show that A embeds in the leading principal positions. It suffices to embed the matrix $\begin{pmatrix} 1 & b \\ b & c \end{pmatrix}$, where $0 < b < \sqrt{c}$, inside the square matrix

$$\frac{1}{K} V^T V = \frac{1}{K} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_K \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{n-1} & x_2^{n-1} & \cdots & x_K^{n-1} \end{pmatrix} \begin{pmatrix} 1 & x_1 & \cdots & x_1^{n-1} \\ 1 & x_2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_K & \cdots & x_K^{n-1} \end{pmatrix},$$

where $K \geq n$ and $V_{K\times n}$ is (part of) a 'usual' Vandermonde matrix (x_j^{k-1}). In terms of probability, this amounts to finding a uniform random variable, supported on $\{x_1, \dots, x_K\}$, with mean b and variance $c - b^2$. (See e.g. [1].) That is, a discrete inverse moment problem.

In fact this motivates the proof-strategy, even for the stronger result (3) $\implies$ (2), which should similarly involve continuous distributions. Thus, the proof involves continuous random variables, and uses the Generalized Cauchy–Binet formula (one can also use Andréief's identity).

Proof. Clearly (1) or (2) implies (3), and (1) is a special case of (2), so it suffices to show (3) $\implies$ (2). The proof is similar to that of Theorem 7.1: first embed X inside $\mathbb{R}$ via Lemma 7.2. Now it suffices to embed any A as above in a continuous Hankel (hence symmetric) TP kernel $K : \mathbb{R} \times \mathbb{R} \to (0, \infty)$, and then use an increasing linear map $\varphi_X : X \to \mathbb{R}$ to 'change locations'. To construct K, we use Proposition 6.2. Thus, we need an increasing function $u : \mathbb{R} \to (0, \infty)$, scalars $s < t \in \mathbb{R}$, and a positive measure μ on $\mathbb{R}$ such that

$$\int_{\mathbb{R}} u(y)^{2s}\ d\mu(y) = a, \qquad \int_{\mathbb{R}} u(y)^{s+t}\ d\mu(y) = b, \qquad \int_{\mathbb{R}} u(y)^{2t}\ d\mu(y) = c.$$

A solution is as follows, verified via direct computations. Define $B := \log(b/a)$, $C := \log(c/a)$ and note that $(C/2) - B = \frac{1}{2}\log(ac/b^2) > 0$. Now verify that the following works:

$$s := \frac{4B - C}{8\sqrt{(C/2) - B}}, \quad t := \frac{3C - 4B}{8\sqrt{(C/2) - B}}, \quad u(y) := e^{2y}, \quad \mu(y) := \frac{a}{\sqrt{\pi} e^{4s^2}}\ e^{-y^2}\ dy. \quad \square$$

The third TP completion problem in this section extends Theorem 7.1 differently, to completions of matrices of arbitrary sizes that are totally positive of arbitrary order:

Theorem 7.5. *Suppose $m, n \geq 1$ and $1 \leq p \leq \min(m, n)$ are integers, and $J \subset [m]$, $K \subset [n]$ are 'sub-intervals' containing m', n' consecutive integers, respectively. A real $m' \times n'$ matrix A' can be completed to a TP_p real $m \times n$ matrix, in positions $J \times K$, if and only if A' is TP_p.*

Proof. One implication is obvious. Conversely, suppose $A'_{m' \times n'}$ is TP_p. It suffices to show that one can add an extra row either above or below A' and obtain a TP_p matrix. Then the result follows by induction and taking transposes.

We first show how to add a row $(a_1, \dots, a_{n'})$ at the bottom. Choose any $a_1 > 0$; having defined $a_1, \dots, a_k > 0$ for some $0 < k < n'$, we inductively define a_{k+1} as follows. Define the $(m'+1) \times (k+1)$ matrix $B_k := \begin{pmatrix} A'_{[m'] \times [k+1]} \\ a_1 \ \cdots \ a_{k+1} \end{pmatrix}$ (with unknown $a_{k+1} > 0$), and consider the 1×1, $\dots$, $p \times p$ submatrices B' of B_k which contain entries from the last row and last column, whence a_{k+1}. Compute $\det(B')$ by expanding along the last row, and from right to left. Requiring $\det(B') > 0$ yields a strict lower bound for a_{k+1}, since the cofactor corresponding to a_{k+1} is a lower-order minor of A', hence positive. Working over all such minors $\det(B')$ yields a finite set of lower bounds, so that it is possible to define a_{k+1} and obtain all minors with 'bottom corner' a_{k+1} and size at most $p \times p$ to be positive. By the induction hypothesis, all other minors of B_k of size at most $p \times p$ are positive, so B_k is TP_p. Proceeding inductively, we obtain the desired $(m'+1) \times n'$ completion of A' that is TP_p.

The argument is similar to add a row $(a_1, \dots, a_{n'})$ on top of A': this time we proceed sequentially from right to left. First, $a_{n'}$ is arbitrary; then to define a_k (given $a_{k+1}, \dots, a_{n'}$), we require a_k to satisfy a finite set of inequalities (obtained by expanding $\det(B')$ along the first row from left to right), and each inequality is again a strict lower bound. $\square$

Remark 7.6. To conclude, we explain how Whitney density and the above TP completion problems are used in a later part of the text, in classifying the preservers $F \circ -$ of TP matrices and kernels on arbitrary domains $X \times Y$. The first step will be to deduce from Theorem 7.1 that any 2×2 TP matrix can be embedded in a TP kernel on $X \times Y$. This will help show that the preserver F must be continuous. We then use results akin to Whitney's density theorem 6.7 to show that F preserves TN kernels on $X \times Y$. These latter will turn out to be easier to classify. Similarly for the preservers of symmetric TP kernels on $X \times X$.

8. Theorems of Perron and Kronecker. Spectra of TP and TN matrices.

We conclude this part by studying the spectra of TP/TN matrices. For real square symmetric matrices $A_{n\times n}$, recall Sylvester's criterion 2.8, which says modulo Theorem 2.5 that *such a matrix $A_{n\times n}$ has all principal minors non-negative (or positive), if and only if all eigenvalues of A are non-negative (positive).*

The goal in this section is to show a similar result for TP/TN matrices. More precisely, we will show the same statement as above, removing the words 'symmetric' and 'principal' from the preceding paragraph. In particular, not only are all minors of TP and TN matrices positive and non-negative respectively, but moreover, so are their eigenvalues. A slightly more involved formulation of this result is:

Theorem 8.1. *Given integers $m, n \geq p \geq 1$ and $A \in \mathbb{R}^{m\times n}$, the following are equivalent:*

1. *For every square submatrix B of A of size $\leq p$, we have $\det(B)$ is non-negative (respectively positive). In other words, A is TN_p (respectively TP_p).*
2. *For every square submatrix B of A of size $\leq p$, the eigenvalues of B are non-negative (respectively positive and simple).*

Note that the analogous statement for positive semidefinite matrices clearly holds (as mentioned above), by Sylvester's theorem.

We will follow the original proof, written out by Gantmacher and Krein in their 1937 paper in *Compositio Math.* This approach is also found in Chapter XIII.9 of F.R. Gantmacher's book *The theory of matrices*; and in an expository account by A. Pinkus [279] found in the conference proceedings *Total positivity and its applications* (of the 1994 Jaca meeting in honor of Sam Karlin), edited by M. Gasca and C.A. Micchelli. (The above paper of Pinkus also features at the end of this section, when we discuss spectra of TN kernels.)

This approach relies on two well-known theorems, which are interesting in their own right. The first was shown by O. Perron in his 1907 paper in *Math. Ann.*:

Theorem 8.2 (Perron)**.** *Let $A_{n\times n}$ be a square, real matrix with all positive elements. Then A has a simple, positive eigenvalue λ with an eigenvector $u_0 \in \mathbb{R}^n$, such that:*

(a) For the remaining $n-1$ eigenvalues $\mu \in \mathbb{C}$, we have $|\mu| < \lambda$.
(b) The coordinates of u_0 are all non-zero and of the same sign.

This result has been studied and extended by many authors in the literature; notably, the *Perron–Frobenius theorem* is a key tool used in one of the approaches to studying discrete time Markov chains over finite state-space. As these extensions are not central to the present discussion, we do not pursue them further.

Proof. Write $v \geq u$ (or $v > u$) for $u, v \in \mathbb{R}^n$ to denote the (strict) coordinatewise ordering: $v_j \geq u_j$ (or $v_j > u_j$) for all $1 \leq j \leq n$. A first observation, used below, is:

$$u \leq v \text{ in } \mathbb{R}^n,\ u \neq v \qquad \Longrightarrow \qquad Au < Av. \tag{8.3}$$

We now proceed to the proof. Define

$$\lambda := \sup\{\mu \in \mathbb{R} : Au \geq \mu u \text{ for some non-zero vector } 0 \leq u \in \mathbb{R}^n\}.$$

Now verify that

$$0 < n \min_{j,k} a_{jk} \leq \lambda \leq n \max_{j,k} a_{jk};$$

in particular, λ is well-defined. Now for each $k \geq 1$, there exist (rescaled) vectors $u_k \geq 0$ in $\mathbb{R}^n$ whose coordinates sum to 1, and such that $Au_k \geq (\lambda - 1/k)u_k$. But then the u_k belong to a compact simplex S, so there exists a subsequence converging to some vector $u_0 \in S$. It follows that $Au_0 \geq \lambda u_0$; if $Au_0 \neq \lambda u_0$, then an application of (8.3) leads to a contradiction to the maximality of λ. Thus $Au_0 = \lambda u_0$ for non-zero $u_0 \geq 0$. But then $Au_0 > 0$, hence $u_0 = \lambda^{-1} A u_0$ has all positive coordinates. This proves part (b).

It remains to show part (a) and the simplicity of λ. First if $Av = \mu v$ for any eigenvalue μ of A (and $v \neq 0$), then defining $|v| := (|v_1|, \dots, |v_n|)^T$, we deduce:

$$A|v| \geq |Av| = |\mu v| = |\mu||v| \qquad \Longrightarrow \qquad \lambda \geq |\mu|.$$

Suppose for the moment that $|\mu| = \lambda$. Then $A|v| = \lambda|v|$, else (as above) an application of (8.3) leads to a contradiction to the maximality of λ. But then $A|v| = |Av|$ from the preceding computation. By the triangle inequality over $\mathbb{C}$, this shows all coordinates of v have the same argument, which we can take to be $e^{i0} = 1$ by normalizing v. It follows that $Av = \lambda v$ from above, since now $v = |v|$. Hence $\mu = \lambda$.

Thus we have shown that if $Av = \mu v$ for $|\mu| = \lambda$ (and $v \neq 0$), then $\mu = \lambda$ and we may rescale to get $v \geq 0$. In particular, this shows part (a) modulo the simplicity of the eigenvalue λ. Moreover, if u_0, u_0' are linearly independent λ-eigenvectors for A, then one can come up with a linear combination $v \in \mathbb{R}u_0 + \mathbb{R}u_0'$ with at least one positive and one negative coordinate. This contradicts the previous paragraph, so it follows that λ has geometric multiplicity one.

The final remaining task is to show that λ is a simple eigenvalue of A. If not, then by the preceding paragraph there exists $u_1 \notin \mathbb{R}u_0$ such that ($Au_0 = \lambda u_0$ and) $Au_1 = \lambda u_1 + \mu u_0$ for some non-zero scalar μ. Now since A^T has the same eigenvalues as A, the above analysis there exists $v_0 \in \mathbb{R}^n$ such that $v_0^T A = \lambda v_0^T$. Hence:

$$\lambda v_0^T u_1 = v_0^T A u_1 = v_0^T (\lambda u_1 + \mu u_0).$$

But then $\mu \cdot v_0^T u_0 = 0$, which is impossible since $\mu \neq 0$ and $u_0, v_0 > 0$. This shows that λ is simple, and concludes the proof. $\square$

The second result we require is folklore: Kronecker's theorem on compound matrices. We begin by introducing this family of auxiliary matrices, associated to each given matrix.

Definition 8.4. Fix a matrix $A_{m \times n}$ (which we take to be real, but the entries can lie in any unital commutative ring), and an integer $1 \leq r \leq \min(m, n)$.

(1) Let $S_1, \dots, S_{\binom{m}{r}}$ denote the r-element subsets of $[m] = \{1, \dots, m\}$, ordered lexicographically. (Thus $S_1 = \{1, \dots, r\}$ and $S_{\binom{m}{r}} = \{m - r + 1, \dots, m\}$.) Similarly, let $T_1, \dots, T_{\binom{n}{r}}$ denote the r-element subsets of $[n]$ in lexicographic order.

Now define the *rth compound matrix of A* to be a matrix $C_r(A)$ of dimension $\binom{m}{r} \times \binom{n}{r}$, whose (j, k)th entry is the minor $\det(A_{S_j \times T_k})$.

(2) For $r = 0$, define $C_0(A) := \mathrm{Id}_{1 \times 1}$.

We now collect together some basic properties of compound matrices:

Lemma 8.5. *Suppose $m, n \geq 1$ and $0 \leq r \leq \min(m, n)$ are integers, and $A_{m \times n}$ a matrix.*

(1) Then $C_1(A) = A$, and $C_r(cA) = c^r C_r(A)$ for all scalars c.
(2) $C_r(A^T) = C_r(A)^T$.
(3) $C_r(\mathrm{Id}_{n \times n}) = \mathrm{Id}_{\binom{n}{r} \times \binom{n}{r}}$.
(4) The Cauchy–Binet formula essentially says:

$$C_r(AB) = C_r(A) C_r(B), \qquad \forall A \in \mathbb{R}^{m \times n},\ B \in \mathbb{R}^{n \times p},\ p \geq 1.$$

(5) *As a consequence,* $\det(C_r(AB)) = \det(C_r(A))\det(C_r(B))$ *when* $m = n = p$ *(i.e.,* A, B *are square).*
(6) *As another consequence of the multiplicativity of* C_r*, if* A *has rank* $0 \le r \le \min(m,n)$*, then* $C_j(A)$ *has rank* $\binom{r}{j}$ *for* $j = 0, 1;\ r, r+1, \dots, \min(m,n)$*.*
(7) *If* A *is square, then* $C_n(A) = \det(A)$*; if* A *is moreover invertible, then* $C_r(A)^{-1} = C_r(A^{-1})$*.*
(8) *If* A *is upper/lower triangular, diagonal, symmetric, orthogonal, or normal, then* $C_r(A)$ *has the same property.*

Proof. We only sketch a couple of the proofs, and leave the others as exercises. If A has rank r, then one can write $A = M_{m\times r} N_{r\times n}$, where the columns of M are linearly independent, as are the rows of N. But then $C_r(A)$ is the product of a non-zero column vector $C_r(M)$ and a non-zero row vector $C_r(N)$, hence has rank 1. (Here we require the underlying ground ring to be an integral domain.)

The other case we consider here is when A is ($n \times n$ and) upper triangular. In this case let $J = \{j_1 < \cdots < j_r\}$ and $K = \{k_1 < \cdots < k_r\}$ be subsets of $[n]$, with $J > K$ in the lexicographic order. Hence there exists a unique $l \in [1, r]$ such that

$$j_1 = k_1, \quad \cdots, \quad j_{l-1} = k_{l-1}, \quad k_l < j_l < j_{l+1} < \cdots < j_r.$$

It follows that $A_{J\times K}$ is a block triangular matrix of the form $\begin{pmatrix} C_{(l-1)\times(l-1)} & D \\ 0 & E_{(r-l+1)\times(r-l+1)} \end{pmatrix}$, and that the leftmost column of E is the zero vector. Hence $\det(A_{J\times K}) = 0$ if $J > K$. $\square$

With Lemma 8.5 in hand, one can state and prove

Theorem 8.6 (Kronecker)**.** *Let* $n \ge 1$ *and suppose the complex matrix* $A_{n\times n}$ *has the multiset of eigenvalues* $\{\lambda_1, \dots, \lambda_n\}$*. For all* $0 \le r \le n$*, the* $\binom{n}{r}$ *eigenvalues of* $C_r(A)$ *are precisely of the form* $\prod_{j\in S} \lambda_j$*, where* S *runs over all* r*-element subsets of* $[n]$*.*

In words, the eigenvalues of $C_r(A)$ are the $\binom{n}{r}$ products of r distinct eigenvalues of A.

Proof. Let J denote an (upper triangular) Jordan canonical form of A. That is, there exists an invertible matrix M satisfying: $MJM^{-1} = A$, with the diagonal entries of J given by $\lambda_1, \dots, \lambda_n$. Applying various parts of Lemma 8.5,

$$C_r(A) = C_r(M)\, C_r(J)\, C_r(M)^{-1},$$

with $C_r(J)$ upper triangular. Thus the eigenvalues of $C_r(A)$ are precisely the diagonal entries of $C_r(J)$, and these are precisely the claimed set of scalars. $\square$

These ingredients help show that TP square matrices have simple, positive eigenvalues:

Proof of Theorem 8.1. Clearly, (2) implies (1). Conversely, first note that by focussing on a fixed square submatrix B and all of its minors, the implication (1) $\implies$ (2) for general $m, n \ge p$ reduces to the special case $m = p = n$, which we assume henceforth.

First suppose $A_{n\times n}$ is TP. Relabel its eigenvalues $\lambda_1, \dots, \lambda_n$ such that $|\lambda_1| \ge \cdots \ge |\lambda_n|$. Now let $1 \le r \le n$; then the compound matrix $C_r(A)$ has positive entries, so by Perron's theorem 8.2, there exists a unique largest positive eigenvalue $\lambda_{\max,r}$, and all others are smaller in modulus. Hence by Kronecker's theorem 8.6, $\lambda_{\max,r} = \lambda_1 \cdots \lambda_r$, and we have

$$\lambda_1 \cdots \lambda_r > 0, \qquad \forall 1 \le r \le n.$$

It follows that each λ_j is (real and) positive. Moreover, from Perron and Kronecker's results it also follows for each $1 \leq r \leq n-1$ that

$$\lambda_1 \cdots \lambda_r > \lambda_1 \cdots \lambda_{r-1} \cdot \lambda_{r+1},$$

and so $\lambda_r > \lambda_{r+1}$, as desired.

This shows the result for TP matrices. Now suppose $A_{n\times n}$ is TN. By Whitney's density theorem 6.7, we may approximate A by a sequence A_k of TP matrices. Hence the characteristic polynomials converge: $p_{A_k}(t) := \det(t \operatorname{Id}_{n\times n} - A_k) \to p_A(t)$ coefficientwise, as $k \to \infty$. Since $\deg(p_{A_k}) = n$ for all $k \geq 1$, it follows by the 'continuity of roots' – proved below – that the eigenvalues of p_A also avoid the open set $\mathbb{C} \setminus [0, \infty)$. This concludes the proof. $\square$

Thus, it remains to show that the roots of a real or complex polynomial are continuous functions of its coefficients. This is in fact a consequence of Hurwitz's theorem in complex analysis, but we restrict ourselves here to mentioning a simpler result. We provide two proofs, which can both be found online or in books.

Proposition 8.7. *Suppose $p_k \in \mathbb{C}[t]$ is a sequence of polynomials, with $\deg(p_k)$ uniformly bounded over all $k \geq 1$. If $U \subset \mathbb{C}$ is an open set on which no p_k vanishes, and $p_k(t) \to p(t)$ coefficientwise, then either $p \equiv 0$ on U, or p is nonvanishing on U.*

Proof. We restrict ourselves to outlining this argument, as this direction is not our main focus. Suppose $p|_U$ is not identically zero, and $p(w) = 0$ for some $w \in U$. Choose $\delta > 0$ such that the closed disk $\overline{D} := \overline{D(w, \delta)} \subset U$ and $p(t)$ has no roots in $\overline{D} \setminus \{w\}$. Then each p_k is uniformly continuous on the compact boundary $\partial D = \overline{D} \setminus D$, where $D = D(w, \delta)$ is the open disk. For sufficiently large k, $\deg(p_k) = \deg(p) \geq 0$ by the hypotheses. This is used to show that the p_k converge uniformly on ∂D to p, and similarly, $p_k' \to p'$ uniformly on ∂D.

Since p is nonvanishing on ∂D, we have

$$m := \min_{z \in \partial D} |p(z)| > 0,$$

and hence for sufficiently large k, we have

$$\min_{z \in \partial D} |p_k(z)| > \frac{m}{2}, \qquad \forall k \gg 0.$$

Using this, one shows that the sequence $\{p_k'/p_k : k > 0\}$ converges uniformly on ∂D to p'/p.

Now integrate on ∂D: since p_k'/p_k equals $\sum_j 1/(z - \lambda_j(p_k))$ where one sums over the multiset of roots λ_j of each p_k, and since p_k does not vanish in $U \supset \overline{D}$, we have

$$0 = \oint_{\partial D} \frac{p_k'(z)}{p_k(z)}\, dz \to \oint_{\partial D} \frac{p'(z)}{p(z)}\, dz.$$

Hence the right-hand integral vanishes. On the other hand, that same integral equals a positive integer – namely, the multiplicity of the root w of $p(t)$. This yields the desired contradiction, and so p does not vanish on U. $\square$

Alternate proof. This proof is even simpler, and does not use complex analysis – but assuming that the polynomials are all monic. In this case, the hypotheses imply that $\deg p_k$ is independent of k, say equal to $m \geq 0$. (This indeed subsumes the case when each p_k is the characteristic polynomial of an $m \times m$ complex matrix A_k, and $A_k \to A$ entrywise.) Now if $m = 0$ or $m = 1$ then the result is immediate – e.g. for $m = 1$, if $p_k(x) = x - \alpha_k$, then $\alpha_k \in \mathbb{C} \setminus U$. But then $\lim_k \alpha_k$ also lies outside U, as desired.

Henceforth, we suppose $m \geq 2$. The first claim is that every root α of a complex polynomial

$$a_0 + a_1 z + \cdots + a_{m-1} z^{m-1} + z^m$$

satisfies: $|\alpha| < 1 + m \max_{j<m} |a_j|$. Indeed, if $a_0 = \cdots = a_{m-1} = 0$, then we have $\alpha^m = 0$, so $\alpha = 0$. Else if $\max_{j<m} |a_j| > 0$, then there are two cases. If $|\alpha| \leq 1$ then the result is immediate. Otherwise $|\alpha| > 1$, and we compute via the triangle inequality in $\mathbb{C}$:

$$1 = \frac{|\alpha^m|}{|\alpha^m|} = \left| \sum_{j=0}^{m-1} a_j \alpha^{j-m} \right| \leq \sum_{j=0}^{m-1} \max_{j<m} |a_j| \cdot |\alpha|^{j-m} < \max_{j<m} |a_j| \cdot m |\alpha|^{-1}.$$

This shows $|\alpha| < m \max_{j<m} |a_j|$, proving the claim in the case $|\alpha| > 1$.

We now come to the proof of the continuity of roots (for the case $m \geq 2$). Thus, here p_k is monic of degree $m \geq 2$, for all $k \geq 1$. Denote the roots of p_k by $\alpha_k^{(1)}, \dots, \alpha_k^{(m)}$, for each $k \geq 1$. Since $p_k \to p$ coefficientwise, the coefficients of all p_k, p are uniformly bounded, say in $D(0, M)$ for some $M > 0$. By the claim above, it follows that all roots $\alpha_k^{(j)}$ lie in a bounded disk:

$$|\alpha_k^{(j)}| < 1 + mM, \qquad \forall 1 \leq j \leq m, \ k \geq 1.$$

Thus, there exists a subsequence $k_l, \ l \geq 1$ and a tuple $(\alpha^{(1)}, \dots, \alpha^{(m)}) \in \mathbb{C}^m$ such that

$$\lim_{l \to \infty} \alpha_{k_l}^{(j)} = \alpha^{(j)}, \qquad \forall 1 \leq j \leq m.$$

But then,

$$p(z) = \lim_{l \to \infty} p_{k_l}(z) = \lim_{l \to \infty} \prod_{j=1}^{m} (z - \alpha_{k_l}^{(j)}) = \prod_{j=1}^{m} (z - \alpha^{(j)}).$$

Now since $\alpha_{k_l}^{(j)} \notin U$ for all j, l, the same holds for all $\alpha^{(j)}$, and the result follows. $\square$

Remark 8.8 (Spectra of continuous TN kernels)**.** We conclude with some remarks on the spectral properties of totally non-negative *kernels*. These were studied even before the 1937 paper of Gantmacher–Krein on the spectra of TN matrices; for a detailed historical account with complete proofs, see the article by Pinkus [279] in the compilation [141]. As Pinkus mentions, the case of symmetric kernels was studied by Kellogg in his 1918 paper [211] in *Amer. J. Math.* The non-symmetric case was studied in 1936 by Gantmacher [135], following prior work by the 1909 work [328] of Schur in *Math. Ann.* and the 1912 work [195] of Jentzsch in *J. reine angew. Math.*

Bibliographic notes and references

Most of the material in the first four sections is standard and can be found in other textbooks on matrix theory; see e.g. Bapat and Raghavan [18], Bhatia [45, 46], Fallat and Johnson [112], Gantmacher [136], Hiai and Petz [172], Horn and Johnson [183, 184], Horn and Johnson [183, 184], Karlin [200], Pinkus [282], Zhan [376], and Zhang [378].

About the rest: the matrix factorization in (2.33) involving Schur complements was observed by Schur in [331]. Theorems 2.32 and 2.38, on the positivity of a block-matrix in terms of Schur complements, were shown by Albert [9]. Remark 3.8 on applications of Schur products to other areas is taken from discussions in the books [183, 184]. More broadly, a discussion of the legacy of Schur's contributions in analysis can be found in the comprehensive survey [105].

The Schur product theorem 3.12 was shown by Schur [329] (the proof involving Kronecker products is by Marcus and Khan [253]), and its non-zero lower bound, Theorem 3.17, is by Khare [213] (following a prior bound by Vybíral [356]). Remark 3.18 is by Vybíral [356]; and the previous non-zero lower bounds on the Schur product in (3.16) are from [120] and [296]. For more on the Hamburger and Stieltjes moment problems (see Remarks 2.23 and 4.4, respectively), see the monographs [8, 310, 335].

The notion of TN and TP matrices and kernels was introduced by Schoenberg in [311], where he showed that TN matrices satisfy the variation diminishing property. (Schoenberg then proved in [312] the Budan–Fourier theorem 10.6 using TN matrices.) The characterization in Theorem 3.22 of this property is from Motzkin's thesis [263]. (Instances of total non-negativity and of variation diminution had appeared in earlier works, e.g., by Fekete [118], Hurwitz [188], Laguerre [229], and others.) Theorem 4.1, relating positivity and total non-negativity for a Hankel matrix, appears first in [282] for TP Hankel matrices, then in detail in [113] for the TN, TP_p, and TN_p variants. (Neither of these works uses contiguous minors, which have the advantage of only needing to work with Hankel submatrices.) The lemmas used in the proof above are given in [136, 137], and the result of Fekete and its extension by Schoenberg are in [118] and [325], respectively. Corollary 4.3 – on the total non-negativity of moment matrices of measures on $[0, \infty)$ – is the easy half of the Stieltjes moment problem, and was also proved differently, by Heiligers [163].

Theorem 5.1 on the total positivity of generalized Vandermonde matrices is found in [136]. The "weak" Descartes' rule of signs (Lemma 5.2) was first shown by Descartes in 1637 [100] for polynomials; the proof given in this text via Rolle's theorem is by Laguerre in 1883 [229] and holds equally well for the extension to real powers. The Basic Composition formula (5.14) can be found in the book by Pólya and Szegő [289] (see also Karlin [200]), while Andréief's identity is from [13]. The subsequent observations on the total positivity of "most" Hankel moment matrices are taken from [200]. Whitney's density theorem is from [368]. Theorem 7.5 is due to Johnson and Smith [196]; all other results in Section 7, on TP completions of 2×2 matrices to TP matrices/kernels on arbitrary domains, are from Belton–Guillot–Khare–Putinar [32]. Theorem 8.1 on the eigenvalues of TP and TN matrices is due to Gantmacher and Krein [137], and is also found in numerous sources – to list a few, [136, 138], and Pinkus's article [279] in the collection [141]. (The original result was for oscillatory matrices, and immediately follows from Theorem 8.1.) Perron's theorem 8.2 is from [278]. Hurwitz's theorem, or the continuity of zeros shown in Proposition 8.7, can be found in in standard textbooks on complex analysis, see e.g. [88].

Part 2:

Entrywise powers preserving (total) positivity in a fixed dimension

Part 2: Entrywise powers preserving (total) positivity in a fixed dimension

9. Entrywise powers preserving positivity in a fixed dimension: I.

In the rest of this text (except Part 4), we discuss operations that preserve the notions of positivity that have been discussed earlier. Specifically, we will study functions that preserve positive semidefiniteness, or TP/TN, when applied via composition operators on positive kernels. In this part and the next, we deal with kernels on finite domains – aka matrices – which translates to the functions being applied *entrywise* to various classes of matrices. This part of the text discusses the important special case of entrywise *powers* preserving positivity on matrices; to understand some of the modern and classical motivations behind this study, we refer the reader to Sections 13.1 and 16.1 below, respectively.

We begin with some preliminary definitions.

Definition 9.1. Given a subset $I \subset \mathbb{R}$, define $\mathbb{P}_n(I) := \mathbb{P}_n \cap I^{n\times n}$ to be the set of $n \times n$ positive semidefinite matrices, all of whose entries are in I.

A function $f : I \to \mathbb{R}$ acts *entrywise* on vectors/matrices with entries in I via $A = (a_{jk}) \mapsto f[A] := (f(a_{jk}))$. We say f is *Loewner positive on* $\mathbb{P}_n(I)$ if $f[A] \in \mathbb{P}_n$ whenever $A \in \mathbb{P}_n(I)$.

Note that the entrywise operator $f[-]$ differs from the usual holomorphic calculus (except when acting on diagonal matrices by functions that vanish at the origin).

Remark 9.2. The entrywise calculus was initiated by Schur in the same paper [329] in *J. reine angew. Math.* (1911) where he proved the Schur product theorem. Schur defined $f[A]$ – but using different notation – and proved the first result involving entrywise maps; see e.g. [105, Page cxii] for additional commentary.

We fix the following **notation** for future use. If $f(x) = x^\alpha$ for some $\alpha \geq 0$ and $I \subset [0,\infty)$, then we write $A^{\circ\alpha}$ for $f[A]$, where A is any vector or matrix. By convention we shall take $0^0 = 1$ whenever required, so that $A^{\circ 0}$ is the matrix $\mathbf{1}$ of all ones, and this is positive semidefinite whenever A is square.

At this point, one can ask the following question: *Which entrywise power functions preserve positive semidefiniteness, total positivity or total-negativity on $n \times n$ matrices?* (We will also study in Sections 17 and 18, the case of general functions.) The first of these questions was considered by Loewner [242, 243] while studying the Green's function of the unit circle and schlicht/univalent functions (on a separate note, the coefficients of such functions feature in the Bieberbach conjecture). The question – i.e., which entrywise powers preserve positivity – was eventually answered in *J. Math. Anal. Appl.* (1977) by two of Loewner's students, C.H. FitzGerald and R.A. Horn:

Theorem 9.3 (FitzGerald–Horn)**.** *Given an integer $n \geq 2$ and a scalar $\alpha \in \mathbb{R}$, $f(x) = x^\alpha$ preserves positive semidefiniteness on $\mathbb{P}_n((0,\infty))$ if and only if $\alpha \in \mathbb{Z}^{\geq 0} \cup [n-2,\infty)$.*

Remark 9.4. We will in fact show that if α is not in this set, there exists a rank-2 Hankel TN matrix $A_{n\times n}$, such that $A^{\circ\alpha} \notin \mathbb{P}_n$. (In fact, it is the (partial) moment matrix of a non-negative measure on two points.) Also notice that Theorem 9.3 holds for entrywise powers applied to $\mathbb{P}_n([0,\infty))$, since as we show, $\alpha < 0$ never works while $\alpha = 0$ always does work by convention; and for $\alpha > 0$ the power x^α is continuous on $[0,\infty)$, and we use the density of $\mathbb{P}_n((0,\infty))$ in $\mathbb{P}_n([0,\infty))$.

The "phase transition" at $n-2$ in Theorem 9.3 is a remarkable and oft-repeated phenomenon in the entrywise calculus (we will see additional examples of such events in Section 14). The value $n-2$ is called the *critical exponent* for the given problem of preserving positivity.

(Loewner's aforementioned papers [242, 243] are perhaps the first time that the term *critical exponent* was used in this context.)

To prove Theorem 9.3, we require a preliminary lemma, also by FitzGerald and Horn. Recall the preliminaries in Section 2.4.

Lemma 9.5. *Given a matrix $A \in \mathbb{P}_n(\mathbb{R})$ with last column ζ, the matrix $A - a_{nn}^\dagger \zeta\zeta^T$ is positive semidefinite with last row and column zero.*

Here, $a_{nn}^\dagger$ denotes the Moore–Penrose inverse of the 1×1 matrix (a_{nn}).

Proof. If $a_{nn} = 0$, then $\zeta = 0$ by positive semidefiniteness and $a_{nn}^\dagger = 0$ as well. The result follows. Now suppose $a_{nn} > 0$ and write $A = \begin{pmatrix} B & \omega \\ \omega^T & a_{nn} \end{pmatrix}$. Then a straightforward computation shows that

$$A - a_{nn}^{-1}\zeta\zeta^T = \begin{pmatrix} B - \frac{\omega\omega^T}{a_{nn}} & 0 \\ 0 & 0 \end{pmatrix}.$$

Notice that $B - \frac{\omega\omega^T}{a_{nn}}$ is the Schur complement of A with respect to $a_{nn} > 0$. Now since A is positive semidefinite, so is $B - \frac{\omega\omega^T}{a_{nn}}$ by Theorem 2.32. □

Proof of Theorem 9.3. Notice that x^α preserves positivity on $\mathbb{P}_n((0,\infty))$ for all $\alpha \in \mathbb{Z}^{\geq 0}$, by the Schur product theorem 3.12. Now we prove by induction on $n \geq 2$ that if $\alpha \geq n-2$, then x^α preserves positivity on $\mathbb{P}_n((0,\infty))$. If $n = 2$ and $A = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \in \mathbb{P}_2((0,\infty))$, then $ac \geq b^2 \implies (ac)^\alpha \geq b^{2\alpha}$ for all $\alpha \geq 0$. It follows that $A^{\circ\alpha} \in \mathbb{P}_2((0,\infty))$, proving the base case.

For the induction step, assume that the result holds for $n-1 \geq 2$. Suppose $\alpha \geq n-2$ and $A \in \mathbb{P}_n((0,\infty))$; thus, $a_{nn} > 0$. Consider the following elementary definite integral

$$x^\alpha - y^\alpha = \alpha(x-y)\int_0^1 (\lambda x + (1-\lambda)y)^{\alpha-1}\, d\lambda. \tag{9.6}$$

Let ζ denote the final column of A; applying (9.6) entrywise to x, an entry of A, and y, the corresponding entry of $B := \frac{\zeta\zeta^T}{a_{nn}}$ yields

$$A^{\circ\alpha} - B^{\circ\alpha} = \alpha\int_0^1 (A-B)\circ(\lambda A + (1-\lambda)B)^{\circ(\alpha-1)}\, d\lambda. \tag{9.7}$$

By the induction hypothesis, the leading principal $(n-1)\times(n-1)$ submatrix of the matrix $(\lambda A + (1-\lambda)B)^{\circ(\alpha-1)}$ is positive semidefinite (even though the entire matrix need not be so). By Lemma 9.5, $A - B$ is positive semidefinite and has last row and column zero. It follows by the Schur product theorem that the integrand on the right-hand side is positive semidefinite. Since $B^{\circ\alpha}$ is a rank-1 positive semidefinite matrix (this is easy to verify), it follows that $A^{\circ\alpha}$ is also positive. This concludes one direction of the proof.

To prove the other half, suppose $\alpha \notin \mathbb{Z}^{\geq 0} \cup [n-2,\infty)$; now consider H_μ where $\mu = \delta_1 + \epsilon\delta_x$ for $\epsilon, x > 0$, $x \neq 1$. Note that $(H_\mu)_{jk} = s_{j+k}(\mu) = 1 + \epsilon x^{j+k}$; and as shown previously, H_μ is positive semidefinite of rank 2.

First, suppose $\alpha < 0$. Then consider the leading principal 2×2 submatrix of $H_\mu^{\circ\alpha}$

$$B := \begin{pmatrix} (1+\epsilon)^\alpha & (1+\epsilon x)^\alpha \\ (1+\epsilon x)^\alpha & (1+\epsilon x^2)^\alpha \end{pmatrix}.$$

We claim that $\det B < 0$, which shows $H_\mu^{\circ\alpha}$ is not positive semidefinite. Indeed, note that

$$(1+\epsilon)(1+\epsilon x^2) - (1+\epsilon x)^2 = \epsilon(x-1)^2 > 0,$$

so $\det B = (1+\epsilon)^\alpha (1+\epsilon x^2)^\alpha - (1+\epsilon x)^{2\alpha} < 0$ because $\alpha < 0$.

Next, suppose that $\alpha \in (0, n-2) \setminus \mathbb{N}$. Given $x > 0$, for small ϵ we know by the binomial theorem that

$$(1+\epsilon x)^\alpha = 1 + \sum_{k \geq 1} \binom{\alpha}{k} \epsilon^k x^k, \qquad \text{where} \qquad \binom{\alpha}{k} = \frac{\alpha(\alpha-1)\cdots(\alpha-k+1)}{k!}.$$

We will produce $u \in \mathbb{R}^n$, such that $u^T H_\mu^{\circ\alpha} u < 0$; note this shows that $H_\mu^{\circ\alpha} \notin \mathbb{P}_n$.

Starting with the matrix $H_\mu = \mathbf{1}\mathbf{1}^T + \epsilon v v^T$ where $v = (1, x, \dots, x^{n-1})^T$, we obtain:

$$H_\mu^{\circ\alpha} = \mathbf{1}\mathbf{1}^T + \sum_{k=1}^{\lfloor \alpha \rfloor + 2} \epsilon^k \binom{\alpha}{k} (v^{\circ k})(v^{\circ k})^T + o(\epsilon^{\lfloor \alpha \rfloor + 2}), \tag{9.8}$$

where $o(\epsilon^{\lfloor \alpha \rfloor + 2})$ is a matrix, such that the quotient of any entry by $\epsilon^{\lfloor \alpha \rfloor + 2}$ goes to zero as $\epsilon \to 0^+$.

Note that the first term and the sum together contain at most n terms. Since the corresponding vectors $\mathbf{1}, v, v^{\circ 2}, \dots, v^{\circ(\lfloor \alpha \rfloor + 2)}$ are linearly independent (by considering the – possibly partial – usual Vandermonde matrix formed by them), there exists a vector $u \in \mathbb{R}^n$ satisfying

$$u^T \mathbf{1} = u^T v = u^T v^{\circ 2} = \cdots = u^T v^{\circ(\lfloor \alpha \rfloor + 1)} = 0, \qquad u^T v^{\circ(\lfloor \alpha \rfloor + 2)} = 1.$$

Substituting these into the above computation, we obtain

$$u^T H_\mu^{\circ\alpha} u = \epsilon^{\lfloor \alpha \rfloor + 2} \binom{\alpha}{\lfloor \alpha \rfloor + 2} + u^T \cdot o(\epsilon^{\lfloor \alpha \rfloor + 2}) \cdot u.$$

Since $\binom{\alpha}{\lfloor \alpha \rfloor + 2}$ is negative if α is not an integer, it follows that $\lim_{\epsilon \to 0^+} \dfrac{u^T H_\mu^{\circ\alpha} u}{\epsilon^{\lfloor \alpha \rfloor + 2}} < 0$. Hence one can choose a small $\epsilon > 0$, such that $u^T H_\mu^{\circ\alpha} u < 0$. For this ϵ, $H_\mu^{\circ\alpha}$ is not positive semidefinite. $\square$

Remark 9.9. As the above proof reveals, the following are equivalent for $n \geq 2$ and $\alpha \in \mathbb{R}$:

(1) The entrywise map x^α preserves positivity on $\mathbb{P}_n((0,\infty))$ (or $\mathbb{P}_n([0,\infty))$).
(2) $\alpha \in \mathbb{Z}^{\geq 0} \cup [n-2, \infty)$.
(3) The entrywise map x^α preserves positivity on the (leading principal $n \times n$ truncations of) Hankel moment matrices of non-negative measures supported on $\{1, x\}$, for any fixed $x > 0,\ x \neq 1$.

The use of the Hankel moment matrix counterexample $\mathbf{1}\mathbf{1}^T + \epsilon v v^T$ for $v = (1, x, \dots, x^{n-1})^T$ and small $\epsilon > 0$ was not due to FitzGerald and Horn – who used $v = (1, 2, \dots, n)^T$ instead – but due to Fallat, Johnson, and Sokal [113]. In fact, the above proof can be made to work if one uses any vector v with distinct positive real coordinates, and small enough $\epsilon > 0$.

As these remarks show, to isolate the entrywise powers preserving positivity on $\mathbb{P}_n((0,\infty))$, it suffices to consider a much smaller family – namely, the one-parameter family of truncated moment matrices of the measures $\delta_1 + \epsilon\delta_x$ – or the one-parameter family $\mathbf{1}_{n\times n} + \epsilon v v^T$, where $v = (x_1, \dots, x_n)^T$ for pairwise distinct $x_j > 0$. In fact, a stronger result is true. In her 2017 paper [191] in *Linear Algebra Appl.*, Jain was able to eliminate the dependence on ϵ:

Theorem 9.10 (Jain)**.** *Suppose $n > 0$ is an integer, and $x_1, x_2, \dots, x_n$ are pairwise distinct positive real numbers. Let $C := (1 + x_j x_k)_{j,k=1}^n$. Then $C^{\circ\alpha}$ is positive semidefinite if and only if $\alpha \in \mathbb{Z}^{\geq 0} \cup [n-2, \infty)$.*

In other words, this result identifies a multiparameter family of matrices, *each one of which* encodes the positivity preserving powers in the original result of FitzGerald–Horn.

We defer the proof of Theorem 9.10 (in fact, a stronger form shown by Jain in 2020, [192]) to Section 15. We then use this stronger variant to prove the corresponding result for entrywise powers preserving other Loewner properties: monotonicity (again shown by Jain in 2020), and hence convexity, both with the same multiparameter family of matrices.

The next result is an application of Theorem 9.3 to classify the entrywise powers that preserve positive definiteness:

Corollary 9.11. *Given an integer $n \geq 2$ and a scalar $\alpha \in \mathbb{R}$, the following are equivalent:*

1. *The entrywise αth power preserves positive definiteness for $n \times n$ matrices with positive entries.*
2. *The entrywise αth power preserves positive definiteness for $n \times n$ Hankel matrices with positive entries.*
3. *$\alpha \in \mathbb{Z}^{>0} \cup [n-2, \infty)$.*

Proof. Clearly, (1) $\implies$ (2). Next, if $\alpha = 0$ then x^α sends every matrix with positive entries to $\mathbf{1}_{n \times n}$, which is not positive definite. Now suppose α is not in $\mathbb{Z}^{\geq 0} \cup [n-2, \infty)$. Then given $1 \neq x \in (0, \infty)$, there exists $\epsilon > 0$, such that $A^{\circ\alpha}$ has a negative principal minor, where $A := (1 + \epsilon x^{j+k})_{j,k=0}^{n-1}$ is Hankel. Now perturb A by $\delta H_1'$ for small enough $\delta > 0$, where $H_1' := (e^{(j+k)^2})_{j,k=0}^{n-1}$ is a Hankel "principal submatrix" drawn from the kernel in Lemma 6.9. By Theorem 4.1, $A + \delta H_1'$ is TP Hankel for all $\delta > 0$, hence positive definite; and for small enough $\delta > 0$, its αth power also has a negative principal minor. This shows the contrapositive to (2) $\implies$ (3).

Finally, suppose (3) holds. Since the Schur product is a principal submatrix of the Kronecker product, it follows from the first proof of Theorem 3.12 that positive integer powers entrywise preserve positive definiteness. Now suppose $\alpha \geq n-2$ and $A_{n \times n}$ is positive definite. Then all eigenvalues of A are positive, so there exists $\epsilon > 0$, such that $A - \epsilon \operatorname{Id}_{n \times n} \in \mathbb{P}_n([0, \infty))$. Now we have

$$A^{\circ\alpha} = (A - \epsilon \operatorname{Id})^{\circ\alpha} + \operatorname{diag}(a_{jj}^\alpha - (a_{jj} - \epsilon)^\alpha)_{j=1}^n,$$

and the first term on the right-hand side is in $\mathbb{P}_n$ by Theorem 9.3, so $A^{\circ\alpha}$ is positive definite. □

We conclude by highlighting the power and applicability of the "integration trick" (9.7) of FitzGerald and Horn. First, it in fact applies to general functions, not just to powers. The following observation (by the author and Tao, [217]) will be useful below:

Theorem 9.12 (Extension principle)**.** *Let $0 < \rho \leq \infty$ and $I = (0, \rho)$ or $(-\rho, \rho)$. Fix an integer $n \geq 2$ and a continuously differentiable function $h : I \to \mathbb{R}$. If $h[-]$ preserves positivity on rank-1 matrices in $\mathbb{P}_n(I)$ and $h'[-]$ preserves positivity on $\mathbb{P}_{n-1}(I)$, then $h[-]$ preserves positivity on $\mathbb{P}_n(I)$.*

The proof is exactly as before, but now using the more general integral identity:

$$h(x) - h(y) = \int_y^x h'(t)\, dt = \int_0^1 (x-y) h'(\lambda x + (1-\lambda) y)\, d\lambda.$$

Second, this integration trick is even more powerful, in that it further applies to classify the entrywise powers that preserve other properties of $\mathbb{P}_n$, including monotonicity and superadditivity. See Section 14 for details on these properties, their power-preservers, and their further application to the distinguished sub-cones $\mathbb{P}_G$ for non-complete graphs G.

10. Entrywise powers preserving total positivity: I.

The next goal is to study which entrywise power functions preserve total positivity and total non-negativity. The present section is devoted to proving:

Theorem 10.1. *If $A_{m\times n}$ is TP_3, then so is $A^{\circ t}$ for all $t \geq 1$.*

The proof relies on *Descartes' rule of signs* (also known as *Laguerre's rule of signs* – for the connection, see Section 29.3). Recall that we had shown a 'weak' variant of this in Lemma 5.2. The next variant is stronger, and relies on the following notion.

Definition 10.2. Suppose $F : \mathbb{R} \to \mathbb{R}$ is infinitely differentiable. Given an integer $k \geq 0$, we say F has a *zero of order k* at $t_0 \in \mathbb{R}$, if $F(t_0) = F'(t_0) = \cdots = F^{(k-1)}(t_0) = 0$ and $F^{(k)}(t_0) \neq 0$. (Note that a zero of order 0 means that $F(t_0) \neq 0$.)

Descartes' rule of signs bounds the number of real zeros of *generalized Dirichlet polynomials*, which are functions of the form

$$F : \mathbb{R} \to \mathbb{R}, \qquad F(t) = \sum_{j=1}^{n} c_j e^{\alpha_j t}, \quad c_j, \alpha_j \in \mathbb{R}.$$

These functions are so named because changing variables to $x = e^t$ gives

$$f(x) = \sum_{j=1}^{n} c_j x^{\alpha_j} : (0, \infty) \to \mathbb{R},$$

which are known as *generalized polynomials*. Another special case of $F(t)$ is when one uses $\alpha_j = -\log(j)$, to obtain $F(t) = \sum_{j=1}^{n} c_j / j^t$; these are called *Dirichlet polynomials*. The generalized Dirichlet polynomials subsume both of these families of examples.

We can now state

Theorem 10.3 (Decartes' rule of signs)**.** *Suppose $F(t) = \sum_{j=1}^{n} c_j e^{\alpha_j t}$ as above, with $c_j \in \mathbb{R}$ not all zero, and $\alpha_1 > \alpha_2 > \cdots > \alpha_n$ also real. Then the number of real zeros of F, counting multiplicities, is at most the number of sign changes, or 'variations', in the sequence $c_1, c_2, \dots, c_n$ (after removing all zero terms).*

For instance, the polynomial $x^6 - 8 = (x^2 - 2)(x^4 + 2x^2 + 4)$ has only one sign change, so at most one positive root – which is at $x = e^t = 2$.

To prove Theorem 10.3, we require a couple of preliminary lemmas.

Lemma 10.4 (Generalized Rolle's theorem)**.** *Given an open interval I and a smooth function $F : I \to \mathbb{R}$, let $Z(F, I)$ denote the number of zeros of F in I, counting orders. If $Z(F, I)$ is finite, then we have $Z(F', I) \geq Z(F, I) - 1$.*

Proof. Suppose F has a zero of order $k_r > 0$ at $x_r, 1 \leq r \leq n$. Then F' has a zero of order $k_r - 1 \geq 0$ at x_r. These add up to :

$$\sum_{r=1}^{n} (k_r - 1) = Z(F, I) - n$$

We may also suppose $x_1 < x_2 < \cdots < x_n$. Now by Rolle's theorem, F' also has at least $n-1$ zeros in the intervals (x_r, x_{r+1}) between the points x_r. Together, we obtain: $Z(F', I) \geq Z(F, I) - 1$. □

Lemma 10.5. *Let $F, G : I \to \mathbb{R}$ be smooth and $G \neq 0$ on I. If F has a zero of order k at t_0 then so does $F \cdot G$.*

Proof. This is straightforward: use Leibnitz's rule to compute $(F\cdot G)^{(j)}(t_0)$ for $0 \leq j \leq k$. $\square$

With these lemmas in hand, we can prove Descartes' rule of signs.

Proof of Theorem 10.3. The proof again follows Laguerre's argument (1883), by induction on the number s of sign changes in the sequence $c_1, c_2, \dots, c_n$. (Note that not all c_j are zero.) The base case is $s = 0$, in which case $F(t) = \sum_{j=1}^n c_j e^{\alpha_j t}$ has all non-zero coefficients of the same sign, and hence never vanishes.

For the induction step, we first assume without loss of generality that all c_j are non-zero. Suppose the last sign change occurs at c_k, i.e., $c_k c_{k-1} < 0$. Choose and fix $\alpha \in (\alpha_k, \alpha_{k-1})$, and define $G(t) := e^{-\alpha t}$. Then,

$$H(t) := F(t) \cdot G(t) = \sum_{j=1}^n c_j e^{(\alpha_j - \alpha)t}$$

has the same zeros (with orders) as $F(t)$, by Lemma 10.5. Moreover,

$$H'(t) = \sum_{j=1}^n c_j(\alpha_j - \alpha) e^{(\alpha_j - \alpha)t}$$

has exactly one less sign change than $F(t)$, namely, $s-1$. It follows by the induction hypothesis that $Z(H', \mathbb{R}) \leq s - 1$. Hence by Lemma 10.4, $Z(F, \mathbb{R}) = Z(H, \mathbb{R}) \leq 1 + Z(H', \mathbb{R}) \leq s$, and the proof is complete by induction. $\square$

We remark that there are numerous strengthenings of Descartes' rule of signs in the literature, obtained by Budan [75], Fourier [129], Laguerre [229], Segner [333], Sturm [348], and many others – this was popular even in the 20th century, see e.g. the articles by Curtiss [97] in *Ann. of Math.* (1918) and by Hurwitz [189] in *Math. Ann.* (1920). Here we restrict ourselves to mentioning some of these variants without proofs (although we remark that their proofs are quite accessible – see for instance the 2006 survey [193] by Jameson in *Math. Gazette*). As above, let $F(t) = \sum_{j=1}^n c_j e^{\alpha_j t}$ with $\alpha_1 > \alpha_2 > \cdots > \alpha_n$.

(1) Then not only is $Z(F, \mathbb{R}) \leq s$, but $s - Z(F, \mathbb{R})$ is an even integer. This was shown by Budan [75] and Fourier [129], and is also attributed to Le Gua. In fact, Budan–Fourier showed a more general result, stated here for polynomials:

Theorem 10.6 (Budan–Fourier)**.** *Suppose f is a polynomial, and $-\infty < \alpha < \beta < \infty$. Denoting by $V(x)$ the number of sign changes in the sequence $(f(x), f'(x), f''(x), \dots)$,*

$$V(\alpha) - V(\beta) - Z(f, (\alpha, \beta))$$

is a non-negative integer, which is moreover even if α, β are not zeros of f.

Notice that $V(\beta) \to 0$ as $\beta \to \infty$, e.g. by considering f monic. From this the above assertion for $Z(f, (0, \infty))$ – or $Z(F, \mathbb{R})$ (for rational α_j) – follows. Curiously, in his 1934 paper [312] in *Math. Z.*, Schoenberg proved this result using totally non-negative matrices.

(2) Define the partial sums

$$C_1 := c_1, \quad C_2 := c_1 + c_2, \quad \dots, \quad C_n := c_1 + c_2 + \cdots + c_n.$$

Then the number of positive roots of $F(t)$, counting orders, is at most the number of sign changes in $C_1, C_2, \dots, C_n$.

(3) Similarly, the number of negative roots of $F(t)$, counting orders, is the number of positive roots of $F(-t)$, hence at most the number of sign changes in the 'reverse sequence'

$$D_1 := c_n, \quad D_2 := c_n + c_{n-1}, \quad \dots, \quad D_n := c_n + c_{n-1} + \cdots + c_1.$$

Finally, we use Descartes' rule of signs to show the result stated above: that all powers ≥ 1 preserve total positivity of order 3.

Proof of Theorem 10.1. It is easy to check that all entrywise powers $\alpha \geq 1$ preserve the TP_2 property. We now show that the positivity of all 3×3 minors is also preserved by entrywise applying $x^t,\ t \geq 1$. Without loss of generality, we may assume $m = n = 3$ and work with a TP matrix $B_{3\times 3} = (b_{jk})_{j,k=1}^3$.

The first claim is that we may assume $B = \begin{pmatrix} 1 & 1 & 1 \\ 1 & a & b \\ 1 & c & d \end{pmatrix}$. Indeed, define the diagonal matrices

$$D_1 := \begin{pmatrix} 1/b_{11} & 0 & 0 \\ 0 & 1/b_{21} & 0 \\ 0 & 0 & 1/b_{31} \end{pmatrix}, \qquad D_2 := \begin{pmatrix} 1 & 0 & 0 \\ 0 & b_{11}/b_{12} & 0 \\ 0 & 0 & b_{11}/b_{13} \end{pmatrix}.$$

Using the Cauchy–Binet formula, one shows that $B = (b_{jk})_{j,k=1}^3$ is totally positive if and only if $D_1 B D_2$ is TP. But check that $D_1 B D_2$ has only ones in its first row and column, as desired.

The next observation is that a matrix $A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & a & b \\ 1 & c & d \end{pmatrix}$ is TP_2 if and only if $b, c > a > 1$ and $ad > bc$. (This is easily verified, and in turn implies $d > b, c$.)

Now consider $A^{\circ t}$. For $t \geq 1$ this is TP_2 by above, so we only need to consider $\det A^{\circ t}$ for $t \geq 1$. Define the generalized Dirichlet polynomial

$$F(t) := \det(A^{\circ t}) = (ad)^t - d^t - (bc)^t + b^t + c^t - a^t, \qquad t \in \mathbb{R}.$$

Notice from the above inequalities involving a, b, c, d that regardless of whether or not $d > bc$, and whether or not $b > c$, the sign sequence remains unchanged when arranging the exponents in F (namely, $\log ad, \log d, \log bc, \log b, \dots$) in decreasing order. It follows by Theorem 10.3 that F has at most three real roots.

As $t \to \infty$, $F(t) \to \infty$. Now one can carry out a Taylor expansion of F and check that the constant and linear terms vanish, yielding:

$$F(t) = e^{t \log(ad)} - e^{t \log(d)} - \cdots = t^2 (\log(a)\log(d) - \log(b)\log(c)) + o(t^2).$$

It follows that F has (at least) a double root at 0. Now claim that F is indeed positive on $(1, \infty)$, as desired. For if F is negative on $(1, \infty)$, then since $F(1) = \det A > 0$, it follows by continuity that F has at least two more roots in $(1, \infty)$, which is false. Hence $F \geq 0$ on $(1, \infty)$. If $F(t_0) = 0$ for some $t_0 > 1$, then t_0 is a global minimum point in $[1, t_0 + 1]$ for F, hence $F'(t_0) = 0$. But then F has at least two zeros at $t_0 \in (1, \infty)$, which is false. $\square$

While Theorem 10.3 sufficed in proving Theorem 10.1, we will need in Section 30.4 below a similar variant for 'usual' polynomials – more precisely, for Laurent series with degrees bounded below. This is now shown:

Theorem 10.7. *Fix an integer $n_0 \geq 0$ and an open interval $I \subset (0, \infty)$. Suppose $F : I \to \mathbb{R}$, sending $t \mapsto \sum_{j=-n_0}^{\infty} c_j t^j$ is a convergent power series with not all $c_j \in \mathbb{R}$ zero. Then the number of zeros of F in I, counting multiplicities, is either infinite or at most the number of sign changes in the sequence $c_{-n_0}, c_{1-n_0}, \dots$ (after removing all zero terms).*

Proof. If the Maclaurin coefficients c_j have infinitely many sign changes then the result is immediate. Otherwise suppose there are only finitely many sign changes in the c_j, say s many. We show the result by induction on $0 \leq s < \infty$, with the result immediate for $s = 0$. For the induction step, suppose the last sign change occurs sat c_k, i.e., c_k has sign opposite to the immediately preceding non-zero Maclaurin coefficient of F.

Now let $R > 0$ denote the radius of convergence of the power series $t^{n_0} F(t)$, so that F is smooth on $(0, R)$ by (repeatedly) using the quotient rule. Consider the function

$$G(u) := u^{1-2k} F(u^2), \quad 0 < u < \sqrt{R}.$$

Then G is smooth on $(0, \sqrt{R})$, and we now proceed as in (Laguerre's 1883) proof of Theorem 10.3 via Rolle's theorem – now working solely in $(0, \sqrt{R})$. The Laurent series $G'(u)$ has one less sign change than does F, so at most $s - 1$ roots (counting multiplicities) in $(0, \sqrt{R})$ by the induction hypothesis. Hence, G has at most s roots in $(0, \sqrt{R})$. But then so does F in $(0, R)$, hence in I. $\square$

11. Entrywise powers preserving total positivity: II.

In the previous section, we used Descartes' rule of signs to show that x^α entrywise preserves the 3×3 TP matrices, for all $\alpha \geq 1$. Here the goal is twofold: first, to completely classify the entrywise powers that preserve TP/TN for $m \times n$ matrices for each fixed $m, n \geq 1$; and second, to then classify *all continuous functions* that do the same (at present, only for TN).

Corollary 11.1. *If $\alpha \geq 1$, then x^α entrywise preserves the 3×3 TN matrices.*

Proof. Let $A_{3\times 3}$ be TN and $\alpha \geq 1$. By Whitney's density theorem 6.7, there exist 3×3 TP matrices B_m that entrywise converge to A, as $m \to \infty$. Hence $B_m^{\circ\alpha} \to A^{\circ\alpha}$ for $\alpha \geq 1$. Since $B_m^{\circ\alpha}$ is TP by Theorem 10.1, it follows that $A^{\circ\alpha}$ is TN, as claimed. □

The next result classifies all entrywise powers preserving total non-negativity for matrices of any fixed size.

Theorem 11.2. *Given integers $m, n > 0$, define $d := \min(m, n)$. The following are equivalent for $\alpha \in \mathbb{R}$.*

(1) x^α preserves (entrywise) the $m \times n$ TN matrices.
(2) x^α preserves (entrywise) the $d \times d$ TN matrices.
(3) Either $\alpha = 0$ (where we set $0^0 := 1$), or
  - *(a) For $d = 1, 2$: $\alpha \geq 0$.*
  - *(b) For $d = 3$: $\alpha \geq 1$.*
  - *(c) For $d \geq 4$: $\alpha = 1$.*

Thus we see that in contrast to the entrywise preservers of positive semidefiniteness (see Theorem 9.3), almost no powers preserve the TN matrices – nor the TP matrices, as we show presently.

Proof. That (2) $\implies$ (1) is straightforward, as is (1) $\implies$ (2) by padding by zeros – noting that negative powers are not allowed (given zero entries of TN matrices). To show (3) $\implies$ (2), we use Theorem 10.1 as well as that x^0 applied to any TN matrix yields the matrix of all ones.

It remains to prove (2) $\implies$ (3). We may rule out negative powers since $(0_{d\times d})^{\circ\alpha}$ is not defined for $d \geq 1$. Similarly, x^0 always preserves total non-negativity. This shows (3) for $d = 1, 2$. For $d = 3$, suppose $\alpha \in (0, 1)$ and consider the matrix

$$A = \begin{pmatrix} 1 & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & 1 & \frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & 1 \end{pmatrix}. \tag{11.3}$$

This is a Toeplitz cosine matrix, hence TN (see Example 3.23, or verify directly). Now compute:

$$\det A^{\circ\alpha} = \det \begin{pmatrix} 1 & (\sqrt{2})^{-\alpha} & 0 \\ (\sqrt{2})^{-\alpha} & 1 & (\sqrt{2})^{-\alpha} \\ 0 & (\sqrt{2})^{-\alpha} & 1 \end{pmatrix} = 1 - 2^{1-\alpha},$$

which is negative if $\alpha < 1$. So $A^{\circ\alpha}$ is not TN (not even positive semidefinite, in fact), for $\alpha < 1$, which shows (3) for $d = 3$.

Next suppose $d = 4$ and consider the matrix

$$N(x) = \mathbf{1}_{4\times 4} + x \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 3 \\ 0 & 2 & 4 & 6 \\ 0 & 3 & 8 & 14 \end{pmatrix}, \qquad x \geq 0.$$

One verifies that: all 2×2 minors are of the form $ax + bx^2$, where $a > 0, b \geq 0$; all 3×3 minors are of the form cx^2, where $c \geq 0$; and $\det N(x) = 0$. This implies $N(x)$ is TN for $x \geq 0$. Moreover, for small $x > 0$, computations similar to the proof of Theorem 10.1 show that

$$\det N(x)^{\circ t} = 2(t^3 - t^4)x^4 + o(x^4),$$

so given $t > 1$, it follows that $\det N(x)^{\circ t} < 0$ for sufficiently small $x > 0$. Thus $N(x)^{\circ t}$ is not TN, hence x^α does not preserve 4×4 TN matrices for $\alpha > 1$. If on the other hand $\alpha \in (0,1)$, then we work with the 4×4 TN matrix $C = \begin{pmatrix} 1 & \frac{1}{\sqrt{2}} & 0 & 0 \\ \frac{1}{\sqrt{2}} & 1 & \frac{1}{\sqrt{2}} & 0 \\ 0 & \frac{1}{\sqrt{2}} & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$, proceeding as in the $d = 3$ case. This concludes the proof for $d = 4$.

Finally if $d > 4$ then we use the TN matrices $\begin{pmatrix} N(x) & 0 \\ 0 & 0 \end{pmatrix}_{d \times d}$; for small $x > 0$ this rules out the powers $\alpha > 1$ as above. Similarly, using the TN matrix $\begin{pmatrix} C & 0 \\ 0 & 0 \end{pmatrix}_{d \times d}$ rules out the powers in $(0,1)$. $\square$

In turn, Theorem 11.2 helps classify the powers preserving total positivity in each fixed size.

Corollary 11.4. *Given $m, n > 0$, define $d := \min(m,n)$ as in Theorem 11.2. The following are equivalent for $\alpha \in \mathbb{R}$:*

(1) x^α preserves entrywise the $m \times n$ TP matrices.
(2) x^α preserves entrywise the $d \times d$ TP matrices.
(3) We have:
  - *(a) For $d = 1$: $\alpha \in \mathbb{R}$.*
  - *(b) For $d = 2$: $\alpha > 0$.*
  - *(c) For $d = 3$: $\alpha \geq 1$.*
  - *(d) For $d \geq 4$: $\alpha = 1$.*

Proof. That (2) $\implies$ (1) is straightforward, as is (1) $\implies$ (2) (as above) by now using Theorem 7.5. That (3) $\implies$ (2) was shown in Theorem 10.1 for $d = 3$, and is obvious for $d \neq 3$. Finally, we show (2) $\implies$ (3). The $d = 1$ case is trivial, while the $d = 2$ case follows by considering $\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$, say. Next, if $d \geq 3$ and if x^α preserves the $d \times d$ TP matrices, then $\alpha > 0$, by considering any TP matrix and applying x^α to any of its 2×2 minors. Hence x^α extends continuously to $x = 0$; now x^α preserves the $d \times d$ TN matrices by continuity. Theorem 11.2 now finishes the proof. $\square$

Next, we tackle the more challenging question of classifying *all* functions that entrywise preserve total positivity or total non-negativity in fixed dimension $m \times n$. We will show that (i) every such function must be continuous (barring a one-parameter exceptional family of

TN preservers), and in turn, this implies that (ii) it must be a power function. We first show (ii), beginning with an observation on the *(additive) Cauchy functional equation.*

Remark 11.5 (Additive continuous functions)**.** Suppose $g : \mathbb{R} \to \mathbb{R}$ is continuous and satisfies the Cauchy functional equation $g(x+y) = g(x) + g(y)$ for all $x, y \in \mathbb{R}$. Then we claim that $g(x) = cx$ for some $c \in \mathbb{R}$ (and all x). Indeed, $g(0+0) = g(0) + g(0)$, and hence $g(0) = 0$. Next, one shows by induction that $g(n) = ng(1)$ for integers $n > 0$, and hence for all integers $n < 0$ as well. Now one shows that $pg(1) = g(p) = g(q \cdot p/q) = q \cdot g(p/q)$ for integers p, q with $q \neq 0$, from which it follows that $g(p/q) = (p/q)g(1)$ for all rationals p/q. Finally, using continuity we conclude that $g(x) = xg(1)$ for all $x \in \mathbb{R}$.

Proposition 11.6. *Suppose $f : [0, \infty) \to \mathbb{R}$ is continuous and entrywise preserves the 2×2 TN matrices. Then $f(x) = f(1)x^\alpha$ for some $\alpha \geq 0$.*

We recall here that $0^0 := 1$ by convention.

Proof. Define the matrices

$$A(x,y) = \begin{pmatrix} x & xy \\ 1 & y \end{pmatrix}, \qquad B(x,y) = \begin{pmatrix} xy & x \\ y & 1 \end{pmatrix}, \qquad x, y \geq 0.$$

Clearly, these matrices are TN, so by the hypotheses,

$$\det f[A(x,y)] = f(x)f(y) - f(1)f(xy) \geq 0,$$
$$\det f[B(x,y)] = f(1)f(xy) - f(x)f(y) \geq 0.$$

It follows that

$$f(x)f(y) = f(1)f(xy), \qquad \forall x, y \geq 0. \tag{11.7}$$

There are two cases. First if $f(1) = 0$ then choosing $x = y \geq 0$ in (11.7) gives $f \equiv 0$ on $[0, \infty)$. Else if $f(1) > 0$ then we claim that f is always positive on $(0, \infty)$. Indeed, if $f(x_0) = 0$ for $x_0 > 0$, then set $x = x_0, y = 1/x_0$ in (11.7) to get: $0 = f(1)^2$, which is false.

Now define the functions

$$g(x) := f(x)/f(1),\ x > 0, \qquad h(y) := \log g(e^y),\ y \in \mathbb{R}.$$

Then (11.7) can be successively reformulated as:

$$\begin{aligned} g(xy) &= g(x)g(y), \qquad \forall x, y > 0, \\ h(a+b) &= h(a) + h(b), \qquad \forall a, b \in \mathbb{R}. \end{aligned} \tag{11.8}$$

Moreover, both g, h are continuous. Since h satisfies the additive Cauchy functional equation, it follows by Remark 11.5 that $h(y) = yh(1)$ for all $y \in \mathbb{R}$. Translating back, we get $g(x) = x^{h(1)}$ for all $x > 0$. It follows that $f(x) = f(1)x^\alpha$ for $x > 0$, where $\alpha = h(1)$. Finally, since f is also continuous at 0^+, it follows that $\alpha \geq 0$; and either $\alpha = 0$ and $f \equiv 1$ (so we set $0^0 := 1$), or $f(0) = 0 < \alpha$. (Note that α cannot be negative, since $f[-]$ preserves TN on the zero matrix, say.) $\square$

Corollary 11.9. *Suppose $f : [0, \infty) \to \mathbb{R}$ is continuous and entrywise preserves the $m \times n$ TN matrices, for some $m, n \geq 2$. Then $f(x) = f(1)x^\alpha$ for some $\alpha \geq 0$, with $f(1) \geq 0$.*

Proof. Given $m, n \geq 2$, every 2×2 TN matrix can be embedded as a leading principal submatrix in a $m \times n$ TN matrix, by padding it with (all other) zero entries. Hence the hypotheses imply that $f[-]$ preserves the 2×2 TN matrices, and we are done by the above Proposition 11.6. $\square$

12. Entrywise functions preserving total positivity. Mid-convex implies continuity. The test set of Hankel TN matrices.

We continue working toward the classification of all entrywise functions preserving $m \times n$ TP/TN matrices. Thus far, we have classified the power functions among these preservers; and we also showed that every continuous map that preserves $m \times n$ TN matrices is a multiple of a power function.

We now show that every function that entrywise preserves the $m \times n$ TP/TN matrices is automatically continuous on $(0, \infty)$ – which allows us to classify *all* such preservers. The continuity will follow from a variant of a 1929 result by Ostrowski [275] on mid-convex functions on normed linear spaces, and we begin by proving this result.

12.1. Mid-convex functions and continuity.

Definition 12.1. Given a convex subset U of a real vector space, a function $f : U \to \mathbb{R}$ is said to be *mid-convex* if

$$f\left(\frac{x+y}{2}\right) \leq \frac{f(x)+f(y)}{2}, \quad \forall x, y \in U;$$

and f is *convex* if $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$ for all $x, y \in U$ and $\lambda \in (0,1)$.

Notice that convex functions are automatically mid-convex. The converse need not be true in general. However, if a mid-convex function is continuous, then it is easy to see that it is also convex. Thus, a natural question for mid-convex functions is to find sufficient conditions under which they are continuous. We now discuss two such conditions, both classical results. The first condition is mild: f is locally bounded, *on one neighborhood of one point.*

Theorem 12.2. *Let $\mathbb{B}$ be a normed linear space (over $\mathbb{R}$) and let U be a convex open subset. Suppose $f : U \to \mathbb{R}$ is mid-convex and f is bounded above in an open neighborhood of a single point $x_0 \in U$. Then f is continuous on U, and hence convex.*

This generalizes to normed linear spaces a special case of a result by Ostrowski, who showed in *Jber. Deut. Math. Ver.* (1929) the same conclusion, but over $\mathbb{B} = \mathbb{R}$ and assuming f is bounded above in a measurable subset.

The proof requires the following useful observation:

Lemma 12.3. *If $f : U \to \mathbb{R}$ is mid-convex, then f is rationally convex, i.e., $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$ for all $x, y \in U$ and $\lambda \in (0,1) \cap \mathbb{Q}$.*

Proof. Inductively using mid-convexity, it follows that

$$f\left(\frac{x_1 + x_2 + \cdots + x_{2^n}}{2^n}\right) \leq \frac{f(x_1) + \cdots + f(x_{2^n})}{2^n}, \qquad \forall n \in \mathbb{N},\ x_1, \dots, x_{2^n} \in U.$$

Now suppose that $\lambda = \frac{p}{q} \in (0,1)$, where $p, q > 0$ are integers and $2^{n-1} \leq q < 2^n$ for some $n \in \mathbb{N}$. Let $x_1, \dots, x_q \in U$ and define $\overline{x} = \frac{1}{q}(x_1 + \cdots + x_q)$. Setting $x_{q+1} = \cdots = x_{2^n} = \overline{x}$, we obtain

$$\begin{aligned}
& f\left(\frac{x_1 + \cdots + x_q + (2^n - q)\bar{x}}{2^n}\right) \leq \frac{f(x_1) + \cdots + f(x_q) + (2^n - q) f(\overline{x})}{2^n} \\
\implies \quad & 2^n f(\overline{x}) \leq f(x_1) + \cdots + f(x_q) + (2^n - q) f(\overline{x}) \\
\implies \quad & q f(\overline{x}) \leq f(x_1) + \cdots + f(x_q) \\
\implies \quad & f\left(\frac{x_1 + \cdots + x_q}{q}\right) \leq \frac{f(x_1) + \cdots + f(x_q)}{q}
\end{aligned}$$

In this inequality, set $x_1 = \cdots = x_p = x$ and $x_{p+1} = \cdots = x_q = y$ to complete the proof:

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y). \qquad \square$$

With Lemma 12.3 in hand, we can prove the theorem above.

Proof of Theorem 12.2. We may assume without loss of generality that $x_0 = \mathbf{0} \in U \subset \mathbb{B}$, and also that $f(x_0) = f(\mathbf{0}) = 0$.

We claim that f is continuous at $\mathbf{0}$, where f was assumed to be bounded above in an open neighborhood of $\mathbf{0}$. Write this as: $f(B(\mathbf{0}, r)) < M$ for some $r, M > 0$, where $B(x, r) \subset \mathbb{B}$ denotes the open ball of radius r centered at $x \in \mathbb{B}$. Now given $\epsilon \in (0,1) \cap \mathbb{Q}$ rational and $x \in B(\mathbf{0}, \epsilon r)$, we compute using Lemma 12.3:

$$x = \epsilon \left(\frac{x}{\epsilon} \right) + (1-\epsilon)\mathbf{0} \qquad \Longrightarrow \qquad f(x) \leq \epsilon f \left(\frac{x}{\epsilon} \right) + 0 < \epsilon M.$$

Moreover,

$$\mathbf{0} = \left(\frac{\epsilon}{1+\epsilon} \right) \left(\frac{-x}{\epsilon} \right) + \frac{x}{1+\epsilon},$$

so applying Lemma 12.3 once again, we obtain:

$$0 \leq \left(\frac{\epsilon}{1+\epsilon} \right) f \left(\frac{-x}{\epsilon} \right) + \frac{f(x)}{1+\epsilon} < \frac{\epsilon M}{1+\epsilon} + \frac{f(x)}{1+\epsilon} \qquad \Longrightarrow \qquad f(x) > -\epsilon M.$$

Therefore, we have $x \in B(\mathbf{0}, \epsilon r) \implies |f(x)| < \epsilon M$.

Now given $\epsilon > 0$, choose $0 < \epsilon' < \min(M, \epsilon)$, such that ϵ'/M is rational, and set $\delta := r\epsilon'/M$. Then $\delta < r$, so $|f(x)| < \delta M / r = \epsilon' < \epsilon$ whenever $x \in B(\mathbf{0}, \delta)$. Hence, f is continuous at x_0.

We have shown that if f is bounded above in some open neighborhood of $x_0 \in U$, then f is continuous at x_0. To finish the proof, we claim that for all $y \in U$, f is bounded above on some open neighborhood of y. This would show that f is continuous on U, which combined with mid-convexity implies convexity.

To show the claim, choose a rational $\rho > 1$, such that $\rho y \in U$ (this is possible as U is open), and set $U_y := B(y, (1-1/\rho)r)$. Note that $U_y \subset U$ since for every $v \in U_y$ there exists $x \in B(\mathbf{0}, r)$, such that

$$v = y + (1 - 1/\rho)x = \frac{1}{\rho}(\rho y) + \left(1 - \frac{1}{\rho}\right) x.$$

Thus, v is a convex combination of $\rho y \in U$ and $x \in B(\mathbf{0}, r) \subset U$. Hence, $U_y \subset U$; in turn,

$$f(v) \leq \frac{1}{\rho} f(\rho y) + \left(1 - \frac{1}{\rho}\right) f(x) \leq \frac{f(\rho y)}{\rho} + \left(1 - \frac{1}{\rho}\right) M, \qquad \forall v \in U_y$$

by Lemma 12.3. Since the right-hand side is independent of $v \in U_y$, the above claim follows. Hence, by the first claim, f is indeed continuous at every point in U. $\square$

The second condition, which will be used in a later part in this text, is that f is Lebesgue measurable. Its sufficiency was proved a decade before Ostrowski's result, independently by Blumberg in *Trans. Amer. Math. Soc.* (1919) and Sierpińsky in *Fund. Math.* (1920). However, the following proof goes via Theorem 12.2:

Theorem 12.4. *If $I \subset \mathbb{R}$ is an open interval and $f : I \to \mathbb{R}$ is Lebesgue measurable and mid-convex, then f is continuous, hence convex.*

Proof. Suppose f is not continuous at a point $x_0 \in I$. Fix $c > 0$, such that $(x_0 - 2c, x_0 + 2c) \subset I$. By Theorem 12.2, f is unbounded on $(x_0 - c, x_0 + c)$. Now let $B_n := \{x \in I : f(x) > n\}$ for $n \geq 1$; note this is Lebesgue measurable. Choose $u_n \in B_n \cap (x_0 - c, x_0 + c)$ and $\lambda \in [0,1]$; then by mid-convexity,

$$n < f(u_n) = f\left[\frac{u_n + \lambda c}{2} + \frac{u_n - \lambda c}{2}\right] \leq \frac{1}{2}(f(u_n + \lambda c) + f(u_n - \lambda c)).$$

Thus, B_n contains at least one of the points $u_n \pm \lambda c \in I$, i.e., one of $\pm\lambda c$ lies in $B_n - u_n$. We now **claim** that each B_n has Lebesgue measure $\mu(B_n) \geq c$. Assuming this claim,

$$c \leq \lim_{n\to\infty} \mu(B_n) = \mu\left(\cap_{n\geq 1} B_n\right),$$

since the B_n are a nested family of subsets. But then $S := \cap_{n\geq 1} B_n$ is non-empty, so for any $v \in S$, we have $f(v) > n$ for all n, which produces the desired contradiction.

Thus, it remains to show the above claim. Fix $n \geq 1$ and note from above that $M_n := B_n - u_n$ is a Lebesgue measurable set, such that for every $\lambda \in [0,1]$, at least one of $\lambda c, -\lambda c$ lies in M_n. Define the measurable sets $A_1 := M_n \cap [-c, 0]$ and $A_2 := M_n \cap [0, c]$, so that $-A_1 \cup A_2 = [0, c]$. This implies

$$c \leq \mu(-A_1) + \mu(A_2) = \mu(A_1) + \mu(A_2) = \mu(A_1 \cup A_2) \leq \mu(M_n). \qquad \square$$

12.2. **Functions preserving total non-negativity.** With Theorem 12.2 in hand, it is possible to classify all entrywise functions that preserve total non-negativity or total positivity in a fixed size, or even positive semidefiniteness on 2×2 matrices. A major portion of the work is carried out by the next result. To state this result, we need the following notion.

Definition 12.5. Suppose $I \subset [0, \infty)$ is an interval. A function $f : I \to [0, \infty)$ is *multiplicatively mid-convex* on I if and only if $f(\sqrt{xy}) \leq \sqrt{f(x)f(y)}$ for all $x, y \in I$.

Remark 12.6. A straightforward computation yields that if $f : I \to \mathbb{R}$ is always positive and $0 \notin I$, then f is multiplicatively mid-convex on I if and only if the auxiliary function $g(y) := \log f(e^y)$ is mid-convex on $\log(I)$.

We now prove the following important result, which is also crucial later.

Theorem 12.7. *Suppose $I = [0, \infty)$ and $I^+ := I \setminus \{0\}$. A function $f : I \to \mathbb{R}$ satisfies $\begin{pmatrix} f(a) & f(b) \\ f(b) & f(c) \end{pmatrix}$ is positive semidefinite whenever $a, b, c \in I$ and $\begin{pmatrix} a & b \\ b & c \end{pmatrix}$ is TN, if and only if f is non-negative, non-decreasing, and multiplicatively mid-convex on I. In particular,*

(1) $f|_{I^+}$ is never zero or always zero.
(2) $f|_{I^+}$ is continuous.

The same results hold if $I = [0, \infty)$ is replaced by $I = (0, \infty), [0, \rho)$, or $(0, \rho)$ for $0 < \rho < \infty$.

This result was essentially proved by H.L. Vasudeva [353], under some reformulation. In the result, note that TN is the same as 'positive semidefinite with non-negative entries', since we are dealing with 2×2 matrices; thus, the test set of matrices is precisely $\mathbb{P}_2(I)$, and the hypothesis can be rephrased as:

$$f[-] : \mathbb{P}_2(I) \to \mathbb{P}_2 = \mathbb{P}_2(\mathbb{R}).$$

Moreover, all of these matrices are clearly Hankel. This result will therefore also play an important role when we classify the entrywise preservers of positive semidefiniteness on low-rank Hankel matrices (see Theorems 17.1 and 19.1).

Proof. Let I be any of the domains mentioned in the theorem. We begin by showing the equivalence. Given a TN matrix

$$\begin{pmatrix} a & b \\ b & c \end{pmatrix}, \qquad a,b,c \in I, \quad 0 \leq b \leq \sqrt{ac},$$

compute via the non-negativity, monotonicity, and multiplicative mid-convexity respectively:

$$0 \leq f(b) \leq f(\sqrt{ac}) \leq \sqrt{f(a)f(c)}.$$

It follows that $\begin{pmatrix} f(a) & f(b) \\ f(b) & f(c) \end{pmatrix}$ is TN and hence positive semidefinite.

Conversely, if (via the above remarks) $f[-] : \mathbb{P}_2(I) \to \mathbb{P}_2$, then apply $f[-]$ entrywise to the matrices

$$\begin{pmatrix} a & b \\ b & a \end{pmatrix}, \qquad \begin{pmatrix} a & \sqrt{ac} \\ \sqrt{ac} & c \end{pmatrix}, \qquad a,b,c \in I, \tag{12.8}$$

with $0 \leq b \leq a$. From the hypotheses, it successively (and respectively) follows that f is non-negative, non-decreasing, and multiplicatively mid-convex. This proves the equivalence.

As a brief digression that will be useful later, we remark that the test matrices $\begin{pmatrix} a & b \\ b & a \end{pmatrix}$ can be replaced by

$$\begin{pmatrix} a & b \\ b & b \end{pmatrix}, \qquad \begin{pmatrix} a & \sqrt{ac} \\ \sqrt{ac} & c \end{pmatrix}, \qquad a,b,c \in I, \tag{12.9}$$

with $0 \leq b \leq a$, to conclude as above that f is non-decreasing and multiplicatively mid-convex on I. Indeed, we obtain $f(a), f(b) \geq 0$, and either $f(b) = 0 \leq f(a)$, or $0 < f(b)^2 \leq f(b)f(a)$, leading to the same conclusion.

We now show the two final assertions (1) and (2) in the theorem, again on I^+ for any of the domains I above; in other words, $I^+ = (0, \rho)$ for $0 < \rho \leq \infty$. For (1), suppose $f(x) = 0$ for some $x \in I^+$. Since f is non-negative and non-decreasing on I^+, it follows that $f \equiv 0$ on $(0, x)$. Now claim that $f(y) = 0$ if $y > x,\ y \in I^+ = (0, \rho)$. Indeed, choose a large enough $n > 0$, such that $y \sqrt[n]{y/x} < \rho$. Set $\zeta := \sqrt[n]{y/x} > 1$ and consider the following rank-1 matrices in $\mathbb{P}_2(I^+)$:

$$A_1 := \begin{pmatrix} x & x\zeta \\ x\zeta & x\zeta^2 \end{pmatrix}, \qquad A_2 := \begin{pmatrix} x\zeta & x\zeta^2 \\ x\zeta^2 & x\zeta^3 \end{pmatrix}, \qquad \dots, \qquad A_n := \begin{pmatrix} x\zeta^{n-1} & x\zeta^n \\ x\zeta^n & x\zeta^{n+1} \end{pmatrix}.$$

The inequalities $\det f[A_k] \geq 0,\ 1 \leq k \leq n$ yield:

$$0 \leq f(x\zeta^k) \leq \sqrt{f(x\zeta^{k-1})f(x\zeta^{k+1})}, \qquad k = 1, 2, \dots, n.$$

From this inequality for $k = 1$, it follows that $f(x\zeta) = 0$. Similarly, these inequalities inductively yield: $f(x\zeta^k) = 0$ for all $1 \leq k \leq n$. In particular, we have $f(y) = f(x\zeta^n) = 0$. This shows that $f \equiv 0$ on I^+, as claimed.

We provide two proofs of (2). If $f \equiv 0$ on I^+, then f is continuous on I^+. Otherwise by (1), f is strictly positive on $(0, \rho) = I^+$. Now the "classical" proof uses the above "Ostrowski-result": define the function $g : \log I^+ := (-\infty, \log \rho) \to \mathbb{R}$ via

$$g(y) := \log f(e^y), \qquad y < \log \rho.$$

By the assumptions on f and the observation in Remark 12.6, g is mid-convex and non-decreasing on $(-\infty, \log \rho)$. In particular, g is bounded above on compact sets. Now apply Theorem 12.2 to deduce that g is continuous. It follows that f is continuous on $(0, \rho)$.

A more recent, shorter proof is by Hiai in 2009 [171]: given f as above which is strictly positive and non-decreasing on $(0,\rho)$, fix $t \in (0,\rho)$ and let $0 < \epsilon < \min(t/5, (\rho - t)/4)$. Then $0 < t + \epsilon \leq \sqrt{(t+4\epsilon)(t-\epsilon)} < \rho$, so

$$f(t+\epsilon) \leq f\left(\sqrt{(t+4\epsilon)(t-\epsilon)}\right) \leq \sqrt{f(t+4\epsilon)f(t-\epsilon)}.$$

Now letting $\epsilon \to 0^+$, this implies $f(t^+) \leq f(t^-)$, so

$$0 < f(t) \leq f(t^+) \leq f(t^-) \leq f(t), \qquad \forall t \in (0,\rho).$$

Since $t \in (0,\rho)$ was arbitrary, this shows f is continuous as claimed. $\square$

Remark 12.10. From the proof – see (12.8) – it follows that the assumptions may be further weakened to not work with all symmetric 2×2 TN matrices, but with only the rank-1 symmetric and the Toeplitz symmetric 2×2 TN matrices.

As an application, Theorem 12.7 allows us to complete the classification of all entrywise maps that preserve total non-negativity in each fixed size.

Theorem 12.11. *Suppose $m, n \geq 2$ and $f : [0,\infty) \to \mathbb{R}$ entrywise preserves the $m \times n$ TN matrices. Then either $f(x) = f(1)x^\alpha$ for $f(1), \alpha \geq 0$ and all $x \geq 0$ (and these powers were classified in Theorem 11.2), or* $\min(m,n) = 2$ *and $f(x) = f(1)\mathrm{sgn}(x)$ for $x \geq 0$ and $f(1) > 0$.*

If instead $\min(m,n) = 1$, *then f can be any function that maps $[0,\infty)$ into itself.*

Proof. The result is trivial for $\min(m,n) = 1$, so we assume henceforth that $m, n \geq 2$. By embedding 2×2 TN matrices inside $m \times n$ TN matrices, it follows that $f[-]$ preserves the 2×2 TN matrices. In particular, f is continuous on $(0,\infty)$ by Theorem 12.7, and non-negative and non-decreasing on $[0,\infty)$. Now one can repeat the proof of Proposition 11.6 above, to show that

$$f(x)f(y) = f(xy)f(1), \qquad \forall x, y \geq 0, \tag{12.12}$$

and moreover, either $f \equiv 0$ on $[0,\infty)$, or $f(x) = f(1)x^\alpha$ for $x > 0$ and some $\alpha \geq 0$.

We assume henceforth that $f \not\equiv 0$ on $[0,\infty)$, so $f(x) = f(1)x^\alpha$ as above, with $f(1) > 0$. If now $f(0) \neq 0$, then substituting $x = 0, y \neq 1$ in (12.12) shows that $\alpha = 0$, and now using $x = y = 0$ in (12.12) shows $f(0) = f(1)$, i.e., $f|_{[0,\infty)}$ is constant (and positive).

Otherwise $f(0) = 0$. Now if $\alpha > 0$ then $f(x) = f(1)x^\alpha$ for all $x \geq 0$ and f is continuous on $[0,\infty)$. The final case is where $f(0) = 0 = \alpha$, but $f \not\equiv 0$. Then $f(0) = 0$ while $f(x) = f(1) > 0$ for all $x > 0$. Now if $\min(m,n) = 2$ then it is easy to verify that $f[-]$ preserves $TN_{m \times n}$. On the other hand, if $m, n \geq 3$, then computing $\det f[A]$ for the matrix $A = \begin{pmatrix} 1 & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & 1 & \frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & 1 \end{pmatrix}$ shows that $f[-]$ is not a positivity preserver on $A \oplus 0_{(m-3)\times(n-3)} \in TN_{m \times n}$. $\square$

12.3. **Functions preserving total positivity.** Akin to the above results, we can also classify the entrywise functions preserving total positivity in any fixed size, and they too are essentially power functions.

Theorem 12.13. *Suppose $f : (0,\infty) \to \mathbb{R}$ is such that $f[-]$ preserves the $m \times n$ TP matrices for some $m, n \geq 2$. Then f is continuous and $f(x) = f(1)x^\alpha$, with $\alpha > 0$ and $f(1) > 0$.*

Recall that the powers preserving the $m \times n$ TP matrices were classified in Corollary 11.4.

To show the theorem, we make use of the following intermediate lemma, which is also useful later in studying preservers of TP Hankel kernels. A cruder version of the next result says that if $f[-]$ preserves total positivity on 2×2 matrices, then f is continuous. More strongly, we have:

Lemma 12.14. *Suppose $p \in (1, \infty)$, and $f : (0, \infty) \to \mathbb{R}$. The following are equivalent:*

(1) $f[-]$ preserves total positivity on all symmetric 2×2 TP matrices.

(2) $f[-]$ preserves positive definiteness on the symmetric TP Hankel matrices

$$\begin{pmatrix} a & b \\ b & c \end{pmatrix}, \qquad a, c > 0, \ \sqrt{ac/p} < b < \sqrt{ac}.$$

(3) f is positive, increasing, and multiplicatively mid-convex on $(0, \infty)$.

In particular, f is continuous.

Proof. Clearly, (1) $\implies$ (2), and that (3) $\implies$ (1) is left to the reader as it is similar to the proof of Theorem 12.7. We now assume (2) and show (3). The first step is to claim that f is positive and strictly increasing on $(0, \infty)$. Suppose $0 < x < y < \infty$. Choose $n > \log_p(y/x)$, and define the increasing sequence

$$x_0 = x, \quad x_1 = x(y/x)^{1/n}, \quad x_2 = x(y/x)^{2/n}, \quad \dots, \quad x_n = y.$$

Now the matrix $\begin{pmatrix} x_{k+1} & x_k \\ x_k & x_k \end{pmatrix}$ is in the given test set, by choice of n, so applying $f[-]$ and taking determinants, we have

$$f(x_k), \ f(x_{k+1}), \ f(x_k)(f(x_{k+1}) - f(x_k)) > 0, \qquad 0 \leq k \leq n - 1.$$

It follows that f is positive on $(0, \infty)$, hence also strictly increasing, since $f(x) = f(x_0) < f(x_1) < \cdots < f(x_n) = f(y)$.

We next show continuity, proceeding indirectly. From above, $f : (0, \infty) \to (0, \infty)$ has at most countably many discontinuities, and they are all jump discontinuities. Let $f(x^+) := \lim_{y \to x^+} f(y)$, for $x > 0$. Then $f(x^+) \geq f(x)\ \forall x$, and $f(x^+)$ coincides with $f(x)$ at all points of right continuity and has the same jumps as f. Thus, it suffices to show that $f(x^+)$ is continuous (since this implies f is also continuous).

Now given $0 < x < y < \infty$, apply $f[-]$ to the matrices

$$M(x, y, \epsilon) := \begin{pmatrix} x + \epsilon & \sqrt{xy} + \epsilon \\ \sqrt{xy} + \epsilon & y + \epsilon \end{pmatrix}, \qquad x, y, \epsilon > 0,$$

where $\epsilon > 0$ is small enough that $(x + \epsilon)(y + \epsilon) < p(\sqrt{xy} + \epsilon)^2$. Then an easy verification shows that $M(x, y, \epsilon)$ is in the given test set. It follows that $\det f[M(x, y, \epsilon)] > 0$, i.e., $f(x + \epsilon) f(y + \epsilon) > f(\sqrt{xy} + \epsilon)^2$. Taking $\epsilon \to 0^+$, we obtain

$$f(x^+) f(y^+) \geq f(\sqrt{xy}^+)^2, \qquad \forall x, y > 0.$$

Thus, $f(x^+)$ is positive, non-decreasing and multiplicatively mid-convex on $(0, \infty)$. From the proof of Theorem 12.7(2), we conclude that $f(x^+)$ is continuous on $(0, \infty)$, so $f(x) = f(x^+)$ is also continuous and multiplicatively mid-convex on $(0, \infty)$. $\square$

Using this lemma, we now show:

Proof of Theorem 12.13. We first show the result for $m = n = 2$. By Lemma 12.14, f is continuous, positive, and strictly increasing on $(0, \infty)$. Now claim that $f(x) = f(1)x^\alpha$ for all $x > 0$ (and some $\alpha > 0$). For this, consider the matrices

$$A(x, y, \epsilon) := \begin{pmatrix} x & xy \\ 1 - \epsilon & y \end{pmatrix}, \qquad B(x, y, \epsilon) := \begin{pmatrix} xy & y \\ x & 1 + \epsilon \end{pmatrix}, \qquad \text{where } x, y, \epsilon > 0.$$

These are both TP matrices, hence so are $f[A(x,y,\epsilon)]$ and $f[B(x,y,\epsilon)]$. The positivity of both determinants yields:

$$f(x)f(y) > f(xy)f(1-\epsilon), \qquad f(xy)f(1+\epsilon) > f(x)f(y), \qquad \forall x,y,\epsilon > 0.$$

Taking $\epsilon \to 0^+$, the continuity of f and the assumptions imply that $\frac{f(x)}{f(1)}$ is multiplicative, continuous, positive, non-constant, and strictly increasing on $(0,\infty)$. Hence (e.g. as in the proof of Proposition 11.6), $f(x) = f(1)x^\alpha$ for all $x > 0$, where $\alpha > 0$ and $f(1) > 0$.

This completes the proof for $m = n = 2$. Now suppose more generally that $m, n \geq 2$. Recall by a TP completion problem (see Theorem 7.1) that every 2×2 TP matrix can be completed to an $m \times n$ TP matrix. It follows from the assumptions that $f[-]$ must preserve the 2×2 TP matrices, and we are done. ☐

12.4. **Symmetric TN and TP matrix preservers.** Having classified the preservers of total positivity on all matrices of a fixed size, we turn to TN symmetric matrices:

Theorem 12.15. *Suppose $f : [0,\infty) \to \mathbb{R}$ and $d \geq 1$. Then $f[-]$ preserves the symmetric TN $d \times d$ matrices if and only if f is a non-negative constant, or:*

1. *($d = 1$). The function f is non-negative.*
2. *($d = 2$). f is non-negative, non-decreasing, and multiplicatively mid-convex on $[0,\infty)$. In particular, f is continuous on $(0,\infty)$.*
3. *($d = 3$). $f(x) = cx^\alpha$ for some $c > 0$ and $\alpha \geq 1$.*
4. *($d = 4$). $f(x) = cx^\alpha$ for some $c > 0$ and $\alpha \in \{1\} \cup [2,\infty)$.*
5. *($d = 5$). $f(x) = cx$ for some $c > 0$.*

Proof. For $d = 1$ the result is immediate. If $d = 2$, the result follows from Theorem 12.7. Now suppose $d = 3$. One implication follows from Theorem 12.11. Conversely, every symmetric 2×2 can be padded by a row and column of zeros to remain TN, so by Theorem 12.7, f is continuous, non-negative, and non-decreasing on $(0,\infty)$. We next show that f is continuous at 0 and that $f(0) = 0$ for non-constant f. First, the matrix $f[x\,\mathrm{Id}_{3\times 3}]$ is TN for each $x > 0$. If $f(0) > 0$, considering various 2×2 minors yields $f(x) = f(0)$ for all $x > 0$, so f is constant. The remaining case is $f(0) = 0$. Now let $A_{3\times 3}$ be as in Equation (11.3). By the hypotheses, $f[xA]$ is TN for all $x > 0$, so

$$0 \leq \det f[xA] = -f(0^+)^3.$$

Thus $f(0^+) = 0 = f(0)$, and f is continuous.

Next, consider the symmetric TN matrices

$$A'(x,y) := \begin{pmatrix} x^2 & x & xy \\ x & 1 & y \\ xy & y & y^2 \end{pmatrix}, \qquad B'(x,y) := \begin{pmatrix} x^2y & xy & x \\ xy & y & 1 \\ x & 1 & 1/y \end{pmatrix}, \qquad x \geq 0,\ y > 0.$$

Since these contain $\begin{pmatrix} x & xy \\ 1 & y \end{pmatrix}$ and $\begin{pmatrix} xy & x \\ y & 1 \end{pmatrix}$ as non-principal submatrices, we can repeat the proof of Proposition 11.6 to conclude that f is either a constant or $f(x) = cx^\alpha$ for some $c > 0, \alpha \geq 0$. Finally, again using the matrix A in (11.3) and the computations following it, we conclude that $\alpha \geq 1$.

The next case is $d = 4$. As above, embedding 3×3 matrices via padding by zeros shows that $f(x) = cx^\alpha$, with $c > 0$ and $\alpha \geq 1$. From the proof of Theorem 9.3, given $\alpha \in (1,2)$ one obtains a 4×4 Hankel moment matrix (hence this is TN), corresponding to the measure $\delta_1 + \epsilon\delta_x$ for $1 \neq x \in (0,\infty)$ and small $\epsilon > 0$, whose αth entrywise power is not positive semidefinite. This proves one implication; for the reverse, the preceding parts for $d = 2, 3$ imply that the determinants of all proper submatrices of $f[A]$ are non-negative, for every 4×4 symmetric TN matrix A. That $\det f[A] \geq 0$ follows from Theorem 9.3.

The final case is $d \geq 5$. In this case, one implication is trivial, and the reverse implication for $d = 5$ implies the same for all $d > 5$, by padding 5×5 TN matrices by zeros. Thus, it suffices to classify the non-constant preservers of 5×5 symmetric TN matrices. By the preceding part, these are of the form cx^α for $\alpha = 1$ or $\alpha \geq 2$, and $c > 0$. Now suppose $\alpha \geq 2$, and consider the family of 5×5 matrices

$$T(x) := \mathbf{1}_{5\times 5} + x \begin{pmatrix} 2 & 3 & 6 & 14 & 36 \\ 3 & 6 & 14 & 36 & 98 \\ 6 & 14 & 36 & 98 & 276 \\ 14 & 36 & 98 & 284 & 842 \\ 36 & 98 & 276 & 842 & 2604 \end{pmatrix}, \qquad x > 0. \tag{12.16}$$

Straightforward computations show that all $k \times k$ minors of $M(x)$ are of the form $ax^{k-1} + bx^k$ for $a, b \geq 0$, for $1 \leq k \leq 4$, and $\det M(x) = 0$. Thus $M(x)$ is TN for all $x > 0$. Let $N(x) := M(x)^{(1)}$ be the truncation of $M(x)$, i.e., with its first row and last column removed. Another computation reveals that for small $x > 0$,

$$\det N(x)^{\circ\alpha} = 28584(\alpha^3 - \alpha^4)x^4 + O(x^5),$$

so if $\alpha > 1$, then there exists small $x > 0$ such that $M(x)^{\circ\alpha}$ is not TN. $\square$

From this result, it is possible to deduce the classification of TP preservers on symmetric matrices of each fixed size:

Corollary 12.17. *Suppose $f : (0, \infty) \to (0, \infty)$. Then $f[-]$ preserves total positivity on symmetric TP $d \times d$ matrices, if and only if f satisfies:*

(1) $(d = 1)$. *The function f is positive.*
(2) $(d = 2)$. *f is positive, increasing, and multiplicatively mid-convex on $(0, \infty)$. In particular, f is continuous.*
(3) $(d = 3)$. *$f(x) = cx^\alpha$ for some $c > 0$ and $\alpha \geq 1$.*
(4) $(d = 4)$. *$f(x) = cx^\alpha$ for some $c > 0$ and $\alpha \in \{1\} \cup [2, \infty)$.*
(5) $(d = 5)$. *$f(x) = cx$ for some $c > 0$.*

Proof. The equivalence is obvious for $d = 1$, and was shown for $d = 2$ in Lemma 12.14. Now suppose $d \geq 3$ and A is any symmetric TP 2×2 matrix. By Theorem 7.4, A extends to a symmetric TP $d \times d$ matrix, hence $f[A]$ is TP. The $d = 2$ case now implies that f is continuous, increasing, and positive on $(0, \infty)$, and hence extends to a continuous, non-negative, increasing function $\widetilde{f} : [0, \infty) \to [0, \infty)$. By Whitney density for symmetric matrices (Proposition 6.14), $\widetilde{f}[-]$ preserves symmetric TN 2×2 matrices, so f is of the desired form for each $d \geq 3$ by Theorem 12.15. Conversely, the $d = 3$ case follows from Theorem 10.1; the $d \geq 5$ case is obvious; and for $d = 4$, given $A_{4\times 4}$ symmetric TP and $\alpha \geq 2$, note that all 3×3 submatrices of $A^{\circ\alpha}$ are TP by Theorem 10.1, while $\det A^{\circ\alpha} > 0$ by Corollary 9.11. $\square$

12.5. **Totally non-negative Hankel matrices – entrywise preservers.** We have seen that if the entrywise map $f[-]$ preserves the $m \times n$ TP/TN matrices for $m, n \geq 4$, then f is either constant on $(0, \infty)$ (and $f(0)$ equals either this constant or zero) or $f(x) = f(1)x$ for all x. In contrast, the powers x^α that entrywise preserve positive semidefiniteness on $\mathbb{P}_n((0, \infty))$ (for fixed $n \geq 2$) are $\mathbb{Z}^{\geq 0} \cup [n - 2, \infty)$.

This discrepancy is also supported by the fact that $\mathbb{P}_n$ is closed under the Schur (or entrywise) product, but already the 3×3 TN matrices are not. (Hence neither are the $m \times n$ TN or TP matrices for $m, n \geq 3$, by using completions and density arguments.) For example, $A := \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}$, $B := A^T = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$ are both TN, but $A \circ B = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}$ has determinant -1, and hence cannot be TN.

Thus, a more refined (albeit technical) question would be to isolate and work with a class of TN matrices that is a closed, convex cone, and which is further closed under the Schur product. In fact, such a class has already been discussed earlier: the family of *Hankel TN* matrices (see Corollary 4.2). With those results in mind, and for future use, we introduce the following notation:

Definition 12.18. Given an integer $n \geq 1$, let $\mathtt{HTN}_n$ denote the $n \times n$ Hankel TN matrices.

We also study in this text the entrywise preservers of TN on $\mathtt{HTN}_n$ for a fixed n and for all n – see Remark 9.9, Corollary 44.11, and Corollary 19.4. This study turns out to be remarkably similar (and related) to the study of positivity preservers on $\mathbb{P}_n$ – which is not surprising, given Theorem 4.1. For now, we work in the setting under current consideration: entrywise power-preservers.

Theorem 12.19. *For $n \geq 2$ and $\alpha \in \mathbb{R}$, x^α entrywise preserves TN on $\mathtt{HTN}_n$, if and only if $\alpha \in \mathbb{Z}^{\geq 0} \cup [n-2, \infty)$.*

In other words, x^α preserves total non-negativity on $\mathtt{HTN}_n$ if and only if it preserves positive semidefiniteness on $\mathbb{P}_n([0, \infty))$.

Proof. If $\alpha \in \mathbb{Z}^{\geq 0} \cup [n-2, \infty)$, then we use Theorem 4.1 together with Theorem 9.3. Conversely, suppose $\alpha \in (0, n-2) \setminus \mathbb{Z}$. We have previously shown that the moment matrix $H := (1 + \epsilon x^{j+k-2})_{j,k=1}^n$ lies in $\mathtt{HTN}_n$ for $x, \epsilon > 0$; but if $x \neq 1$ and $\epsilon > 0$ is small, then $H^{\circ \alpha} \notin \mathbb{P}_n$, as shown in the proof of Theorem 9.3. (Alternately, this holds for all $\epsilon > 0$ by Theorem 9.10.) It follows that $H^{\circ \alpha} \notin TN_n$. $\square$

13. Entrywise powers (and functions) preserving positivity: II. Matrices with zero patterns.

Having completed the classification of entrywise functions preserving the TP/TN matrices in any fixed size, in this part of the text and the next we restrict ourselves to understanding the entrywise functions preserving positive semidefiniteness – henceforth termed *positivity* – either in a fixed dimension or in all dimensions. (As mentioned above, there will be minor detours studying the related notion of entrywise preservers of $\mathtt{HTN}_n$.)

In this section and the next, we continue to study entrywise powers preserving positivity in a fixed dimension, by refining the test set of positive semidefinite matrices. The plan for these two sections is as follows:

(1) We begin by recalling the test set $\mathbb{P}_G([0,\infty))$ associated to any graph G, and discussing some of the modern-day motivations in studying entrywise functions (including powers) that preserve positivity.
(2) We then prove some results on general entrywise functions preserving positivity on $\mathbb{P}_G$ for arbitrary non-complete graphs. (The case of complete graphs is the subject of the remainder of the text.) As a consequence, the powers – in fact, the *functions* – preserving $\mathbb{P}_G([0,\infty))$ where G is any tree (or collection of trees) are completely classified.
(3) We show how the integration trick of FitzGerald and Horn (see the discussion around Equations (9.6) and (9.7)) extends to help classify the entrywise powers preserving other Loewner properties, including monotonicity, and in turn, super-additivity.
(4) Using these results, we classify the powers preserving $\mathbb{P}_G$ for G the *almost complete graph* (i.e., the complete graph minus any one edge).
(5) We then state some recent results on powers preserving $\mathbb{P}_G$ for other G (all chordal graphs; cycles), and conclude with some questions for general graphs G, which arise naturally from these results.

13.1. **Modern-day motivations: graphical models and high-dimensional covariance estimation.** As we discuss in Section 16, the question of which functions preserve positivity when applied entrywise has a long history, having been studied for the best part of a century within the analysis literature. For now, we explain why this question has attracted renewed attention owing to its importance in high-dimensional covariance estimation.

In modern-day scientific applications, one of the most important challenges involves understanding complex multivariate structures and dependencies. Such questions naturally arise in various domains: understanding the interactions of financial instruments, studying markers of climate parameters to understand climate patterns, and modeling gene-gene associations in cancer and cardiovascular disease, to name a few. In such applications, one works with very large random vectors $X \in \mathbb{R}^p$, and a fundamental measure of dependency that is commonly used (given a sample of vectors) is the covariance matrix (or correlation matrix) and its inverse. Unlike traditional regimes, where the sample size n far exceeds the dimension of the problem p (i.e., the number of random variables in the model), these modern applications – among others – involve the reverse situation: $n \ll p$. This is due to the high cost of making, storing, and working with observations, for instance; but moreover, an immediate consequence is that the corresponding covariance matrix built out of the samples $x_1, \dots, x_n \in \mathbb{R}^p$

$$\widehat{\Sigma} := \frac{1}{n-1} \sum_{j=1}^n (x_j - \overline{x})(x_j - \overline{x})^T,$$

is highly singular. (Its rank is bounded above by the sample size $n \ll p$.) This makes $\widehat{\Sigma}$ a poor estimator of the true underlying covariance matrix Σ.

A second shortcoming of the sample covariance matrix has to do with zero patterns. In the underlying model, there is often additional domain-specific knowledge which leads to *sparsity*. In other words, certain pairs of variables are known to be independent, or conditionally independent given other variables. For instance, in probability theory one has the notion of a *Markov random field*, or graphical model, in which the nodes of a graph represent random variables and the edges the dependency structure between them. Or, in the aforementioned climate-related application – specifically, temperature reconstruction – the temperature at one location is assumed to not influence that at another (perhaps far away) location, at least when conditioned on the neighboring points. Such (conditional) independences are reflected in zero entries in the associated (inverse) covariance matrix. In fact, in the aforementioned applications, several models assume most of the entries ($\sim 90\%$ or more) to be zero.

However, in the observed sample covariance matrix, there is almost always some noise, as a result of which very few entries are zero. This is another reason why sample covariance is a poor estimator in modern applications.

For such reasons, it is common for statistical practitioners to *regularize* the sample covariance matrix (or other estimators), in order to improve its properties for a given application. Popular state-of-the-art methods involve inducing sparsity – i.e., zero entries – via convex optimization techniques that impose an ℓ^1-penalty (since ℓ^0-penalties are not amenable to such techniques). While these methods induce sparsity and are statistically consistent as $n, p \to \infty$, they are iterative and hence require solving computationally expensive optimization problems. In particular, they are not scalable to ultra high-dimensional data, say for $p \sim 100,000$ or $500,000$, as one often encounters in the aforementioned situations.

A recent promising alternative is to apply entrywise functions on the entries of sample covariance matrices (see, e.g., [15, 48, 110, 167, 168, 236, 304, 377] and numerous follow-up papers). For example, the hard and soft thresholding functions set very small entries to zero (operating under the assumption that these often come from noise, and do not represent the most important associations). Another popular family of functions used in applications consists of entrywise powers. Indeed, powering up the entries provides an effective way in applications to separate signal from noise.

Note that these entrywise operations do not suffer from the same drawback of scalability, since they operate directly on the entries of the matrix, and do not involve optimization-based techniques. The key question now, is to understand *when such entrywise operations preserve positive semidefiniteness*. Indeed, the regularized matrix that these operations yield must serve as a proxy for the sample covariance matrix in further statistical analyses, and hence is required to be positive semidefinite.

It is thus crucial to understand when these entrywise operations preserve positivity – and in a fixed dimension, since in a given application one knows the dimension of the problem. Note that while the motivation here comes from downstream applications, the heart of the issue is very much a mathematical question involving analysis on the cone $\mathbb{P}_n$.

With these motivations, the current and last/next few sections deal with entrywise powers preserving positivity in a fixed dimension; progress on these questions impacts applied fields. At the same time, the question of when entrywise powers and functions preserve positivity, has been studied in the mathematics literature for almost a century. Thus (looking slightly ahead), in the next part and the last part of this text, we return to the mathematical advances, both classical and recent. This includes proving some of the celebrated characterization results in this area – by Schoenberg, Rudin, Loewner/Horn, and Vasudeva – using fairly accessible mathematical machinery.

13.2. **Entrywise functions preserving positivity on $\mathbb{P}_G$ for non-complete graphs.** In this section and the next, we continue with the theme of entrywise powers and functions preserving positivity in a fixed dimension, now under additional sparsity constraints – i.e., on $\mathbb{P}_G$ for a fixed graph G. In this section, we obtain certain necessary conditions on general functions preserving positivity on $\mathbb{P}_G$.

As we will see in Section 16.1, the functions preserving positive semidefiniteness on $\mathbb{P}_n$ for all n (and those preserving TN on $\mathtt{HTN}_n$) for all integers $n \geq 1$ can be classified, and they are precisely the power series with non-negative coefficients

$$f(x) = \sum_{k=0}^{\infty} c_k x^k, \qquad \text{with } c_k \geq 0 \ \forall \ k.$$

This is a celebrated result of Schoenberg and Rudin (see Theorems 16.2 and 16.3). However, the situation is markedly different for entrywise preservers of $\mathbb{P}_n$ for a *fixed* dimension $n \geq 1$:

- For $n = 1$, clearly any $f : [0, \infty) \to [0, \infty)$ works.
- For $n = 2$, the entrywise preservers of positive semidefiniteness (or of total non-negativity) on $\mathbb{P}_2((0, \infty))$ have been classified by Vasudeva in *Ind. J. Pure Appl. Math.* (1979): see Theorem 12.7.
- For $n \geq 3$, the problem remains open to date.

Given the open (and challenging!) nature of the problem in fixed dimension, efforts along this direction have tended to work on refinements of the problem: either restricting the class of entrywise functions (to, e.g., power functions, or polynomials as we study later), or restricting the class of matrices: to TP/TN matrices, to Toeplitz matrices (by Rudin), or Hankel TN matrices, or to matrices with rank bounded above (by Schoenberg, Rudin, Loewner and Horn, and subsequent authors), or to matrices with a given sparsity pattern – i.e., $\mathbb{P}_G$ for fixed G. It is this last approach that we focus on, in this section and Section 14.

Given a (finite simple) graph $G = (V, E)$, with $V = [n] = \{1, \dots, n\}$ for some $n \geq 1$, and a subset $0 \in I \subset \mathbb{R}$, the subset $\mathbb{P}_G(I)$ is defined to be:

$$\mathbb{P}_G(I) := \{A \in \mathbb{P}_n(I) \ : \ a_{jk} = 0 \text{ if } j \neq k \text{ and } (j,k) \notin E\}. \tag{13.1}$$

For example, when $G = A_3$ (the path graph on three nodes), $\mathbb{P}_G = \left\{ \begin{pmatrix} a & b & e \\ d & b & 0 \\ e & 0 & c \end{pmatrix} \in \mathbb{P}_3 \right\}$, and when $G = K_n$ (the complete graph on n vertices), we have $\mathbb{P}_G(I) = \mathbb{P}_n(I)$.

We now study the entrywise preservers of $\mathbb{P}_G$ for a graph G. To begin, we extend the notion of entrywise functions to $\mathbb{P}_G$, by acting only on the "unconstrained" entries:

Definition 13.2. Let $0 \in I \subset \mathbb{R}$. Given a graph G with vertex set $[n]$, and $f : I \to \mathbb{R}$, define $f_G[-] : \mathbb{P}_G(I) \to \mathbb{R}^{n \times n}$ via

$$(f_G[A])_{jk} := \begin{cases} 0, & \text{if } j \neq k, \ (j,k) \notin E, \\ f(a_{jk}), & \text{otherwise.} \end{cases}$$

Here are some straightforward observations on entrywise preservers of $\mathbb{P}_G([0, \infty))$:

(1) When G is the empty graph, i.e., $G = (V, \emptyset)$, the functions f, such that $f_G[-]$ preserves $\mathbb{P}_G$ are precisely the functions sending $[0, \infty)$ to itself.

(2) When G is the disjoint union of a positive number of disconnected copies of K_2 and isolated nodes, $\mathbb{P}_G$ consists of block diagonal matrices of the form $\oplus_{j=1}^k A_j$, where the A_j are either 2×2 or 1×1 matrices (blocks), corresponding to copies of K_2 or isolated points respectively, and $\oplus$ denotes a block diagonal matrix of the form

$$\begin{pmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_k \end{pmatrix}.$$

(The remaining entries are zero.) By assumption, at least one of the A_j must be a 2×2 block. For such graphs, we conclude by Theorem 12.7 that $f_G[-] : \mathbb{P}_G([0,\infty)) \to \mathbb{P}_G([0,\infty))$ if and only if f is non-negative, non-decreasing, multiplicatively mid-convex, and $0 \leq f(0) \leq \lim_{x\to 0^+} f(x)$.

(3) More generally, if G is a disconnected union of graphs: $G = \bigsqcup_{j\in J} G_j$, then $f_G[-] : \mathbb{P}_G([0,\infty)) \to \mathbb{P}_G([0,\infty))$ if and only if the entrywise map $f_{G_j}[-]$ preserves $\mathbb{P}_{G_j}([0,\infty))$ for all j.

In light of these examples, we shall henceforth consider only connected, non-complete graphs G, and the functions f, such that $f_G[-]$ preserves $\mathbb{P}_G([0,\infty))$. We begin with the following necessary conditions:

Proposition 13.3. *Let $I = [0,\infty)$ and G be a connected, non-complete graph. Suppose $f : I \to \mathbb{R}$ is such that $f_G[-] : \mathbb{P}_G(I) \to \mathbb{P}_G(I)$. Then the following statements hold:*

(1) $f(0) = 0$.
(2) f is continuous on I (and not just on $(0,\infty)$).
(3) f is super-additive on I, i.e., $f(x+y) \geq f(x) + f(y)\ \forall\, x,y \geq 0$.

Remark 13.4. In particular, $f_G[-] = f[-]$ for (non-)complete graphs G. Thus, following the proof of Proposition 13.3, we use $f[-]$ in the sequel.

Proof. Clearly, $f : I \to I$. Assume that G has at least three nodes, since for connected graphs with two nodes, the proposition is vacuous. A small observation – made by Horn [182], if not earlier – reveals that there exist three nodes, which we may relabel as $1, 2$, and 3 without loss of generality, such that 2 and 3 are adjacent to 1 but not to each other. Since $\mathbb{P}_2(I) \hookrightarrow \mathbb{P}_G(I)$ via

$$\begin{pmatrix} a & b \\ b & c \end{pmatrix} \mapsto \begin{pmatrix} a & b \\ b & c \end{pmatrix} \oplus 0_{(|V|-2)\times(|V|-2)},$$

it follows from Theorem 12.7 that $f|_{(0,\infty)}$ is non-negative, non-decreasing, and multiplicatively mid-convex; moreover, $f|_{(0,\infty)}$ is continuous and is identically zero or never zero.

To prove (1), define

$$B(\alpha,\beta) := \begin{pmatrix} \alpha+\beta & \alpha & \beta \\ \alpha & \alpha & 0 \\ \beta & 0 & \beta \end{pmatrix}, \qquad \alpha,\beta \geq 0.$$

Note that $B(\alpha,\beta) \oplus 0_{(|V|-3)\times(|V|-3)} \in \mathbb{P}_G(I)$. Hence, $f_G[B(\alpha,\beta)\oplus 0] \in \mathbb{P}_G(I)$, from which we obtain:

$$f_G[B(\alpha,\beta)] = \begin{pmatrix} f(\alpha+\beta) & f(\alpha) & f(\beta) \\ f(\alpha) & f(\alpha) & 0 \\ f(\beta) & 0 & f(\beta) \end{pmatrix} \in \mathbb{P}_3(I), \qquad \forall \alpha,\beta \geq 0. \tag{13.5}$$

For $\alpha = \beta = 0$, (13.5) yields that $\det f_G[B(0,0)] = -f(0)^3 \geq 0$. But since f is non-negative, it follows that $f(0) = 0$, proving (1).

Now if $f|_{(0,\infty)} \equiv 0$, the remaining assertions immediately follow. Thus, we assume in the sequel that $f|_{(0,\infty)}$ is always positive.

To prove (2), let $\alpha = \beta > 0$. Then (13.5) gives:

$$\det f_G[B(\alpha,\alpha)] \geq 0 \quad \implies \quad f(\alpha)^2(f(2\alpha) - 2f(\alpha)) \geq 0 \quad \implies \quad f(2\alpha) - 2f(\alpha) \geq 0.$$

Taking the limit as $t \to 0^+$, we obtain $-f(0^+) \geq 0$. Since f is non-negative, $f(0^+) = 0 = f(0)$, so f is continuous at 0. The continuity of f on I now follows from the above discussion.

Finally, to prove (3), let $\alpha, \beta > 0$. Invoking (13.5) and again starting with $\det f_G[B(\alpha,\beta)] \geq 0$, we obtain

$$f(\alpha)f(\beta)(f(\alpha+\beta) - f(\alpha) - f(\beta)) \geq 0 \quad \implies \quad f(\alpha+\beta) \geq f(\alpha) + f(\beta).$$

This shows that f is super-additive on $(0,\infty)$; since $f(0) = 0$, we obtain super-additivity on all of I. $\square$

Proposition 13.3 is the key step in classifying *all entrywise functions* preserving positivity on $\mathbb{P}_G$ for every tree G. In fact, apart from the case of $\mathbb{P}_2 = \mathbb{P}_{K_2}$, this is perhaps the only known case (i.e., family of individual graphs) for which a complete classification of the entrywise preservers of $\mathbb{P}_G$ is available – and proved in the next result.

Recall that a tree is a connected graph in which there is a unique path between any two vertices; equivalently, where the number of edges is one less than the number of nodes; or also where there are no cycle subgraphs. For example, the graph A_3 considered above (with $V = \{1,2,3\}$ and $E = \{(1,2),(1,3)\}$) is a tree.

Theorem 13.6. *Suppose $I = [0,\infty)$ and a function $f : I \to I$. Let G be a tree on at least three vertices. Then the following are equivalent:*

1. *$f[-] : \mathbb{P}_G(I) \to \mathbb{P}_G(I)$.*
2. *$f[-] : \mathbb{P}_T(I) \to \mathbb{P}_T(I)$ for all trees T.*
3. *$f[-] : \mathbb{P}_{A_3}(I) \to \mathbb{P}_{A_3}(I)$.*
4. *f is multiplicatively mid-convex and super-additive on I.*

Proof. Note that G contains three vertices on which the induced subgraph is A_3 (consider any induced connected subgraph on three vertices). By padding $\mathbb{P}_{A_3}$ by zeros to embed inside $\mathbb{P}_{|G|}$, we obtain (1) $\implies$ (3). Moreover, that (2) $\implies$ (1) is clear.

To prove that (3) $\implies$ (4), note that $K_2 \hookrightarrow A_3$. Hence, f is multiplicatively mid-convex on $(0,\infty)$ by Theorem 12.7. By Proposition 13.3, $f(0) = 0$ and f is super-additive on I. In particular, f is also multiplicatively mid-convex on all of I.

Finally, we show that (4) $\implies$ (2) by induction on n for all trees T with at least $n \geq 2$ nodes. For the case $n = 2$ by Theorem 12.7, it suffices to show that f is non-decreasing. Given $\gamma \geq \alpha \geq 0$, by super-additivity we have

$$f(\gamma) \geq f(\alpha) + f(\gamma - \alpha) \geq f(\alpha),$$

proving the result.

For the induction step, suppose that (2) holds for all trees on n nodes and let $G' = (V,E)$ be a tree on $n+1$ nodes. Without loss of generality, let $V = [n+1] = \{1,\dots,n+1\}$, such that node $n+1$ is adjacent only to node n. (Note: there always exists such a node in every tree.) Let G be the induced subgraph on the subset $[n]$ of vertices. Then, any $A \in \mathbb{P}_{G'}(I)$

can be written as

$$A = \begin{pmatrix} B & b\mathbf{e}_n \\ b\mathbf{e}_n^T & c \end{pmatrix}_{(n+1)\times(n+1)}, \quad \text{where } b \in \mathbb{R},\ c \in I,\ B \in \mathbb{P}_G(I),$$

and $\mathbf{e}_n := ((0,0,\dots,0,1)_{1\times n})^T$ is a standard basis vector. Since f is non-negative and super-additive, $f(0) = f(0+0) \geq 2f(0) \geq 0$, hence $f(0) = 0$. If $f \equiv 0$, we are done. Thus, we assume that $f \not\equiv 0$, so $f|_{(0,\infty)}$ is positive by Theorem 12.7.

If $c = 0$, then $b_{nn}c - b^2 \geq 0$ implies $b = 0$, and so $f[A] = \begin{pmatrix} f[B] & 0_{n\times 1} \\ 0_{1\times n} & 0 \end{pmatrix} \in \mathbb{P}_G \oplus 0_{1\times 1}$ by the induction hypothesis. Otherwise, $c > 0$, hence $f(c) > 0$. From the properties of Schur complements (Theorem 2.32) we obtain that A is positive semidefinite if and only if $B - \frac{b^2}{c}E_{nn}$ is positive semidefinite, where E_{nn} is the elementary $n \times n$ matrix with (j,k) entry $\delta_{j,n}\delta_{k,n}$; and similarly, $f[A]$ is positive semidefinite if and only if $f[B] - \frac{f(b)^2}{f(c)}E_{nn}$ is positive semidefinite.

By the induction hypothesis, we have that $f[B - \frac{b^2}{c}E_{nn}]$ is positive semidefinite. Thus, it suffices to prove that $f[B] - \frac{f(b)^2}{f(c)}E_{nn} - f[B - \frac{b^2}{c}E_{nn}]$ is positive semidefinite. Now compute:

$$f[B] - \frac{f(b)^2}{f(c)}E_{nn} - f[B - \frac{b^2}{c}E_{nn}] = \alpha E_{nn}, \quad \text{where } \alpha = f(b_{nn}) - \frac{f(b)^2}{f(c)} - f(b_{nn} - \frac{b^2}{c}).$$

Therefore, it suffices to show that $\alpha \geq 0$. But by super-additivity, we have

$$\begin{aligned} \alpha &= f(b_{nn}) - \frac{f(b)^2}{f(c)} - f(b_{nn} - \frac{b^2}{c}) \\ &= f(b_{nn} - \frac{b^2}{c} + \frac{b^2}{c}) - \frac{f(b)^2}{f(c)} - f(b_{nn} - \frac{b^2}{c}) \\ &\geq f(b_{nn} - \frac{b^2}{c}) + f(\frac{b^2}{c}) - \frac{f(b)^2}{f(c)} - f(b_{nn} - \frac{b^2}{c}) \\ &= f(\frac{b^2}{c}) - \frac{f(b)^2}{f(c)}. \end{aligned}$$

Moreover, by multiplicative mid-convexity, we obtain that $f(\frac{b^2}{c})f(c) \geq f(b)^2$. Hence, $\alpha \geq 0$ and $f[A]$ is positive semidefinite, as desired. $\square$

An immediate consequence is the complete classification of entrywise powers preserving positivity on $\mathbb{P}_T([0,\infty))$ for T a tree.

Corollary 13.7. *$f(x) = x^\alpha$ preserves $\mathbb{P}_T([0,\infty))$ for a tree on at least three nodes, if and only if $\alpha \geq 1$.*

The proof follows from the observation that x^α is super-additive on $[0,\infty)$ if and only if $\alpha \geq 1$.

14. Entrywise powers preserving positivity: III. Chordal graphs. Loewner monotonicity and super-additivity.

We next study the set of entrywise powers preserving positivity on matrices with zero patterns. Recall the closed convex cone $\mathbb{P}_G([0,\infty))$ studied in Section 13, for a (finite simple) graph G. Now define

$$\mathcal{H}_G := \{\alpha \geq 0 : A^{\circ\alpha} \in \mathbb{P}_G([0,\infty)) \ \forall A \in \mathbb{P}_G([0,\infty))\}, \tag{14.1}$$

with the convention that $0^0 := 1$. Thus, $\mathcal{H}_G$ is the set of entrywise, or Hadamard, powers that preserve positivity on $\mathbb{P}_G$.

Observe that if $G \subset H$ are graphs, then $\mathcal{H}_G \supset \mathcal{H}_H$. In particular, by the FitzGerald–Horn classification in Theorem 9.3,

$$\mathcal{H}_G \supset \mathcal{H}_{K_n} = \mathbb{Z}^{\geq 0} \cup [n-2,\infty) \tag{14.2}$$

whenever G has n vertices. Specifically, there is always a point $\beta \geq 0$ beyond which every real power preserves positivity on $\mathbb{P}_G$. We are interested in the smallest such point, which leads us to the next definition (following the FitzGerald–Horn theorem 9.3 in the special case $G = K_n$):

Definition 14.3. The *critical exponent of a graph G* is

$$\alpha_G := \min\{\beta \geq 0 : \alpha \in \mathcal{H}_G \ \forall \alpha \geq \beta\}.$$

Example 14.4. We saw earlier that if G is a tree (but not a disjoint union of copies of K_2), then $\alpha_G = 1$; and FitzGerald–Horn [123] showed that $\alpha_{K_n} = n-2$ for all $n \geq 2$.

In this section we are interested in closed-form expressions for α_G and $\mathcal{H}_G$. Not only is this a natural mathematical refinement of Theorem 9.3, but as discussed in Section 13.1, this moreover impacts applied fields, providing modern motivation to study the question. Somewhat remarkably, the above examples were the only known cases until very recently.

On a more mathematical note, we are also interested in understanding a combinatorial interpretation of the critical exponent α_G. This is a graph invariant that arises out of positivity; it is natural to ask if it is related to previously known (combinatorial) graph invariants, and more broadly, how it relates to the geometry of the graph.

We explain in this section that there is a uniform answer for a large family of graphs, which includes complete graphs, trees, split graphs, banded graphs, cycles, and other classes; and moreover, there are no known counterexamples to this answer. Before stating the results, we remark that the question of computing $\mathcal{H}_G, \alpha_G$ for a given graph is easy to formulate, and one can carry out easy numerical simulations by running (software code) over large sets of matrices in $\mathbb{P}_G$ (possibly chosen randomly), to better understand which powers preserve $\mathbb{P}_G$. This naturally leads to accessible research problems for various classes of graphs: say triangle-free graphs, or graphs with small numbers of vertices. For instance, there is a graph on five vertices for which the critical exponent is not known!

Now on to the known results. We begin by computing the critical exponent α_G – and $\mathcal{H}_G$, more generally – for a family of graphs that turns out to be crucial in understanding several other families (split, Apollonian, banded, and in fact all chordal graphs).

Definition 14.5. The *almost complete graph $K_n^{(1)}$* is the complete graph on n nodes, with one edge missing.

We will choose a specific labeling of the nodes in $K_n^{(1)}$; note this does not affect the set $\mathcal{H}_G$ or the threshold α_G. Specifically, we set the $(1,n)$ and $(n,1)$ entries to be zero, so that $\mathbb{P}_{K_n^{(1)}}$ consists of matrices of the form $\begin{pmatrix} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \end{pmatrix} \in \mathbb{P}_n$. Our goal is to prove:

Theorem 14.6. *For all $n \geq 2$, we have $\mathcal{H}_{K_n^{(1)}} = \mathcal{H}_{K_n} = \mathbb{Z}^{\geq 0} \cup [n-2, \infty)$.*

14.1. **Other Loewner properties.** In order to prove Theorem 14.6, we need to understand the powers that preserve super-additivity on $n \times n$ matrices under the positive semidefinite ordering. We now define this notion, as well as a related notion of monotonicity.

Definition 14.7. Let $I \subset \mathbb{R}$ and $n \in \mathbb{N}$. A function $f : I \to \mathbb{R}$ is said to be

(1) *Loewner monotone* on $\mathbb{P}_n(I)$ if we have $A \geq B \geq 0_{n \times n} \implies f[A] \geq f[B]$.
(2) *Loewner super-additive* on $\mathbb{P}_n(I)$ if $f[A+B] \geq f[A] + f[B]$ for all $A, B \in \mathbb{P}_n(I)$.

In these definitions, we are using the *Loewner ordering* (or positive semidefinite ordering) on $n \times n$ matrices: $A \geq B$ if $A - B \in \mathbb{P}_n(\mathbb{R})$.

Remark 14.8. A few comments to clarify these definitions are in order. First, if $n = 1$, then these notions both reduce to their usual counterparts for real functions defined on $[0, \infty)$. Second, if $f(0) \geq 0$, then Loewner monotonicity implies Loewner positivity. Third, a Loewner monotone function differs from – in fact is the entrywise analogue of – the more commonly studied *operator monotone* functions, which have the same property but for the functional calculus: $A \geq B \geq 0 \implies f(A) \geq f(B) \geq 0$.

Note that if $n = 1$ and f is continuously differentiable, then f is non-decreasing if and only if f' is non-negative. The following result generalizes this fact to powers acting entrywise on $\mathbb{P}_n$ and classifies the Loewner monotone powers.

Theorem 14.9 (FitzGerald–Horn)**.** *Given an integer $n \geq 2$ and a scalar $\alpha \in \mathbb{R}$, the power x^α is Loewner monotone on $\mathbb{P}_n([0,\infty))$ if and only if $\alpha \in \mathbb{Z}^{\geq 0} \cup [n-1, \infty)$. In particular, the critical exponent for Loewner monotonicity on $\mathbb{P}_n$ is $n-1$.*

We will see in Section 15 a strengthening of Theorem 14.9 by using individual matrices from a multiparameter family, in the spirit of Jain's theorem 9.10 for Loewner positive powers.

Proof. The proof strategy is similar to that of Theorem 9.3: use the Schur product theorem for non-negative integer powers, perform induction on n for the powers above the critical exponent, and employ (the same) rank-2 Hankel moment matrix counterexample for the remaining powers. First, if $\alpha \in \mathbb{N}$ and $0 \leq B \leq A$, then repeated application of the Schur product theorem yields

$$0_{n \times n} \leq B^{\circ \alpha} \leq B^{\circ(\alpha-1)} \circ A \leq B^{\circ(\alpha-2)} \circ A^{\circ 2} \leq \cdots \leq A^{\circ \alpha}.$$

Now, suppose $\alpha \geq n-1$. We prove that x^α is Loewner monotone on $\mathbb{P}_n$ by induction on n; the base case of $n = 1$ is clear. For the induction step, if $\alpha \geq n-1$, then recall the integration trick (9.7) of FitzGerald and Horn:

$$A^{\circ \alpha} - B^{\circ \alpha} = \alpha \int_0^1 (A - B) \circ (\lambda A + (1-\lambda) B)^{\circ(\alpha-1)} \, d\lambda.$$

Since $\alpha - 1 \geq n-2$, the matrix $(\lambda A + (1-\lambda)B)^{\circ(\alpha-1)}$ is positive semidefinite by Theorem 9.3, and thus, $A^{\circ \alpha} - B^{\circ \alpha} \in \mathbb{P}_n$. Therefore, $A^{\circ \alpha} \geq B^{\circ \alpha}$, and we are done by induction.

Finally, to show that the threshold $\alpha = n-1$ is sharp, suppose $\alpha \in (0, n-1) \setminus \mathbb{N}$ (we leave the case of $\alpha < 0$ as an easy exercise). Consider again the Hankel moment matrices

$$A(\epsilon) := H_\mu \text{ for } \mu = \delta_1 + \epsilon\delta_x, \qquad B := A(0) = \mathbf{1}_{n\times n},$$

where $x, \epsilon > 0,\ x \neq 1$, and H_μ is understood to denote the leading principal $n \times n$ submatrix of the Hankel moment matrix for μ. Clearly, $A(\epsilon) \geq B \geq 0_{n\times n}$. As above, let $v = (1, x, \dots, x^{n-1})^T$, so that $A(\epsilon) = \mathbf{1}\mathbf{1}^T + \epsilon vv^T$. Choose a vector $u \in \mathbb{R}^n$ that is orthogonal to $v, v^{\circ 2}, \dots, v^{\circ(\lfloor\alpha\rfloor+1)}$, and $u^T v^{\circ(\lfloor\alpha\rfloor+2)} = 1$. (Note, this is possible since the vectors $v, v^{\circ 2}, \dots, v^{\circ(\lfloor\alpha\rfloor+2)}$ are linearly independent, forming the columns of a possibly partial generalized Vandermonde matrix.)

We claim that $u^T(A(\epsilon)^{\circ\alpha} - B^{\circ\alpha})u < 0$ for small $\epsilon > 0$, which will show that x^α is not Loewner monotone on $\mathbb{P}_n([0,\infty))$. Indeed, compute using the binomial series for $(1+x)^\alpha$:

$$\begin{aligned}
u^T A(\epsilon)^{\circ\alpha} u - u^T B u &= u^T(\mathbf{1}\mathbf{1}^T + \epsilon vv^T)^{\circ\alpha} u - u^T \mathbf{1}\mathbf{1}^T u \\
&= u^T \cdot \sum_{k=1}^{\lfloor\alpha\rfloor+2} \binom{\alpha}{k} \epsilon^k v^{\circ k} (v^{\circ k})^T \cdot u + u^T \cdot o(\epsilon^{\lfloor\alpha\rfloor+2}) \cdot u \\
&= \binom{\alpha}{\lfloor\alpha\rfloor+2} \epsilon^{\lfloor\alpha\rfloor+2} + u^T o(\epsilon^{\lfloor\alpha\rfloor+2}) u \\
&= \epsilon^{\lfloor\alpha\rfloor+2} \left(\binom{\alpha}{\lfloor\alpha\rfloor+2} + u^T \cdot o(1) \cdot u \right),
\end{aligned}$$

and this is negative for small $\epsilon > 0$. (Here, $o(\cdot)$ always denotes a matrix, as in (9.8).) $\square$

Theorem 14.9 is now used to classify the powers preserving Loewner super-additivity. Note that if $n = 1$, then x^α is super-additive on $\mathbb{P}_n([0,\infty)) = [0,\infty)$ if and only if $\alpha \geq n = 1$. The following result generalizes this to all integers $n \geq 1$:

Theorem 14.10 (Guillot, Khare, and Rajaratnam, [151])**.** *Given an integer $n \geq 1$ and a scalar $\alpha \in \mathbb{R}$, the power x^α is Loewner super-additive on $\mathbb{P}_n([0,\infty))$ if and only if $\alpha \in \mathbb{N} \cup [n,\infty)$. Moreover, for each $\alpha \in (0,n) \setminus \mathbb{N}$ and $x \in (0,1)$, for $\epsilon > 0$ small enough the matrix*

$$(\mathbf{1}\mathbf{1}^T + \epsilon vv^T)^{\circ\alpha} - \mathbf{1}\mathbf{1}^T - (\epsilon vv^T)^{\circ\alpha}$$

is not positive semidefinite, where $v = (1, x, \dots, x^{n-1})^T$. In particular, the critical exponent for Loewner super-additivity on $\mathbb{P}_n$ is n.

Thus, once again the same rank-2 Hankel moment matrices provide the desired counterexamples, for non-integer powers α below the critical exponent.

Proof. As above, we leave the proof of the case $\alpha < 0$ or $n = 1$ to the reader. Next, if $\alpha = 0$, then super-additivity fails, since we always get $-\mathbf{1}\mathbf{1}^T$ from the super-additivity condition, and this is not positive semidefinite.

Thus, henceforth $\alpha > 0$ and $n \geq 2$. If α is an integer, then by the binomial theorem and the Schur product theorem,

$$(A+B)^{\circ\alpha} = \sum_{k=0}^{\alpha} \binom{\alpha}{k} A^{\circ k} \circ B^{\circ(\alpha-k)} \geq A^{\circ\alpha} + B^{\circ\alpha}, \qquad \forall A, B \in \mathbb{P}_n.$$

Next, if $\alpha \geq n$ and $A, B \in \mathbb{P}_n([0,\infty))$, then $x^{\alpha-1}$ preserves Loewner monotonicity on $\mathbb{P}_n$, by Theorem 14.9. Again, use the integration trick (9.7) to compute

$$\begin{aligned}(A+B)^{\circ\alpha} - A^{\circ\alpha} &= \alpha \int_0^1 B \circ (\lambda(A+B) + (1-\lambda)A)^{\circ(\alpha-1)}\, d\lambda \\ &\geq \alpha \int_0^1 B \circ (\lambda B)^{\circ(\alpha-1)}\, d\lambda = B^{\circ\alpha}.\end{aligned}$$

The final case is if $\alpha \in (0,n) \setminus \mathbb{Z}$. As above, we fix $x > 0,\ x \neq 1$, and define

$$v := (1, x, \dots, x^{n-1})^T, \qquad A(\epsilon) := \epsilon v v^T \ (\epsilon > 0), \qquad B := A(0) = \mathbf{1}_{n\times n}.$$

Clearly, $A(\epsilon), B \geq 0_{n\times n}$. Now, since $\alpha \in (0,n)$, the vectors $v, v^{\circ 2}, \dots, v^{\circ\lfloor\alpha\rfloor}, v^{\circ\alpha}$ are linearly independent (since the matrix with these columns is part of a generalized Vandermonde matrix). Thus, we may choose $u \in \mathbb{R}^n$ that is orthogonal to $v, \dots, v^{\circ\lfloor\alpha\rfloor}$ (if $\alpha \in (0,1)$, this is vacuous) and such that $u^T v^{\circ\alpha} = 1$. Now compute as in the previous proof, using the binomial theorem

$$(A(\epsilon)+B)^{\circ\alpha} - A(\epsilon)^{\circ\alpha} - B^{\circ\alpha} = \sum_{k=1}^{\lfloor\alpha\rfloor} \binom{\alpha}{k} \epsilon^k v^{\circ k} (v^{\circ k})^T - \epsilon^\alpha v^{\circ\alpha}(v^{\circ\alpha})^T + o(\epsilon^\alpha);$$

the point here is that the last term shrinks at least as fast as $\epsilon^{\lfloor\alpha\rfloor+1}$. Hence, by the choice of u,

$$u^T\left((A(\epsilon)+B)^{\circ\alpha} - A(\epsilon)^{\circ\alpha} - B^{\circ\alpha}\right) u = -\epsilon^\alpha + u^T \cdot o(\epsilon^\alpha) \cdot u,$$

and this is negative for small $\epsilon > 0$. Hence, x^α is not Loewner super-additive even on rank-1 matrices in $\mathbb{P}_n((0,1])$. $\square$

Remark 14.11. The above proofs of Theorems 14.9 and 14.10 apply for arbitrary $v = (v_1, \dots, v_n)^T$, consisting of pairwise distinct positive real scalars.

14.2. **Entrywise powers preserving $\mathbb{P}_G$.** We now apply the above results to compute the set of entrywise powers preserving positivity on $\mathbb{P}_{K_n^{(1)}}$ (the almost complete graph).

Proof of Theorem 14.6. The result is straightforward for $n = 2$, so we assume henceforth that $n \geq 3$. It suffices to show that $\mathcal{H}_{K_n^{(1)}} \subset \mathbb{Z}^{\geq 0} \cup [n-2,\infty)$, since the reverse inclusion follows from Theorem 9.3 via (14.2). Fix $x > 0,\ x \neq 1$, and define

$$v := (1, x, \dots, x^{n-3})^T \in \mathbb{R}^{n-2}, \qquad A(\epsilon) := \begin{pmatrix} 1 & \mathbf{1}^T & 0 \\ \mathbf{1} & \mathbf{1}\mathbf{1}^T + \epsilon v v^T & \sqrt{\epsilon} v \\ 0 & \sqrt{\epsilon} v^T & 1 \end{pmatrix}_{n\times n}, \qquad \epsilon > 0.$$

Note that if $p, q > 0$ are scalars, $\mathbf{a}, \mathbf{b} \in \mathbb{R}^{n-2}$ are vectors, and B is an $(n-2)\times(n-2)$ matrix, then using Schur complements (Theorem 2.32),

$$\begin{aligned}\begin{pmatrix} p & \mathbf{a}^T & 0 \\ \mathbf{a} & B & \mathbf{b} \\ 0 & \mathbf{b}^T & q \end{pmatrix} \in \mathbb{P}_n \quad &\iff \quad \begin{pmatrix} p & \mathbf{a}^T \\ \mathbf{a} & B - q^{-1}\mathbf{b}\mathbf{b}^T \end{pmatrix} \in \mathbb{P}_{n-1} \\ &\iff \quad B - p^{-1}\mathbf{a}\mathbf{a}^T - q^{-1}\mathbf{b}\mathbf{b}^T \in \mathbb{P}_{n-2}.\end{aligned} \tag{14.12}$$

Applying this to the matrices $A(\epsilon)$ and $A(\epsilon)^{\circ\alpha}$, we obtain: $A(\epsilon) \in \mathbb{P}_n$, and

$$A(\epsilon)^{\circ\alpha} \in \mathbb{P}_n \quad \iff \quad (\mathbf{1}\mathbf{1}^T + \epsilon v v^T)^{\circ\alpha} - (\mathbf{1}\mathbf{1}^T)^{\circ\alpha} - (\epsilon v v^T)^{\circ\alpha} \in \mathbb{P}_{n-2}.$$

For small $\epsilon > 0$, Theorem 14.10 now shows that $\alpha \in \mathbb{Z}^{\geq 0} \cup [n-2,\infty)$, as desired. $\square$

In the remainder of this section, we present what is known about the critical exponents α_G and power-preserver sets $\mathcal{H}_G$ for various graphs. We do not provide proofs below, instead referring the reader to the 2016 paper [153] by Guillot, Khare, and Rajaratnam in *J. Combin. Theory, Ser. A*.

The first family of graphs is that of *chordal graphs*, and it subsumes not only complete graphs, trees, and almost complete graphs (for all of which we have computed $\mathcal{H}_G, \alpha_G$ with full proofs above), but also other graphs including split, banded, and Apollonian graphs, which we shall now discuss.

Definition 14.13. A graph is *chordal* if it has no induced cycle of length ≥ 4.

Chordal graphs are important in many fields. They are also known as triangulated graphs, decomposable graphs, and rigid circuit graphs. They occur in spectral graph theory, but also in network theory, optimization, and Gaussian graphical models. Chordal graphs play a fundamental role in areas including maximum likelihood estimation in Markov random fields, perfect Gaussian elimination, and the matrix completion problem.

The following is the main result of the aforementioned 2016 paper [153], and it computes $\mathcal{H}_G$ for every chordal graph:

Theorem 14.14. *Let G be a chordal graph with $n \geq 2$ nodes and at least one edge. Let r denote the largest integer, such that K_r or $K_r^{(1)} \subset G$. Then $\mathcal{H}_G = \mathbb{Z}^{\geq 0} \cup [r-2, \infty)$.*

The point of the theorem is that the study of powers preserving positivity reduces solely to the geometry of the graph and can be understood combinatorially rather than through matrix analysis (given the theorem). While we do not prove this result here, we remark that the proof crucially uses Theorem 14.6 and the "clique-tree decomposition" of a chordal graph.

As applications of Theorem 14.14, we mention several examples of chordal graphs G and their critical exponents α_G; by the preceding theorem, the only powers below α_G that preserve positivity on $\mathbb{P}_G$ are the non-negative integers.

(1) The complete and almost complete graph on n vertices are chordal and have critical exponent $n-2$.
(2) Trees are chordal and have critical exponent 1.
(3) Let C_n denote a cycle graph (for $n \geq 4$), which is clearly not chordal. Any minimal planar triangulation G of C_n is chordal, and one can check that $\alpha_G = 2$ regardless of the size of the original cycle graph or the locations of the additional chords drawn.
(4) A banded graph with bandwidth $d > 0$ is a graph with vertex set $[n] = \{1, \dots, n\}$ and edges $(j, j+x)$ for $x \in \{-d, -d+1, \dots, d-1, d\}$, such that $1 \leq j, j+x \leq n$. Such graphs are chordal, and one checks (combinatorially) that $\alpha_G = \min(d, n-2)$ if $n > d$.
(5) A split graph consists of a clique $C \subset V$ and an independent (i.e., pairwise disconnected) set $V \setminus C$, whose nodes are connected to various nodes of C. Split graphs are an important class of chordal graphs, because it can be shown that the proportion of (connected) chordal graphs with n nodes that are split graphs grows to 1 as $n \to \infty$. Theorem 14.14 implies that for a split graph G,
$$\alpha_G = \max(|C| - 2, \max \deg(V \setminus C)).$$
(6) Apollonian graphs are constructed as follows: start with a triangle as the first iteration. Given any iteration, which is a subdivision of the original triangle by triangles, choose an interior point of any of these "atomic" triangles, and connect it to the three vertices of the corresponding atomic triangle. This increases the number of atomic

triangles by 2 at each step. If G is an Apollonian graph on $n \geq 3$ nodes, one shows that $\alpha_G = \min(2, n-2)$. Notice, for $n \geq 4$ this is independent of n or the specific triangulation.

It is natural to ask what is known for non-chordal graphs. We mention one such result, also shown in the aforementioned 2016 paper [153].

Theorem 14.15. *Let C_n denote the cycle graph on n vertices (which is non-chordal for $n \geq 4$). Then $\mathcal{H}_{C_n} = [1, \infty)$ for all $n \geq 4$.*

Remarkably, this is the same combinatorial recipe as for chordal graphs (in Theorem 14.14)!

We end with some questions, which can be avenues for further research into this nascent topic.

Question 14.16.

(1) Compute the critical exponent (and set of powers preserving positivity) for graphs other than the ones discussed above. In particular, compute α_G for all $G = (V, E)$ with $|V| \leq 5$.
(2) For all graphs G with known critical exponent, the critical exponent turns out to be $r - 2$, where r is the largest integer, such that G contains either K_r or $K_r^{(1)}$. Does the same result hold for all graphs?
(3) In fact, more is true in all known cases: $\mathcal{H}_G = \mathbb{Z}^{\geq 0} \cup [\alpha_G, \infty)$. Is this true for all graphs?
(4) Taking a step back: can one show that the critical exponent of a graph is an integer (perhaps without computing it explicitly)?
(5) Does the critical exponent have connections to – or can it be expressed in terms of – other, purely combinatorial graph invariants?
(6) More generally, is it possible to classify the entrywise functions that preserve positivity on $\mathbb{P}_G$, for G a non-complete, non-tree graph? Perhaps the simplest example is a cycle $G = C_n$.

15. Loewner convexity. Single matrix encoders of entrywise power-preservers of Loewner properties.

In Section 14, we classified the entrywise powers that are Loewner monotone or super-additive on all matrices in $\mathbb{P}_n([0,\infty))$. In this section, we similarly classify the Loewner convex powers on $\mathbb{P}_n$ (a notion that is not yet defined). Before doing so, we show that there exist *individual* matrices that turn out to encode the sets of entrywise powers preserving Loewner positivity and monotonicity.

15.1. **Matrices encoding Loewner positive powers.** We begin with the Loewner positive powers and recall Jain's Theorem 9.10. This was strengthened by Jain in her 2020 paper in *Adv. Oper. Theory*:

Theorem 15.1 (Jain)**.** *Suppose $n \geq 2$ is an integer, and $x_1, x_2, \dots, x_n$ are pairwise distinct real numbers, such that $1 + x_j x_k > 0$ for all j, k. Let $C := (1 + x_j x_k)_{j,k=1}^n$. Then $C^{\circ\alpha}$ is positive semidefinite if and only if $\alpha \in \mathbb{Z}^{\geq 0} \cup [n-2, \infty)$.*

The remainder of this subsection is devoted to proving Theorem 15.1, beginning with the following notation:

Definition 15.2. Given a real tuple $\mathbf{x} = (x_1, \dots, x_n)$, define $A_{\mathbf{x}} := -\infty$ if all $x_j \leq 0$, and $-1/\max_j x_j$ otherwise. Similarly, define $B_{\mathbf{x}} := \infty$ if all $x_j \geq 0$, and $-1/\min_j x_j$ otherwise.

Here are a few properties of $A_{\mathbf{x}}, B_{\mathbf{x}}$; the details are straightforward verifications, which are left to the reader.

Lemma 15.3.

(1) Suppose $x \in \mathbb{R}$. Then $1 + yx > 0$ for a real scalar y, if and only if $\operatorname{sgn}(x) y \in (-1/|x|, \infty)$, where we set $1/|x| := \infty$ if $x = 0$.
(2) Given real scalars $x_1, \dots, x_n, y$, we have $1 + yx_j > 0$ for all j if and only if $y \in (A_{\mathbf{x}}, B_{\mathbf{x}})$.
(3) $A_{-\mathbf{x}} = -B_{\mathbf{x}}$ and $A_{\mathbf{x}} < 0 < B_{\mathbf{x}}$ for all $\mathbf{x} \in \mathbb{R}^n$.

Proof sketch. The first part follows by using $x = \operatorname{sgn}(x)|x|$; the second follows by intersecting the solution-intervals for each x_j. The final assertion is a consequence of the first two. $\square$

We now show an intermediate result that resembles Descartes' rule of signs (Lemma 5.2), except that it holds for powers of $1 + ux$ rather than $\exp(ux)$:

Proposition 15.4. *Fix a real number r, an integer $n \geq 1$, and two real tuples $\mathbf{c} = (c_1, \dots, c_n) \neq 0$ and $\mathbf{x} = (x_1, \dots, x_n)$ with pairwise distinct x_j. Then the function*

$$\varphi_{\mathbf{x},\mathbf{c},r} : (A_{\mathbf{x}}, B_{\mathbf{x}}) \to \mathbb{R}, \qquad u \mapsto \sum_{j=1}^n c_j (1 + ux_j)^r$$

either is identically zero or has at most $n-1$ zeros, counting multiplicities.

Proof. The proof of this Descartes-type result once again employs the trick by Poulain and Laguerre – namely, to use Rolle's theorem and induct. If $r = 0$, then the result is straightforward, so assume henceforth that $r \neq 0$. Let $s = S^-(\mathbf{c})$ denote the number of sign changes in the non-zero tuple $\mathbf{c} = (c_1, \dots, c_n)$. Now claim more strongly that the number of zeros is at most s. The proof is by induction on $n \geq 1$ and then on $s \in [0, n-1]$. The base case of $n = 1$ is clear; and, for any n, the base case of $s = 0$ is also immediate. Thus, suppose $n \geq 2$

and $s \geq 1$, and suppose the result holds for all tuples $\mathbf{c}$ of length $< n$, as well as for all tuples $\mathbf{c} \in \mathbb{R}^n \setminus \{0\}$ with at most $s-1$ sign changes. Because of this, we may also assume that all c_j are non-zero and $S^-(\mathbf{c}) = s$.

We begin by relabeling the x_j if required, to lie in increasing order:

$$x_1 < \cdots < x_n.$$

Now, suppose there does not exist $0 < k \leq n$, such that $c_{k-1}c_k < 0$ (there is a sign change here in the tuple $\mathbf{c}$, which is also relabeled corresponding to the x_j if required) and $x_k > 0$. Then $x_k \leq 0$, so that $x_{k-1} < 0$. Now work with the tuples $-\mathbf{x}$ and $\mathbf{c}' := (c_n, \dots, c_1)$, i.e.,

$$-x_n < -x_{n-1} < \cdots < -x_1, \qquad \varphi_{-\mathbf{x},\mathbf{c}',r}(v) := \sum_{j=n}^{1} c_j(1-ux_j)^r.$$

Here $v = -u \in (-B_{\mathbf{x}}, -A_{\mathbf{x}}) = (A_{-\mathbf{x}}, B_{-\mathbf{x}})$ by Lemma 15.3, so the result for $\varphi_{-\mathbf{x},\mathbf{c}',r}(v)$ would prove that for $\varphi_{\mathbf{x},\mathbf{c},r}(u)$. Using this workaround if needed, it follows that there exists $1 \leq k \leq n$, such that $c_{k-1}c_k < 0$ and $x_k > 0$. In particular, there exists $v > 0$, such that

$$1 - vx_n < \cdots < 1 - vx_k < 0 < 1 - vx_{k-1} < \cdots < 1 - vx_1.$$

Define

$$\psi : (A_{\mathbf{x}}, B_{\mathbf{x}}) \to \mathbb{R}, \qquad \psi(u) := \sum_{j=1}^{n} c_j(1-vx_j)(1+ux_j)^{r-1}.$$

By choice of v, the sequence $(c_1(1-vx_1), \dots, c_n(1-vx_n))$ has precisely $s-1$ sign changes, so ψ has at most $s-1$ zeros. Now, for $u \in (A_{\mathbf{x}}, B_{\mathbf{x}})$, we compute

$$\begin{aligned}\psi(u) = &\sum_{j=1}^{n} c_j(1+ux_j-(u+v)x_j)(1+ux_j)^{r-1}\\ &= \varphi_{\mathbf{x},\mathbf{c},r}(u) - (u+v)\sum_{j=1}^{n} c_jx_j(1+ux_j)^{r-1} = -\frac{(u+v)^{r+1}}{r}h'(u),\end{aligned}$$

where $h(u) := (u+v)^{-r}\varphi_{\mathbf{x},\mathbf{c},r}(u)$ and $r \neq 0$. Since $x_k > 0$, we obtain from above

$$u \in (A_{\mathbf{x}}, B_{\mathbf{x}}) \quad \Longrightarrow \quad u + v > A_{\mathbf{x}} + v = v - x_n^{-1} > v - x_k^{-1} > 0.$$

Thus, $u \mapsto u+v$ is positive on $(A_{\mathbf{x}}, B_{\mathbf{x}})$, hence $h : (A_{\mathbf{x}}, B_{\mathbf{x}}) \to \mathbb{R}$ is well defined. From above, ψ has at most $s-1$ zeros on $(A_{\mathbf{x}}, B_{\mathbf{x}})$, hence so does h'. But then by Rolle's theorem, h has at most s zeros on $(A_{\mathbf{x}}, B_{\mathbf{x}})$, and hence, so does $\varphi_{\mathbf{x},\mathbf{c},r}$. $\square$

A second intermediate result involves a homotopy argument that will be crucial in proving Theorem 15.1:

Proposition 15.5. *Let $n \geq 2$ be an integer and fix real scalars*

$$x_1 < \cdots < x_n, \qquad 0 < y_1 < \cdots < y_n,$$

such that $1 + x_jx_k > 0$ for all j, k. Then there exists $\epsilon_0 > 0$, such that for all $0 < \epsilon \leq \epsilon_0$, the linear homotopies (between x_j and ϵy_j)

$$x_j^{(\epsilon)}(t) := x_j + t(\epsilon y_j - x_j), \qquad t \in [0,1]$$

all satisfy

$$1 + x_j^{(\epsilon)}(t)x_k^{(\epsilon)}(t) > 0, \qquad \forall j, k = 1, \dots, n, \ t \in [0,1].$$

Notice that this is not immediate, nor even true for arbitrary ϵ. For instance, suppose $n = 2$, $\epsilon = 1$, and $(y_1, y_2) = (1, 2)$. If, say, $(x_1, x_2) = (-199, 0)$, then an easy "completion of squares" shows that the assertion is false at most times in the homotopy:

$$1 + x_1^{(1)}(t)x_2^{(1)}(t) \leq 0, \quad \forall t \in \left[\frac{398}{800} - \frac{1}{20}\sqrt{\frac{398^2}{40^2} - 1}, \frac{398}{800} + \frac{1}{20}\sqrt{\frac{398^2}{40^2} - 1}\right] \supset [0.0026, 0.9924].$$

Similarly, if, say, $(x_1, x_2) = (-8.5, 0.1)$, then

$$1 + x_1^{(1)}(t)x_2^{(1)}(t) \leq 0, \quad \forall t \in \left[\frac{8 - \sqrt{61}}{19}, \frac{8 + \sqrt{61}}{19}\right] \supset [0.01, 0.8321].$$

Thus, the ϵ in the statement is crucial for the result to hold.

Proof of Proposition 15.5. We begin with three observations; in all of them, $x_j(t) = x_j^{(\epsilon)}(t)$ for some fixed $\epsilon > 0$, and all j, t. First, we have $x_1(t) < \cdots < x_n(t)$ for all $t \in [0, 1]$.

Second, if $x_1 \geq 0$, then it is clear that $x_j(t) \geq 0$ for all $j \in [1, n]$ and all $t \in [0, 1]$, and the result is immediate. Thus, we suppose henceforth that $x_1 < 0$.

Third, if there exist integers $j < k$ and a time $t \in [0, 1]$, such that $1 + x_j(t)x_k(t) \leq 0$, then $x_j(t) < 0 < x_k(t)$, and hence $x_1(t) < 0 < x_n(t)$. One then verifies easily that $1 + x_1(t)x_n(t) \leq 1 + x_j(t)x_k(t) \leq 0$.

Thus, for every choice of x_j, y_j as above, with $x_1 < 0$, it suffices to produce $\epsilon_0 > 0$, such that $1 + x_1^{(\epsilon)}(t)x_n^{(\epsilon)}(t) > 0$ for all $t \in (0, 1)$ and all $0 < \epsilon \leq \epsilon_0$. There are two cases, depending on the sign of x_n:

Case 1: $x_n \geq 0$. Then $x_n < 1/|x_1|$. We claim that $\epsilon_0 := 1/(|x_1|y_n)$ works; to see this, compute using the known inequalities on x_j, y_j:

$$\begin{aligned} 1 + x_1^{(\epsilon)}(t)x_n^{(\epsilon)}(t) &= 1 + (t\epsilon y_1 + (1-t)x_1)(t\epsilon y_n + (1-t)x_n) \\ &> 1 + (1-t)x_1(t\epsilon y_n + (1-t)x_n) > 1 + (1-t)x_1(t\epsilon y_n + (1-t)/|x_1|), \end{aligned}$$

where the (final) two inequalities are strict because $t \in (0, 1)$. Continuing, this last expression equals

$$= 1 - (1-t)^2 + t(1-t)\epsilon y_n x_1 \geq t\left(2 - t - (1-t)\epsilon_0 y_n |x_1|\right) = t > 0.$$

Case 2: $x_n < 0$. For ϵ close to 0, define

$$f(\epsilon) := 1 - \frac{\epsilon^2(x_n y_1 - x_1 y_n)^2}{4(\epsilon y_1 - x_1)(\epsilon y_n - x_n)}.$$

This function is continuous in ϵ and $f(0) > 0$. Hence, there exists $\epsilon_0 > 0$, such that $f(\epsilon) > 0$ for all $0 \leq \epsilon \leq \epsilon_0$.

We show that ϵ_0 satisfies the desired properties. Let $0 < \epsilon \leq \epsilon_0$, and set

$$t_j^{(\epsilon)} := -x_j/(\epsilon y_j - x_j), \qquad 1 \leq j \leq n.$$

Note that $x_j^{(\epsilon)}(t)$ is negative, zero, or positive when $t < t_j^{(\epsilon)}, t = t_j^{(\epsilon)}$, or $t > t_j^{(\epsilon)}$, respectively. Also note by the observations above, and since $t_j^{(\epsilon)}$ is the time at which $x_j^{(\epsilon)}(\cdot)$ vanishes, that

$$0 < t_n^{(\epsilon)} < t_{n-1}^{(\epsilon)} < \cdots < t_1^{(\epsilon)} < 1.$$

Now, if $0 \leq t \leq t_n^{(\epsilon)}$ or $t_1^{(\epsilon)} \leq t \leq 1$, then $x_1^{(\epsilon)}(t), x_n^{(\epsilon)}(t)$ are both non-positive or non-negative, respectively. Hence, $1 + x_1^{(\epsilon)}(t)x_n^{(\epsilon)}(t) \geq 1$, as desired.

Next, suppose $t \in (t_n^{(\epsilon)}, t_1^{(\epsilon)})$. Observe that

$$x_j^{(\epsilon)}(t) = t\epsilon y_j + (1-t)x_j = (t - t_j^{(\epsilon)})(\epsilon y_j - x_j), \qquad \forall j \in [1,n],\ t \in [0,1].$$

Now, using the AM–GM inequality, the proof is complete:

$$\begin{aligned} 1 + x_1^{(\epsilon)}(t) x_n^{(\epsilon)}(t) &= 1 + (t - t_1^{(\epsilon)})(t - t_n^{(\epsilon)})(\epsilon y_1 - x_1)(\epsilon y_n - x_n) \\ &\geq 1 - \frac{1}{4}(t_1^{(\epsilon)} - t_n^{(\epsilon)})^2 (\epsilon y_1 - x_1)(\epsilon y_n - x_n) = f(\epsilon) > 0. \end{aligned} \qquad \square$$

With all of the above results at hand, we can proceed:

Proof of Theorem 15.1. For ease of exposition, we break up the proof into steps.

Step 1: The first observation is slightly more general than applies here. *Suppose $y_1, \dots, y_n$ are distinct real numbers, such that $1 + y_k x_j > 0$ for all $1 \leq j, k \leq n$. Let $S := (1 + y_k x_j)$, and let r be real. If $r \in \{0, 1, \dots, n-2\}$, then $S^{\circ r}$ has rank $r+1$, else $S^{\circ r}$ is non-singular.*

Indeed, for $r \in \{0, \dots, n-2\}$, we have

$$S^{\circ r} = (W_{\mathbf{y}}^{(r)})^T D W_{\mathbf{x}}^{(r)}, \qquad \text{where} \ \ W_{\mathbf{x}}^{(r)} := \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_1^r & x_2^r & \cdots & x_n^r \end{pmatrix},$$

and where $D_{(r+1)\times(r+1)}$ is a diagonal matrix with entries $\binom{r}{0}, \binom{r}{1}, \dots, \binom{r}{r}$. Now D is non-singular, and $W_{\mathbf{x}}^{(r)}, W_{\mathbf{y}}^{(r)}$ are submatrices of Vandermonde matrices and hence of full rank – so $S^{\circ r}$ has rank $r+1$.

Now suppose $r \neq 0, 1, \dots, n-2$ and $S^{\circ r}\mathbf{c}^T = \mathbf{0}$ for some tuple $\mathbf{c} = (c_1, \dots, c_n) \neq 0$. Rewrite $S^{\circ r}\mathbf{c}^T = \mathbf{0}$ to obtain

$$\varphi_{\mathbf{x},\mathbf{c},r}(y_k) = \sum_{j=1}^n c_j (1 + y_k x_j)^r = 0, \qquad k = 1, \dots, n.$$

Since $1 + y_k x_j > 0$ for all j, k, we have $y_k \in (A_{\mathbf{x}}, B_{\mathbf{x}})$ by Lemma 15.3(2). So $\varphi_{\mathbf{x},\mathbf{c},r} \equiv 0$ on $(A_{\mathbf{x}}, B_{\mathbf{x}})$, by Proposition 15.4, and hence $\varphi_{\mathbf{x},\mathbf{c},r}^{(l)}(0) = 0$ for $l = 0, 1, \dots, n-1$ by Lemma 15.3(3). Reformulating this,

$$\sum_{j=1}^n \left(c_j r(r-1)\cdots(r-l+1)\right) x_j^l = 0, \qquad \forall l = 0, 1, \dots, n-1,$$

i.e., $W_{\mathbf{x}}^{(n-1)} D \mathbf{c}^T = \mathbf{0}$, where D is a diagonal matrix with diagonal entries $1, r, r(r-1), \dots, r(r-1)\cdots(r-n+2)$. Now $W_{\mathbf{x}}^{(n-1)}$ is a non-singular (Vandermonde) matrix, as is D by choice of r. Thus, the tuple $\mathbf{c}$ is zero, i.e., $S^{\circ r}$ is non-singular.

Step 2: We now turn to the proof of the theorem. First, if $\alpha \in \mathbb{Z}^{\geq 0} \cup [n-2, \infty)$, then $C^{\circ\alpha}$ is positive semidefinite by Theorem 9.3. Next, if $\alpha < 0$, then the leading 2×2 principal minor of $C(\mathbf{x})^{\circ\alpha}$ is easily seen to be negative. Finally, suppose $\alpha \in (0, n-2) \setminus \mathbb{Z}$. Given a real vector $\mathbf{y} \in \mathbb{R}^n$, define $C(\mathbf{y}) := \mathbf{1}_{n\times n} + \mathbf{y}\mathbf{y}^T$. Now apply the previous step, fixing $r = \alpha$ and all $y_j = x_j$. Thus, $\det C(\mathbf{x})^{\circ\alpha} \neq 0$ for every $\mathbf{x}$ with all $1 + x_j x_k > 0$.

Now recall the proof of Theorem 9.3, and the subsequent remarks. Thus, if $x_0 < \min_j x_j$, then there exists $\epsilon_1 > 0$, such that $C(\sqrt{\epsilon}(\mathbf{x} - x_0\mathbf{1}))^{\circ\alpha}$ has a negative eigenvalue for all $0 < \epsilon < \epsilon_1$. Now consider the linear homotopy

$$\mathbf{x}(t) := (1 - t + t\sqrt{\epsilon_2})\mathbf{x} - t\sqrt{\epsilon_2}x_0\mathbf{1}, \qquad t \in [0,1],$$

which goes from $\mathbf{x}$ to $\sqrt{\epsilon_2}(\mathbf{x} - x_0\mathbf{1})$ as t goes from 0 to 1. Here we choose $\epsilon_2 \in (0, \epsilon_1)$, such that $\epsilon_0 := \sqrt{\epsilon_2}$ satisfies the conclusions of Proposition 15.5 for x_j as above and $y_j := x_j - x_0$ (suitably relabeled to be in increasing order if desired).

Again applying the previous step (for the same fixed $r = \alpha$), $\det C(\mathbf{x}(t))^{\circ\alpha} \neq 0$ for all $t \in [0,1]$. We also know that $C(\mathbf{x}(1))^{\circ\alpha}$ has a negative eigenvalue, hence $\lambda_{\min}(C(\mathbf{x}(1))) < 0$. Now claim by the "continuity of eigenvalues" that $\lambda_{\min}(C(\mathbf{x}(t))) < 0$ for all $t \in [0,1]$, and in particular at $t = 0$. This is shown in the next step and completes the proof.

Step 3: The claim in the preceding paragraph follows from this more general fact: *Suppose $C : [0,1] \to \mathbb{C}^{n\times n}$ is a continuous Hermitian matrix-valued function, such that each $C(t)$ is non-singular. If $C(1)$ has a negative eigenvalue, then so does $C(t)$ for all $t \in [0,1]$.*

It remains to show this statement, and we use a simpler approach (than the full power of 'continuity of roots' in Proposition 8.7 above) to do so. Let $X := \{t \in [0,1] : \lambda_{\min}(C(t)) \geq 0\}$. Since the cone of positive semidefinite matrices is closed, it follows that X is closed in $[0,1]$. Now the claim follows from the sub-claim that $X^c := [0,1] \setminus X$ is also closed: since $[0,1]$ is connected and $1 \in X^c$, it follows that $X^c = [0,1]$ and so $0 \in X^c$ as desired.

To show the sub-claim, let $\|C(t)\| := \left(\sum_{j,k=1}^n |c_{jk}|^2\right)^{1/2}$. It is clear using the Cauchy–Schwartz inequality that all eigenvalues of $C(t)$ lie in $[-\|C(t)\|, \|C(t)\|]$. Now, given a sequence $t_n \in X^c$ that converges to $t_0 \in [0,1]$, all entries of $\{C(t_n) : n \geq 1\}$ lie in a compact set, hence so do the corresponding minimum eigenvalues $\lambda_{\min}(C(t_n))$. Pick a subsequence n_k, such that the sequence $\lambda_{\min}(C(t_{n_k}))$ is convergent, with limit λ_0, say. Now $\lambda_0 \leq 0$. Also pick a unit-length eigenvector v_n of $C(t_n)$ corresponding to the eigenvalue $\lambda_{\min}(C(t_n))$; as the unit complex sphere is compact, there is a further subsubsequence n_{k_l}, such that $v_{n_{k_l}} \to v_0$ as $l \to \infty$, with v_0 also of unit norm.

With these choices at hand, write the equation $C(t_{n_{k_l}})v_{n_{k_l}} = \lambda_{\min}(C(t_{n_{k_l}}))v_{n_{k_l}}$ and let $l \to \infty$. Then $C(t_0)v_0 = \lambda_0 v_0$, with $\lambda_0 \leq 0$. It follows from the hypotheses that $\lambda_{\min}(C(t_0)) < 0$, and the proof is complete. □

15.2. **Matrices encoding Loewner monotone powers.** We now turn to Loewner monotonicity (recall Theorem 14.9). The next result – again by Jain in 2020 [192] – shows that, akin to Theorem 15.1, there exist individual matrices that encode the Loewner monotone powers:

Corollary 15.6. *Suppose $n \geq 1$ and $x_1, \dots, x_n$ are distinct non-zero real numbers, such that $1 + x_j x_k > 0$ for all j, k. Let $\mathbf{x} := (x_1, \dots, x_n)^T$ and $\alpha \in \mathbb{R}$. Then $(\mathbf{1}_{n\times n} + \mathbf{x}\mathbf{x}^T)^{\circ\alpha} \geq \mathbf{1}_{n\times n}$ if and only if $\alpha \in \mathbb{Z}^{\geq 0} \cup [n-1, \infty)$, if and only if x^α is Loewner monotone on $\mathbb{P}_n((0,\infty))$.*

Notice that here we cannot take $x_j = 0$ for any j; if, for instance, $x_n = 0$ and we call the matrix X, then the monotonicity of $X^{\circ\alpha}$ over $\mathbf{1}_{n\times n}$ is actually equivalent to the positivity of $X^{\circ\alpha}$, and so the result fails to hold.

Proof. If $\alpha \in \mathbb{Z}^{\geq 0} \cup [n-1, \infty)$, then Theorem 14.9 implies x^α is Loewner monotone on $\mathbb{P}_n((0,\infty))$, hence on $X := \mathbf{1}_{n\times n} + \mathbf{x}\mathbf{x}^T \geq \mathbf{1}_{n\times n}$. Conversely, suppose x^α is Loewner monotone on $X \geq \mathbf{1}_{n\times n}$. Let $\mathbf{x}' := (\mathbf{x}^T, 0)^T \in \mathbb{R}^{n+1}$, then

$$\widetilde{X} := \mathbf{1}_{(n+1)\times(n+1)} + \mathbf{x}'(\mathbf{x}')^T = \begin{pmatrix} X & \mathbf{1} \\ \mathbf{1}^T & 1 \end{pmatrix}$$

satisfies the hypotheses of Theorem 15.1. Now, by the theory of Schur complements (Theorem 2.32), $X^{\circ\alpha} \geq \mathbf{1}_{n\times n}$ if and only if $\widetilde{X}^{\circ\alpha} \in \mathbb{P}_{n+1}$. But this is if and only if $\alpha \in \mathbb{Z}^{\geq 0} \cup [n-1, \infty)$, by Theorem 15.1. $\square$

15.3. **Loewner convex powers, and individual matrices encoding them.** Finally, we turn to the entrywise powers preserving Loewner convexity.

Definition 15.7. Let $I \subset \mathbb{R}$ and $n \in \mathbb{N}$. A function $f : I \to \mathbb{R}$ is said to be *Loewner convex* on a subset $V \subset \mathbb{P}_n(I)$ if $f[\lambda A + (1-\lambda)B] \leq \lambda f[A] + (1-\lambda) f[B]$, whenever $A \geq B \geq 0_{n\times n}$ lie in V, and $\lambda \in [0,1]$.

The final theorem in this section classifies the Loewner convex powers in the spirit of Theorems 9.3, 14.9, and 14.10. It shows in particular that there is a critical exponent for convexity as well. It also shows the encoding of these powers by individual matrices, in the spirit of Corollary 15.6:

Theorem 15.8 (Loewner convex entrywise powers)**.** *Fix an integer $n \geq 1$ and a scalar $\alpha \in \mathbb{R}$. The following are equivalent:*

(1) The entrywise power x^α is Loewner convex on $\mathbb{P}_n([0,\infty))$.
(2) Fix distinct non-zero real numbers $x_1, \dots, x_n$, such that $1 + x_j x_k > 0$ for all j,k. Then x^α is Loewner convex on $A := (1 + x_j x_k)_{j,k=1}^n \geq B = \mathbf{1}_{n\times n} \geq 0$.
(3) $\alpha \in \mathbb{Z}^{\geq 0} \cup [n, \infty)$.

In particular, the critical exponent for Loewner convexity on $\mathbb{P}_n$ is n.

Thus, there are rank-2 Hankel TN matrices (with $x_j = x_0^j$ for $x_0 \in (0,\infty) \setminus \{1\}$), which encode the Loewner convex powers.

To prove this result, we require a preliminary result connecting Loewner convex functions with Loewner monotone ones. We also prove a parallel result connecting monotone maps to positive ones.

Proposition 15.9. *Suppose $n \geq 1$ and $A \geq B \geq 0_{n\times n}$ are positive semidefinite matrices with real entries, such that $A - B = uu^T$, with u having all non-zero entries. Fix any open interval I containing the entries of A, B, and suppose $f : I \to \mathbb{R}$ is differentiable.*

(1) Then both notions of the "interval" $[B, A]$ agree, i.e.,

$$\{C : B \leq C \leq A\} = \{\lambda A + (1-\lambda)B : \lambda \in [0,1]\}.$$

(2) If $f[-]$ is Loewner monotone on the interval $[B, A]$, then $f'[-]$ is Loewner positive on (B, A). The converse holds for arbitrary matrices $0 \leq B \leq A$.
(3) If $f[-]$ is Loewner convex on the interval $[B, A]$, then $f'[-]$ is Loewner monotone on (B, A). The converse holds for arbitrary matrices $0 \leq B \leq A$.

Proof.

(1) That the left-hand side contains the right is straightforward. Conversely, if $B \leq C \leq A$, then $0 \leq C - B \leq A - B$, which has rank 1. Write $A - B = uu^T$; now, if $u^T v = 0$, then $\|\sqrt{C-B} \cdot v\|^2 = v^T (C-B) v = 0$, so $(C-B)v = 0$. This inclusion of kernels shows that $\ker(C-B)$ has a codimension of at most one. If $C \neq B$, then $\ker(C-B) = \ker u^T$ and $C - B$ has column space spanned by u, by the orthogonality

of eigenspaces of Hermitian matrices for different eigenvalues. Thus, $C - B = \lambda uu^T$ for some $\lambda \in (0,1]$. But then,

$$C = B + \lambda(A - B) = \lambda A + (1 - \lambda)B,$$

as desired.

(2) Suppose f' is Loewner positive on (B, A) (for any $0 \leq B \leq A$). We show that f is monotone on this interval by using the integration trick (9.7) (see also Theorem 9.12). Indeed,

$$f[A] - f[B] = \int_0^1 (A - B) \circ f'[\lambda A + (1 - \lambda)B] \; d\lambda.$$

By assumption and the Schur product theorem, the integrand is positive semidefinite, hence so is the left-hand side, as desired. The same argument applies to show that $f[A_\lambda] \geq f[A_\mu]$, where $A_\lambda := \lambda A + (1 - \lambda)B$ and $0 \leq \mu \leq \lambda \leq 1$.

Conversely, suppose $f[A_\lambda] \geq f[A_\mu]$ for all $0 \leq \mu \leq \lambda \leq 1$. Now, given $\lambda \in (0, 1)$, let $0 < h \leq 1 - \lambda$, then $f[A_{\lambda+h}] \geq f[A_\lambda]$, so

$$\begin{aligned}
0 &\leq \lim_{h \to 0^+} \frac{1}{h} \left(f[A_{\lambda+h}] - f[A_\lambda] \right) \\
&= \lim_{h \to 0^+} \frac{1}{h} \left(f[\lambda A + (1 - \lambda)B + h(A - B)] - f[\lambda A + (1 - \lambda)B] \right) \\
&= f'[A_\lambda] \circ (A - B).
\end{aligned}$$

By the assumptions, $(A - B)^{\circ(-1)}$ is also a rank-1 positive semidefinite matrix with all non-zero entries, so taking the Schur product, we have $f'[A_\lambda] \geq 0$ for all $\lambda \in (0, 1)$, as desired.

(3) Suppose f' is Loewner monotone on (B, A) (for any $0 \leq B \leq A$). As above, we use the integration trick to show that f is convex, beginning with:

$$\begin{aligned}
f[(A + B)/2] - f[B] &= \frac{1}{2} \int_0^1 (A - B) \circ f' \left[\lambda \frac{A + B}{2} + (1 - \lambda)B \right] \; d\lambda, \\
\frac{f[A] + f[B]}{2} - f[B] &= \frac{f[A] - f[B]}{2} = \frac{1}{2} \int_0^1 (A - B) \circ f' \left[\lambda A + (1 - \lambda)B \right] \; d\lambda.
\end{aligned} \tag{15.10}$$

Now, by the hypotheses on f' and the Schur product theorem, it follows that

$$(A - B) \circ f'[\lambda A + (1 - \lambda)B] \geq (A - B) \circ f' \left[\lambda \frac{A + B}{2} + (1 - \lambda)B \right].$$

This, combined with (15.10), yields $f[(A + B)/2] \leq (f[A] + f[B])/2$. One now proves by induction – first on N and then on j – that

$$f \left[\frac{j}{2^N} A + \left(1 - \frac{j}{2^N} \right) B \right] \leq \frac{j}{2^N} f[A] + \left(1 - \frac{j}{2^N} \right) f[B], \qquad \forall N \geq 1, \; 1 \leq j \leq 2^N.$$

Now, given any $\lambda \in [0, 1]$, approximate λ by a sequence of dyadic rationals $j/2^N$ as above, and use the continuity of $f[-]$ and the preceding inequality to conclude that $f[-]$ is Loewner convex on $\{B, A\}$. The same argument can be adapted to show that $f[-]$ is Loewner convex on $\{A_\lambda, A_\mu\}$, as in the preceding part.

Conversely, since we have $f[\lambda A + (1-\lambda)B] \leq \lambda f[A] + (1-\lambda) f[B]$ for $0 \leq \lambda \leq 1$, it follows for $\lambda \in (0,1)$ that

$$\frac{f[B + \lambda(A-B)] - f[B]}{\lambda} \leq f[A] - f[B],$$
$$\frac{f[A + (1-\lambda)(B-A)] - f[A]}{1-\lambda} \leq f[B] - f[A].$$

Letting $\lambda \to 0^+$ and $\lambda \to 1^-$, respectively, yields

$$(A - B) \circ f'[B] \leq f[A] - f[B], \qquad (B - A) \circ f'[A] \leq f[B] - f[A].$$

Adding these yields $(A - B) \circ (f'[A] - f'[B]) \geq 0$. Finally, $A - B$ has all non-zero entries, so has a rank-1 Schur-inverse; taking the Schur product with this yields $f'[A] \geq f'[B]$. As above, the same argument can be adapted to show that $f'[-]$ is Loewner monotone on $\{A_\lambda, A_\mu\}$. $\square$

Finally, we have:

Proof of Theorem 15.8. Clearly, (1) $\implies$ (2). Now setting $f(x) := x^\alpha$, both (2) $\implies$ (3) and (3) $\implies$ (1) follow via Proposition 15.9(3) and Corollary 15.6. $\square$

The above results on individual (pairs of) matrices encoding the entrywise powers preserving Loewner positivity, monotonicity, and convexity naturally lead to the following question.

Question 15.11. Given an integer $n \geq 1$, do there exist matrices $A, B \in \mathbb{P}_n((0,\infty))$, such that $(A+B)^{\circ\alpha} \geq A^{\circ\alpha} + B^{\circ\alpha}$ if and only if $\alpha \in \mathbb{Z}^{\geq 0} \cup [n, \infty)$? In other words, for each $n \geq 1$, is the set of Loewner super-additive entrywise powers on $\mathbb{P}_n([0,\infty))$ (see Theorem 14.10) also encoded by a single pair of matrices?

We provide a partial solution here, for the matrices studied above. Suppose $u = (u_1, \dots, u_n)^T \in (0,\infty)^n$ has pairwise distinct coordinates. Let $A := \mathbf{1}_{n \times n}, B := uu^T$. By computations similar to the proof of Theorem 14.6, it follows that

$$(A+B)^{\circ\alpha} \geq A^{\circ\alpha} + B^{\circ\alpha} \quad \iff \quad \begin{pmatrix} \mathbf{1} + uu^T & \mathbf{1}_{n\times 1} & u \\ \mathbf{1}_{1\times n} & 1 & 0 \\ u^T & 0 & 1 \end{pmatrix}^{\circ\alpha} \in \mathbb{P}_{n+2}. \tag{15.12}$$

When $\alpha = 1$, denote the matrix on the right in (15.12) by $M(u)$; this is easily seen to have rank 2. Now considering any diagonal entry of the inequality on the left in (15.12), we obtain $\alpha \geq 1$. By Theorem 14.10 and Remark 14.11, it suffices to now assume $\alpha \in (1,n) \setminus \mathbb{Z}$. But if $\alpha \in (1, n-1)$, then Theorem 15.1 yields the desired result, by considering the leading principal $(n+1) \times (n+1)$ submatrices in the preceding inequality on the right in (15.12).

Thus, it remains to show that for $\alpha \in (n-1, n)$, the matrix $M(u)^{\circ\alpha} \notin \mathbb{P}_n$ for all u with pairwise distinct, positive coordinates. In fact, we claim that it suffices to show for $\alpha \in (n-1,n)$ that $M(u)^{\circ\alpha}$ is non-singular for all u as above. Indeed, this would imply by a different homotopy argument that $M(\sqrt{\epsilon}u)^{\circ\alpha}$ is non-singular for all $\epsilon > 0$; but for small enough $\epsilon > 0$ the proof of Theorem 14.10 shows that $M(\sqrt{\epsilon}u)^{\circ\alpha}$ has a negative eigenvalue, so the same holds for all $\epsilon > 0$ by the continuity of eigenvalues (or see Step 3 in the proof of Theorem 15.1).

In light of this discussion, we end this section with a question closely related to the preceding question above.

Question 15.13. Suppose $n \geq 2$ and $\alpha \in (n-1, n)$. Is $M(u)^{\circ\alpha} \in \mathbb{P}_{n+2}$, where $u \in (0,\infty)^n$ has pairwise distinct coordinates and $M(u)$ is as in (15.12)?

Bibliographic notes and references

Entrywise calculus was initiated by Schur in 1911, when he defined the map $f[A]$ (although he did not use this notation), in the same paper [329] where he proved the Schur product theorem. Schur also proved the first result involving entrywise maps; see also page cxii of the survey [105].

Theorem 9.3 and Lemma 9.5, which help classify the Loewner positive powers, are by FitzGerald and Horn [123]. The use of the rank-2 Hankel matrices in the proof, as well as the powers preserving positive definiteness in Corollary 9.11 are by Fallat–Johnson–Sokal [113]. The "individual" matrices encoding Loewner positive powers were constructed in Theorem 9.10 by Jain [191]; the extension principle in Theorem 9.12 is by Khare–Tao [217]. Also note the related papers by Bhatia–Elsner [47], Hiai [171], and Guillot–Khare–Rajaratnam [151], which study "two-sided" powers : $\mathbb{R} \to \mathbb{R}$, and which of these are Loewner positive on $\mathbb{P}_n(\mathbb{R})$ for fixed $n \geq 1$.

The historical account of Descartes' rule of signs in Theorem 10.3 is taken in part from Jameson's article [193]; once again the proof of this result – via Rolle's theorem – can be attributed to Laguerre [229]. The proof provided of Theorem 10.1 – which classifies the powers preserving totally positive 3×3 matrices – is by Fallat–Johnson–Sokal [113]. The Cauchy functional equation (see Remark 11.5) has been studied in numerous papers; we mention Banach [17] and Sierpińsky [337], both papers appearing in the same volume of *Fund. Math.* in 1920. The results in Section 11 prior to Remark 11.5 are shown by Fallat–Johnson–Sokal [113], or essentially follow from there. The results following Remark 11.5 are by Belton–Guillot–Khare–Putinar [32].

Section 12.1 on the continuity of bounded mid-convex functions is taken from the book of Roberts and Varberg [301]; the first main Theorem 12.2 there closely resembles a result by Ostrowski [275], while the second Theorem 12.4 was proved independently by Blumberg [51] and Sierpińsky [337]. Theorem 12.7, classifying the Loewner positive maps on $\mathbb{P}_2((0,\infty))$ and $\mathbb{P}_2([0,\infty))$, is essentially by Vasudeva [353]; see also [30, 155] for the versions that led to the present formulation. The short argument for mid-convexity implying continuity, at the end of that proof, is due to Hiai [171]. The remainder of Section 12, classifying all entrywise maps preserving TN and TP in each fixed dimension (for all matrices and for all symmetric matrices) is taken from Belton–Guillot–Khare–Putinar [32]. The two exceptions are the example in (12.16) due to Fallat–Johnson–Sokal [113], and Theorem 12.19 which classifies the powers preserving Hankel TN $n \times n$ matrices; this latter is from [30].

Section 13 on the entrywise functions preserving positivity on $\mathbb{P}_G$ for G a non-complete graph (and specifically, a tree) follows Guillot–Khare–Rajaratnam [152]. Section 14 classifying the Loewner positive powers on $\mathbb{P}_G$ for G a chordal graph – and hence, computing the critical exponent of G for Loewner positivity – is due to Guillot–Khare–Rajaratnam [153] (see also the summary in [154]). The two exceptions are Theorem 14.9 by FitzGerald and Horn [123] and Theorem 14.10 by Guillot–Khare–Rajaratnam [151], which classify the Loewner monotone and super-additive powers on $\mathbb{P}_n((0,\infty))$, respectively. Also see [197] for a survey of critical exponents in the matrix theory literature.

Theorem 15.1 and Corollary 15.6, about individual matrices encoding the Loewner positive and monotone powers respectively, are by Jain [192]. The arguments proving these results are taken from [192] (some of these are variants of results in her earlier works) and from Khare [215] – specifically, the homotopy argument in Proposition 15.5, which differs from Jain's similar assertion in [192] and avoids SSR (strictly sign regular) matrices. Finally, the classification of the Loewner convex powers on $\mathbb{P}_n$ (i.e., the equivalence (1) $\iff$ (3)

in Theorem 15.8) was shown by Hiai [171], via the intermediate Proposition 15.9; see also Guillot–Khare–Rajaratnam [151] for a rank-constrained version of Theorem 15.8. The further equivalence to Theorem 15.8(2), which obtains individual matrix-encoders of the Loewner convex powers, is taken from [215].

Part 3:

Entrywise functions preserving positivity in all dimensions

Part 3: Entrywise functions preserving positivity in all dimensions

16. History – Schoenberg's theorem. Rudin, Herz, Vasudeva. Metric geometry, positive definite functions, spheres, and correlation matrices.

In the forthcoming sections in this part, we take a step back and explore the foundational results on entrywise preservers of positive semidefiniteness – as well as the rich history that motivated these results.

16.1. **History of the problem.** In the forthcoming sections, we will answer the question: "*Which functions, when applied entrywise, preserve positivity (positive semidefiniteness)?*"

(Henceforth, we use the word 'positivity' to denote "positive semidefiniteness.") This question has been the focus of a concerted effort and significant research activity over the past century. It began with the Schur product theorem (1911, [329]) and the following consequence:

Lemma 16.1 (Pólya and Szegő, 1925, [289]). *Suppose a power series $f(x) = \sum_{k=0}^\infty c_k x^k$ is convergent on $I \subset \mathbb{R}$ and $c_k \geq 0$ for all $k \geq 0$. Then $f[-] : \mathbb{P}_n(I) \to \mathbb{P}_n(\mathbb{R})$ for all $n \geq 1$.*

Proof. By induction and the Schur product theorem 3.12, $f(x) = x^k$ preserves positivity on $\mathbb{P}_n(\mathbb{R})$ for all integers $k \geq 0$ and $n \geq 1$, and hence sends $\mathbb{P}_n(I)$ to $\mathbb{P}_n(\mathbb{R})$. From this the lemma follows, using that $\mathbb{P}_n(\mathbb{R})$ is a closed convex cone. $\square$

With Lemma 16.1 in hand, Pólya and Szegő asked if there exist any other functions that preserve positivity on $\mathbb{P}_n$ for all $n \geq 1$. A negative answer would essentially constitute the converse result to the Schur product theorem; and indeed, this was shown by Schur's student Schoenberg (who features extensively in the next part, and is also well known for his substantial contribution to the theory of splines), for continuous functions:

Theorem 16.2 (Schoenberg, 1942, [318]). *Suppose $I = [-1, 1]$ and $f : I \to \mathbb{R}$ is continuous. The following are equivalent:*

1. *The entrywise map $f[-]$ preserves positivity on $\mathbb{P}_n(I)$ for all $n \geq 1$.*
2. *The function f equals a convergent power series $\sum_{k=0}^\infty c_k x^k$ for all $x \in I$, with the Maclaurin coefficients $c_k \geq 0$ for all $k \geq 0$.*

Schoenberg's 1942 paper (in *Duke Math. J.*) is well-known in the analysis literature. In a sense, his Theorem 16.2 is the (harder) converse to the Schur product theorem, i.e., Lemma 16.1, which is the implication (2) $\implies$ (1). Some of these points were discussed in Section 13.2.

Schoenberg's theorem can also be stated for $I = (-1, 1)$. In this setting, the continuity hypothesis was subsequently removed from assertion (1) by Rudin, who, moreover, showed that in order to prove assertion (2) in Theorem 16.2, one does not need to work with the full test set $\bigcup_{n \geq 1} \mathbb{P}_n(I)$. Instead, it is possible to work only with low-rank Toeplitz matrices:

Theorem 16.3 (Rudin, 1959, [306]). *Suppose $I = (-1, 1)$ and $f : I \to \mathbb{R}$. Then the assertions in Schoenberg's theorem 16.2 are equivalent on I, and further equivalent to:*

3. *$f[-]$ preserves positivity on the Toeplitz matrices in $\mathbb{P}_n(I)$ of rank ≤ 3, for all $n \geq 1$.*

Schoenberg's theorem also has a one-sided variant, over the semi-axis $I = (0, \infty)$:

Theorem 16.4 (Vasudeva, 1979, [353]). *Suppose $I = (0, \infty)$ and $f : I \to \mathbb{R}$. Then the two assertions of Schoenberg's theorem 16.2 are equivalent on I as well.*

Our goal in this part of the text is to prove stronger versions of the theorems of Schoenberg and Vasudeva. Specifically, we will (i) remove the continuity hypothesis, and (ii) work with severely reduced test sets in each dimension, consisting of only the *Hankel matrices of rank at most* 3. For instance, we will show Theorem 16.3, but with the word "Toeplitz" replaced by "Hankel." Similarly, we will show a strengthening of Theorem 16.4, using totally non-negative Hankel matrices of rank at most 2. These results are stated and proved in this part of the text.

16.2. **Digression: the complex case.** In the aforementioned 1959 paper in *Duke Math. J.*, Rudin made some observations about the complex case *a la* Pólya–Szegő, and presented a conjecture, which is now explained. First, observe that the Schur product theorem holds for complex Hermitian matrices as well, with the same proof via the spectral theorem:

"*If A, B are $n \times n$ complex (Hermitian) positive semidefinite matrices, then so is $A \circ B$.*"

As a consequence, every monomial $z \mapsto z^k$ preserves positivity on $\mathbb{P}_n(\mathbb{C})$ for all integers $k \geq 0$ and $n \geq 1$. (Here $\mathbb{P}_n(\mathbb{C})$ comprises the complex Hermitian matrices $A_{n\times n}$, such that $u^* A u \geq 0$ for all $u \in \mathbb{C}^n$.) But more is true: the (entrywise) conjugation map also preserves positivity on $\mathbb{P}_n(\mathbb{C})$ for all $n \geq 1$. Now using the Schur product theorem, the functions

$$z \mapsto z^k (\overline{z})^m, \qquad k, m \geq 0$$

each preserve positivity on $\mathbb{P}_n(\mathbb{C})$, for all $n \geq 1$. Since $\mathbb{P}_n(\mathbb{C})$ is easily seen to be a closed convex cone as well, Rudin observed that if a series

$$f(z) = \sum_{k,m \geq 0} c_{k,m} z^k (\overline{z})^m, \qquad \text{with } c_{k,m} \geq 0,$$

is convergent on the open disk $D(0,\rho) := \{z \in \mathbb{C} : |z| < \rho\}$, then $f[-]$ entrywise preserves positivity on $\mathbb{P}_n(D(0,\rho))$ for $n \geq 1$. Rudin conjectured that there are no other preservers. This was proved soon after:

Theorem 16.5 (Herz, *Ann. Inst. Fourier*, 1963)**.** *Suppose $I = D(0,1) \subset \mathbb{C}$ and $f : I \to \mathbb{C}$. The following are equivalent:*

(1) The entrywise map $f[-]$ preserves positivity on $\mathbb{P}_n(I)$ for all $n \geq 1$.
(2) f is of the form $f(z) = \sum_{k,m\geq 0} c_{k,m} z^k (\overline{z})^m$ on I, with $c_{k,m} \geq 0$ for all $k, m \geq 0$.

For completeness, we also point out [276] for a recent, non-commutative variant of the Schur product and Schoenberg's theorem.

The real and complex cases of Schoenberg/Herz's theorems have been since proved using alternate tools. Christensen and Ressel showed Schoenberg's theorem 16.2 using Choquet's representation theorem, in 1978 in *Trans. Amer. Math. Soc.* [82]. In fact, they removed the continuity assumption of Schoenberg, while continuing to work over the domain $[-1,1]$. In this case, the preservers turn out to be almost the same as in Schoenberg's theorem – i.e., limits of functions in the convex hull of the monomials – but now, one needs to include the pointwise limits on $[-1,1]$ of the odd powers and of the even powers:

Theorem 16.6 (Christensen and Ressel, 1978, [82])**.** *Suppose $I = [-1,1]$ and $f : I \to \mathbb{R}$. The following are equivalent:*

(1) The entrywise map $f[-]$ preserves positivity on $\mathbb{P}_n(I)$ for all $n \geq 1$.
(2) The function f equals a power series plus two other terms:

$$f(x) = \sum_{k=0}^{\infty} c_k x^k + c_{-1}[\mathbf{1}_{\{1\}}(x) - \mathbf{1}_{\{-1\}}(x)] + c_{-2}\mathbf{1}_{\{-1,1\}}(x), \qquad x \in I,$$

with the Maclaurin coefficients $c_k \geq 0$ for all $k \geq -2$ and $\sum_{k \geq -2} c_k < \infty$.

Christensen and Ressel also proved the complex analogue: namely, for preservers of positivity on Gram matrices from unit complex spheres, in *Math. Z.* in 1982:

Theorem 16.7 (Christensen and Ressel, 1982, [83])**.** *Suppose $f : \overline{D}(0,1) \to \mathbb{C}$ is continuous on the closed unit disk, and $\mathcal{H}$ is an infinite-dimensional complex Hilbert space with unit sphere S. Then the following are equivalent:*

(1) f is "positive definite" on S, in that for all $n \geq 1$ and points $z_1, \dots, z_n \in S$, the matrix with (j,k)-entry $f(\langle z_j, z_k \rangle)$ is positive semidefinite.
(2) $f(z)$ has the unique series representation $f(z) = \sum_{k,m \geq 0} c_{k,m} z^k (\overline{z})^m$, with all $c_{k,m} \geq 0$ and $\sum_{k,m \geq 0} c_{k,m} < \infty$.

This resembles Herz's theorem 16.5 (which proved Rudin's conjecture) similar to the relation between Schoenberg's theorem 16.2 and Rudin's theorem 16.3.

As a final remark, the Schoenberg/Rudin/Vasudeva/Herz results are reminiscent of an earlier, famous result, by Loewner in the parallel (and well studied) setting of the matrix functional calculus. Namely, given a complex Hermitian matrix A with eigenvalues in a real interval (a,b), a function $f : (a,b) \to \mathbb{R}$ acts on A as follows: let $A = UDU^*$ be a spectral decomposition of A; then $f(A) := Uf(D)U^*$, where $f(D)$ is the diagonal matrix with diagonal entries $f(d_{jj})$. Now, Loewner showed in *Math. Z.* (1934) even before Schoenberg:

Theorem 16.8 (Loewner)**.** *Let $-\infty \leq a < b \leq \infty$, and $f : (a,b) \to \mathbb{R}$. The following are equivalent:*

(1) f is matrix monotone*: if $A \leq B$ are square matrices with eigenvalues in (a,b), then $f(A) \leq f(B)$.*
(2) f is C^1 on (a,b), and given $a < x_1 < \cdots < x_k < b$ for any $k \geq 1$, the Loewner matrix *given by $L_f(x_j, x_k) := \frac{f(x_j) - f(x_k)}{x_j - x_k}$ if $j \neq k$, else $f'(x_j)$, is positive semidefinite.*
(3) There exist real constants $p \geq 0$ and q, and a finite measure μ on $\mathbb{R} \setminus (a,b)$, such that
$$f(x) = q + px + \int_{\mathbb{R} \setminus (a,b)} \frac{1 + xy}{y - x} \, d\mu(y).$$
(4) There exists a function $\widetilde{f}$ that is analytic on $(\mathbb{C} \setminus \mathbb{R}) \sqcup (a,b)$, such that (a) $f \equiv \widetilde{f}|_{(a,b)}$ and (b) if $\Im z > 0$ then $\Im \widetilde{f}(z) > 0$.

Notice similar to the preceding results, the emergence of analyticity from the dimension-free preservation of a matrix property. (In fact, one shows that Loewner monotone functions on $n \times n$ matrices are automatically C^{2n-3}.) This is also the case with a prior result of Rudin with Helson, Kahane, and Katznelson in *Acta Math.* in 1959, which directly motivated Rudin's 1959 paper discussed above. (The bibliographic notes at the end of this part provide a few more details.)

16.3. **Origins of positive matrices: Menger, Fréchet, Schoenberg, and metric geometry.** In this subsection and the next two, we study some of the historical origins of positive (semi)definite matrices. This class of matrices of course arises as Hessians of twice-differentiable functions at local minima; however, the branch of early twentieth-century mathematics that led to the development of positivity preservers is *metric geometry.* More precisely, the notion of a metric space – emerging from the works of Fréchet and Hausdorff – and isometric embeddings of such structures into Euclidean and Hilbert spaces, spheres, and hyperbolic and homogeneous spaces, were studied by Schoenberg, Bochner, and von Neumann among others; and it is this work that led to the study of matrix positivity and its preservation.

Definition 16.9. A *metric space* is a set X together with a metric $d : X \times X \to \mathbb{R}$, satisfying:

(1) *Positivity:* $d(x, y) \geq 0$ for all $x, y \in X$, with equality if and only if $x = y$.
(2) *Symmetry:* $d(x, y) = d(y, x)$ for all $x, y \in X$.
(3) *Triangle inequality:* $d(x, z) \leq d(x, y) + d(y, z)$ for all $x, y, z \in X$.

In this section, we will state and prove three results by Schoenberg, which explain his motivations in studying positivity and its preservers, and serve to illustrate the (by now well explored) connection between metric geometry and matrix positivity. We begin with a sample result on metric embeddings, shown by Fréchet in *Math. Ann.* in 1910: *If (X, d) is a metric space with $|X| = n + 1$, then (X, d) isometrically embeds into $(\mathbb{R}^n, \|\cdot\|_\infty)$.*

Such results led to exploring which metric spaces isometrically embed into Euclidean spaces. Specifically, in Menger's 1931 paper [260] in *Amer. J. Math.*, and Fréchet's 1935 paper [131] in *Ann. of Math.*, the authors explored the following question: *Given integers $n, r \geq 1$, characterize the tuples of $\binom{n+1}{2}$ positive real numbers that can denote the distances between the vertices of an $(n + 1)$-simplex in $\mathbb{R}^r$ but not in $\mathbb{R}^{r-1}$.* In other words, given a finite metric space X, what is the smallest r, if any, such that X isometrically embeds into $\mathbb{R}^r$?

In his 1935 paper in *Ann. of Math.*, Schoenberg gave an alternate characterization of all such "admissible" tuples of distances. This characterization used ... matrix positivity!

Theorem 16.10 (Schoenberg, 1935, [313])**.** *Fix integers $n, r \geq 1$, and a finite set $X = \{x_0, \dots, x_n\}$ together with a metric d on X. Then (X, d) isometrically embeds into some $\mathbb{R}^r$ (with the Euclidean distance/norm), if and only if the $n \times n$ matrix*

$$A := (d(x_0, x_j)^2 + d(x_0, x_k)^2 - d(x_j, x_k)^2)_{j,k=1}^n \tag{16.11}$$

is positive semidefinite. Moreover, the smallest such r is precisely the rank of the matrix A.

This classical theorem is at the heart of multidimensional scaling; see, e.g., [90]. Additionally, the matrix A features later in this text when we study a result of Menger in Appendix E; it is an alternate form of the *Cayley–Menger matrix* associated to the metric space X. See Section 27, where we also extend this theorem to embeddings of separable metric spaces.

Proof. If (X, d) isometrically embeds into $(\mathbb{R}^r, \|\cdot\| = \|\cdot\|_2)$, then

$$\begin{aligned} d(x_0, x_j)^2 + d(x_0, x_k)^2 - d(x_j, x_k)^2 &= \|x_0 - x_j\|^2 + \|x_0 - x_k\|^2 - \|(x_0 - x_j) - (x_0 - x_k)\|^2 \\ &= 2\langle x_0 - x_j, x_0 - x_k \rangle. \end{aligned} \tag{16.12}$$

But then the matrix A in (16.11) is the Gram matrix of a set of vectors in $\mathbb{R}^r$, and hence, is positive semidefinite. In the rest of this section, we use Theorem 2.5 and Proposition 2.15 (and their proofs) without further reference. Thus, $\frac{1}{2}A = B^T B$, where the columns of B are $x_0 - x_j \in \mathbb{R}^r$. But then A has rank at most the rank of B, hence at most r. Since (X, d) does not embed in $\mathbb{R}^{r-1}$, by the same argument A has rank precisely r.

Conversely, suppose the matrix A in (16.11) is positive semidefinite of rank r. First, consider the case when $r = n$, i.e., A is positive definite. By Theorem 2.5, $\frac{1}{2}A = B^T B$ for a square invertible matrix B. Thus, left-multiplication by B sends the r-simplex with vertices $\mathbf{0}, \mathbf{e}_1, \dots, \mathbf{e}_r$ to an r-simplex, where $\mathbf{e}_j$ comprise the standard basis of $\mathbb{R}^r$.

Now claim that the assignment $x_0 \mapsto \mathbf{0},\ x_j \mapsto B\mathbf{e}_j$ for $1 \leq j \leq r$, is an isometry : $X \to \mathbb{R}^r$ whose image, being the vertex set of an r-simplex, necessarily cannot embed inside $\mathbb{R}^{r-1}$. Indeed, computing these distances proves the claim, and hence the theorem for $r = n$:

$$\begin{aligned} d(B\mathbf{e}_j, \mathbf{0})^2 &= \|B\mathbf{e}_j\|^2 = \frac{1}{2}\mathbf{e}_j^T A \mathbf{e}_j = \frac{a_{jj}}{2} = d(x_0, x_j)^2, \\ d(B\mathbf{e}_j, B\mathbf{e}_k)^2 &= \|B\mathbf{e}_j - B\mathbf{e}_k\|^2 = \frac{a_{jj} + a_{kk}}{2} - a_{jk} \\ &= d(x_0, x_j)^2 + d(x_0, x_k)^2 - (d(x_0, x_j)^2 + d(x_0, x_k)^2 - d(x_j, x_k)^2) = d(x_j, x_k)^2. \end{aligned}$$

Next, suppose $r < n$. Then $\frac{1}{2}A = B^T P B$ for some invertible matrix B, where P is the projection operator $\begin{pmatrix} \mathrm{Id}_{r \times r} & 0 \\ 0 & 0_{(n-r)\times(n-r)} \end{pmatrix}$. Let $\Delta := \{PB\mathbf{e}_j,\ 1 \leq j \leq n\} \sqcup \{\mathbf{0}\}$ denote the projection under P of the vertices of an $(n+1)$ simplex. Repeating the above proof shows that the map : $x_0 \mapsto \mathbf{0},\ x_j \mapsto PB\mathbf{e}_j$ for $1 \leq j \leq n$, is an isometry from X onto Δ. By construction, Δ lies in the image of the projection P, hence in a copy of $\mathbb{R}^r$. But being the image under P of the vertex set of an n-simplex, Δ cannot lie in a copy of $\mathbb{R}^{r-1}$ (otherwise so would its span, which is all of $P(\mathbb{R}^n) \cong \mathbb{R}^r$). $\square$

We end this part with an observation. A (real symmetric square) matrix $A'_{(n+1)\times(n+1)}$ is said to be *conditionally positive semidefinite* if $(\mathbf{u}')^T A' \mathbf{u}' \geq 0$ whenever $\sum_{j=0}^n u'_j = 0$. Such matrices are also studied in the literature (though not as much as positive semidefinite matrices). The following lemma reformulates Theorem 16.10 into the conditional positivity of a related matrix:

Lemma 16.13. *Let $X = \{x_0, \dots, x_n\}$ be a finite set equipped with a metric d. Then the matrix $A_{n\times n}$ as in* (16.11) *is positive semidefinite if and only if the $(n+1)\times(n+1)$ matrix*

$$A' := (-d(x_j, x_k)^2)_{j,k=0}^n \tag{16.14}$$

is conditionally positive semidefinite.

In particular, Schoenberg's papers in the 1930s feature both positive semidefinite matrices (Theorem 16.10) and conditionally positive semidefinite matrices (Theorem 16.17). Certainly, the former class of matrices were a popular and recurrent theme in the analysis literature, with contributions from Carathéodory, Hausdorff, Hermite, Nevanlinna, Pick, Schur, and many others.

Proof of Lemma 16.13. Let $u_1, \dots, u_n \in \mathbb{R}$ be arbitrary, and set $u_0 := -(u_1 + \cdots + u_n)$. Defining $\mathbf{u} := (u_1, \dots, u_n)^T$ and $\mathbf{u}' := (u_0, \dots, u_n)^T$, we compute using that the diagonal entries of A' are zero:

$$\begin{aligned} (\mathbf{u}')^T A' \mathbf{u}' = &\sum_{k=1}^n \left(\sum_{j=1}^n u_j \right) d(x_0, x_k)^2 u_k + \sum_{j=1}^n u_j d(x_j, x_0)^2 \left(\sum_{k=1}^n u_k \right) - \sum_{j,k=1}^n u_j d(x_j, x_k)^2 u_k \\ = &\sum_{j,k=1}^n u_j u_k \left(d(x_0, x_k)^2 + d(x_j, x_0)^2 - d(x_j, x_k)^2 \right) = \mathbf{u}^T A \mathbf{u}, \end{aligned}$$

for all $\mathbf{u} \in \mathbb{R}^n$. This proves the result. $\square$

16.4. **Origins of positivity preservers: Schoenberg, Bochner, and positive definite functions.** We continue with our historical journey, this time into the origins of the entrywise calculus on positive matrices. As Theorem 16.10 and Lemma 16.13 show, applying entrywise the function $-x^2$ to any distance matrix $(d(x_j, x_k))_{j,k=0}^n$ from Euclidean space yields a conditionally positive semidefinite matrix A'.

It is natural to want to remove the word "conditionally" from the above result. Namely: which entrywise maps send distance matrices to positive semidefinite matrices? These are precisely the positive definite functions:

Definition 16.15. Given a metric space (X, d), a function $f : [0, \infty) \to \mathbb{R}$ is *positive definite on* X if for any finite set of points $x_1, x_2, \dots x_n \in X$, the matrix $f[(d(x_j, x_k))_{j,k=1}^n]$ is positive semidefinite.

By considering 2×2 distance matrices, note that positive definite functions are not the same as positivity preservers; no distance matrix is positive semidefinite unless all x_j are equal (in which case we get the zero matrix). On a different note, given any metric space (X, d), the positive definite functions on X form a closed convex cone, by Lemma 3.1.

In arriving at Theorem 16.2, Schoenberg was motivated by metric geometry – as we just studied – as well as the study of positive definite functions. The latter was also of interest to other mathematicians in that era: Bochner, Pólya, and von Neumann, to name a few. In fact, positive definite functions are what led to Schoenberg's Theorem 16.2 and the development of the entrywise calculus. Note that Bochner – and previously Carathéodory, Herglotz, Mathias, and others – studied functions on groups G that were positive definite in the "more standard" sense – namely, where in the above definition $f : G \to \mathbb{C}$, and one substitutes $d(x_j, x_k)$ by $x_j^{-1} x_k$. The above definition seems due to Schoenberg, in his 1938 paper in *Trans. Amer. Math. Soc.*

We now present – from this paper – another characterization by Schoenberg of metric embeddings into a Euclidean space $\mathbb{R}^r$, this time via positive definite functions. This requires a preliminary observation involving the positive definiteness of an even kernel:

Lemma 16.16. *Given $\sigma > 0$, the Gaussian kernel $T_{G_\sigma}(x, y) := \exp(-\sigma \|x - y\|^2)$ – in other words, the function $\exp(-\sigma x^2)$ – is positive definite on $\mathbb{R}^r$ for all $r \geq 1$.*

Proof. Observe that the case of $\mathbb{R}^r$ for general r follows from the $r = 1$ case, via the Schur product theorem. In turn, the $r = 1$ case is a consequence of Pólya's lemma 6.8. ☐

The following result of Schoenberg in *Trans. Amer. Math. Soc.* relates metric space embeddings with this positive definiteness of the Gaussian kernel:

Theorem 16.17 (Schoenberg, 1938, [317])**.** *A finite metric space (X, d) with $X = \{x_0, \dots, x_n\}$ embeds isometrically into $\mathbb{R}^r$ for some $r > 0$ (which turns out to be at most n), if and only if the $(n+1) \times (n+1)$ matrix with (j, k) entry*

$$\exp(-\sigma^2 d(x_j, x_k)^2), \qquad 0 \leq j, k \leq n$$

is positive semidefinite, along any sequence of non-zero scalars σ_m decreasing to 0^+ (equivalently, for all $\sigma \in \mathbb{R}$).

For another application of this result and those in the previous Section 16.3, see Section 27.3.

Proof. Clearly, if (X, d) embeds isometrically into $\mathbb{R}^r$, then identifying the x_j with their images in $\mathbb{R}^r$, it follows by Lemma 16.16 that the matrix with (j,k) entry

$$\exp(-\sigma^2 \|x_j - x_k\|^2) = T_{G_1}(\sigma x_j, \sigma x_k)$$

is positive semidefinite for all $\sigma \in \mathbb{R}$.

Conversely, let $\sigma_m \downarrow 0^+$. From the positivity of the exponentiated distance matrices for σ_m, it follows for any vector $u := (u_0, \dots, u_n)^T \in \mathbb{R}^{n+1}$ that

$$\sum_{j,k=0}^{n} u_j u_k \exp(-\sigma_m^2 d(x_j, x_k)^2) \geq 0.$$

Expanding into Taylor series and interchanging the infinite sum with the two finite sums,

$$\sum_{l=0}^{\infty} \frac{(-\sigma_m^2)^l}{l!} \sum_{j,k=0}^{n} u_j u_k d(x_j, x_k)^{2l} \geq 0, \qquad \forall m \geq 1.$$

Suppose we restrict to the vectors $\mathbf{u}'$ satisfying: $\sum_{j=0}^n u_j = 0$. Then the $l = 0$ term vanishes. Now dividing throughout by σ_m^2 and taking $m \to \infty$, the "leading term" in σ_m must be non-negative. It follows that if $A' := (-d(x_j, x_k)^2)_{j,k=0}^n$, then $(\mathbf{u}')^T A' \mathbf{u}' \geq 0$ whenever $\sum_j u_j = 0$. By Lemma 16.13 and Theorem 16.10, (X, d) embeds isometrically into $\mathbb{R}^r$, where $r \leq n$ denotes the rank of the matrix $A_{n \times n}$ as in (16.11). $\square$

16.5. **Schoenberg: from spheres to correlation matrices, to positivity preservers.** The previous result, Theorem 16.17, says that Euclidean spaces $\mathbb{R}^r$ – or their direct limit / union $\mathbb{R}^\infty$ (which should more accurately be denoted $\mathbb{R}^{\mathbb{N}}$), or even its completion ℓ^2 of square-summable real sequences (which Schoenberg and others called *Hilbert space*) – can be characterized by the property that the maps

$$\exp(-\sigma^2 x^2), \qquad \sigma \in (0, \rho) \tag{16.18}$$

are all positive definite on each (finite) metric subspace. As we saw, such a characterization holds for each $\rho > 0$.[1]

Given this characterization, it is natural to seek out similar characterizations of distinguished submanifolds M in $\mathbb{R}^r$ or $\mathbb{R}^\infty$ or ℓ^2. In fact, in the aforementioned 1935 *Ann. of Math.* paper, Schoenberg showed the first such classification result, for $M = S^{r-1}$ a unit sphere – as well as for the *Hilbert sphere* S^∞. Note here that the unit sphere $S^{r-1} := \{x \in \mathbb{R}^r : \|x\|^2 = 1\}$, while the Hilbert sphere $S^\infty \subset \ell^2$ is the subset of all square-summable sequences with unit ℓ^2-norm. (This is the *closure* of the set of all real sequences with finitely many non-zero coordinates and unit ℓ^2-norm – which is the unit sphere $\bigcup_{r \geq 1} S^{r-1}$ in $\bigcup_{r \geq 1} \mathbb{R}^r$.)

One defines a rotationally invariant metric on S^∞ (hence on each S^{r-1}) as follows. The distance between x and $-x$ is π, and given points $x \neq \pm y$ in S^∞, there exists a unique plane passing through x, y, and the origin. This plane intersects the sphere S^∞ in a unit circle S^1 given by

$$\{\alpha x + \beta y : \alpha, \beta \in \mathbb{R},\ 1 = \|\alpha x + \beta y\|^2 = \alpha^2 + \beta^2 + 2\alpha\beta\langle x, y\rangle\} \subset S^\infty,$$

and we let $d(x, y)$ denote the angle – i.e., arclength – between x and y:

$$d(x, y) := \sphericalangle(x, y) = \arccos(\langle x, y\rangle) \in [0, \pi].$$

[1] A related result on positive definite functions on – or Hilbert space embeddings of – a *topological* space X is by Kolmogorov around 1940 [223]. He showed that a continuous function $K : X \times X \to \mathbb{C}$ is positive definite if and only if there exists a Hilbert space $\mathcal{H}$ and a norm-continuous map $\varphi : X \to \mathcal{H}$, such that $K(x_1, x_2) = \langle \varphi(x_1), \varphi(x_2)\rangle$ for all $x_1, x_2 \in X$.

Now we come to Schoenberg's characterization for metric embeddings into Euclidean spheres. He showed that in contrast to the family (16.18) of positive definite functions for Euclidean spaces, for spheres it suffices to consider *a single function*! This function is the cosine:

Proposition 16.19 (Schoenberg, 1935, [313])**.** *Let (X, d) be a finite metric space with $X = \{x_1, \dots, x_n\}$. Fix an integer $r \geq 2$. Then X isometrically embeds into S^{r-1} but not S^{r-2}, if and only if $d(x_j, x_k) \leq \pi$ for all $1 \leq j, k \leq n$ and the matrix $(\cos d(x_j, x_k))_{j,k=1}^n$ is positive semidefinite of rank r.*

In particular, X embeds isometrically into the Hilbert sphere S^∞ – with the spherical metric – if and only if (a) $\operatorname{diam}(X) \leq \pi$ and (b) $\cos(\cdot)$ is positive definite on X.

Thus, matrix positivity is also intimately connected with spherical embeddings, which may not be surprising given Theorem 16.10.

Proof. If there exists an isometric embedding $\varphi : X \hookrightarrow S^{r-1}$ as claimed, we have as above

$$\cos d(x_j, x_k) = \cos \sphericalangle(\varphi(x_j), \varphi(x_k)) = \langle \varphi(x_j), \varphi(x_k) \rangle,$$

which yields a Gram matrix of rank at most r, hence exactly r (since X does not embed isometrically into S^{r-2}). Moreover, the spherical distance between x_j, x_k (for $j, k > 0$) is at most π, as desired.

Conversely, since $A := (\cos d(x_j, x_k))_{j,k=1}^n$ is positive, it is a Gram matrix (of rank r), and hence $A = B^T B$ for some $B_{r \times n}$ of rank r by Theorem 2.5. Let $y_j \in \mathbb{R}^r$ denote the columns of B; then clearly $y_j \in S^{r-1}\ \forall j$; moreover,

$$\cos \sphericalangle(y_j, y_k) = \langle y_j, y_k \rangle = a_{jk} = \cos d(x_j, x_k), \qquad \forall 1 \leq j, k \leq n.$$

Since $d(x_j, x_k)$ lies in $[0, \pi]$ by assumption, as does $\sphericalangle(y_j, y_k)$, we obtain an isometry $\varphi : X \to S^{r-1}$, sending $x_j \mapsto y_j$ for all $j > 0$. Finally, $\operatorname{im}(\varphi)$ is not contained in S^{r-2}, for otherwise A would have rank at most $r - 1$. This shows the result for S^{r-1}; the case of S^∞ is similar. $\square$

The proof of Proposition 16.19 shows that $\cos(\cdot)$ is a positive definite function on unit spheres of all dimensions.

Note that Proposition 16.19 and the preceding two theorems by Schoenberg in the 1930s:

(1) characterize metric space embeddings into Euclidean spaces via matrix positivity;
(2) characterize metric space embeddings into Euclidean spaces via the positive definite functions $\exp(-\sigma^2(\cdot)^2)$ on $\mathbb{R}^r$ or $\mathbb{R}^\infty$ (so this involves positive matrices); and
(3) characterize metric space embeddings into Euclidean spheres S^{r-1} or S^∞ (with the spherical metric) via the positive definite function $\cos(\cdot)$ on S^∞.

Around the same time (in the 1930s), S. Bochner [56, 57] had classified all of the positive definite functions on $\mathbb{R}$. This result was extended in 1940 simultaneously by Weil, Povzner, and Raikov to classify the positive definite functions on any locally compact abelian group. Amid this backdrop, in his 1942 paper [318] Schoenberg was interested in understanding the positive definite functions of the form $f \circ \cos : [-1, 1] \to \mathbb{R}$ on a unit sphere $S^{r-1} \subset \mathbb{R}^r$, where $r \geq 2$.

To present Schoenberg's result, first consider the $r = 2$ case. As mentioned above, distance (i.e., angle) matrices are not positive semidefinite; but if one applies the cosine function entrywise, then we obtain the matrix with (j, k) entry $\cos(\theta_j - \theta_k)$, and this is positive semidefinite by Lemma 2.17. But now $f[-]$ preserves positivity on a set of Toeplitz matrices (among others), by Lemma 2.17 and the subsequent discussion. For general dimension $r \geq 2$, we have $\cos(d(x_j, x_k)) = \langle x_j, x_k \rangle$ (see also the proof of Proposition 16.19), so $\cos[(d(x_j, x_k))_{j,k}]$

always yields Gram matrices. Hence, $f[-]$ would once again preserve positivity on a set of positive matrices. It was this class of functions that Schoenberg characterized:

Theorem 16.20 (Schoenberg, 1942, [318]). *Suppose $f : [-1,1] \to \mathbb{R}$ is continuous, and $r \geq 2$ is an integer. Then the following are equivalent:*

(1) $(f \circ \cos)$ is positive definite on S^{r-1}.

(2) The function $f(x) = \sum_{k=0}^{\infty} c_k C_k^{(\frac{r-2}{2})}(x)$, where $c_k \geq 0, \forall k$, and $\{C_k^{(\frac{r-2}{2})}(x) : k \geq 0\}$ comprise the first Chebyshev or Gegenbauer family of orthogonal polynomials.

Remark 16.21. Theorem 16.20 has an interesting reformulation in terms of entrywise positivity preservers on correlation matrices. Recall that on the unit sphere S^{r-1}, applying $\cos[-]$ entrywise to a distance matrix of points x_j yields precisely the Gram matrix with entries $\langle x_j, x_k \rangle$, which is positive of rank at most r. Moreover, as the vectors x_j lie on the unit sphere, the diagonal entries are all 1 and hence we obtain a correlation matrix. Putting these facts together, $f \circ \cos$ is positive definite on S^{r-1} if and only $f[-]$ preserves positivity on all correlation matrices of arbitrary size but rank at most r. Thus, Schoenberg's works in 1935 [313] and 1942 [318] already contained connections to entrywise preservers of correlation matrices, which brings us around to the modern-day motivations that arise from precisely this question (now arising in high-dimensional covariance estimation, and discussed in Section 13.1).

Remark 16.22. Schoenberg's work has been followed by numerous papers attempting to understand positive definite functions on locally compact groups, spheres, two-point homogeneous metric spaces, and products of these. See, for example, [19, 21, 22, 41, 42, 58, 79, 111, 150, 256, 257, 373, 374, 379] for a selection of works. The connection to spheres has also led to work in statistics on spatio-temporal covariance functions on spheres, modeling the earth as a sphere [144, 291, 367]. (Note that a metric space X is n-point homogeneous [361] if for all $1 \leq p \leq n$ and subsets $X_1, X_2 \subset X$ of size p, every isometry : $X_1 \to X_2$ extends to a self-isometry of X. This was first studied by Birkhoff [50] and differs from the more widespread usage for spaces G/H. We will study this further in Section 27.)

Remark 16.23. Schoenberg's (and subsequent) work on finite- and infinite-dimensional spheres has many other applications. One area of recent activity involves sphere packing, spherical codes, and configurations of points on spheres that maximize the minimal distance or some potential function. See, for example, the work of Cohn with coauthors in *J. Amer. Math. Soc.* 2007, 2012 [84, 85] and in *Duke Math. J.* 2014, 2018 [86, 87]; and Musin in *Ann. of Math.* 2008 [268].

Returning to the above discussion on Theorem 16.20, if instead we let $r = \infty$, then the corresponding result would classify positivity preservers on all correlation matrices (without rank constraints) by Remark 16.21. And indeed, Schoenberg achieves this goal in the same paper:

Theorem 16.24 (Schoenberg, 1942, [318]). *Suppose $f : [-1,1] \to \mathbb{R}$ is continuous. Then $f \circ \cos$ is positive definite on S^∞ if and only if there exist scalars $c_k \geq 0$, such that*

$$f(\cos\theta) = \sum_{k \geq 0} c_k \cos^k \theta, \qquad \theta \in [0, \pi].$$

Notice here that $\cos^k \theta$ is positive definite on S^∞ for all integers $k \geq 0$, by Proposition 16.19 and the Schur product theorem. Hence, so is $\sum_{k\geq 0} c_k \cos^k \theta$ if all $c_k \geq 0$.

Freed from the sphere context, the preceding theorem says that a continuous function $f : [-1,1] \to \mathbb{R}$ preserves positivity when applied entrywise to all correlation matrices, if and only if $f(x) = \sum_{k \geq 0} c_k x^k$ on $[-1,1]$ with all $c_k \geq 0$. This finally explains how and why Schoenberg arrived at his celebrated converse to the Schur product theorem – namely, Theorem 16.2 on entrywise positivity preservers.

16.6. **Digression on ultraspherical polynomials.** Before proceeding further, we describe the orthogonal polynomials $C_k^{(\alpha)}(x)$ for $k \geq 0$, where $\alpha = \alpha(r) = (r-2)/2$. Given $r \geq 2$, note that $\alpha(r)$ ranges over the non-negative half-integers. Though not used below, here are several different (equivalent) definitions of the polynomials $C_k^{(\alpha)}$ for general real $\alpha \geq 0$.

First, if $\alpha = 0$, then $C_k^{(0)}(x) := T_k(x)$, the Chebyshev polynomials of the first kind

$$T_0(x) = 1, \quad T_1(x) = x, \quad T_2(x) = 2x^2 - 1, \quad \dots, \quad T_k(\cos(\theta)) = \cos(k\theta) \ \forall k \geq 0.$$

A second way to compute the polynomials $C_k^{(0)}(x)$ is through their *generating function*

$$\frac{1 - xt}{1 - 2xt + t^2} = \sum_{k=0}^{\infty} C_k^{(0)}(x) t^k.$$

For higher α: setting $\alpha = \frac{1}{2}$ yields the family of Legendre polynomials. If $\alpha = 1$, we obtain the Chebyshev polynomials of the second kind. For general $\alpha > 0$, the functions $(C_k^{(\alpha)}(x))_{k \geq 0}$ are the *Gegenbauer/ultraspherical polynomials*, defined via their generating function

$$(1 - 2xt + t^2)^{-\alpha} = \sum_{k=0}^{\infty} C_k^{(\alpha)}(x) t^k.$$

For all $\alpha \geq 0$, the polynomials $(C_k^{(\alpha)}(x))_{k \geq 0}$ form a complete orthogonal set in the Hilbert space $L^2([-1,1], w_\alpha)$, where w_α is the weight function

$$w_\alpha(x) := (1 - x^2)^{\alpha - \frac{1}{2}}, \qquad x \in (-1,1).$$

Thus, another definition of $C_k^{(\alpha)}(x)$ is that it is a polynomial of degree k, with $C_0^{(\alpha)}(x) = 1$, and such that the $C_k^{(\alpha)}$ are orthogonal with respect to the bilinear form

$$\langle f, g \rangle := \int_{-1}^{1} f(x) g(x) w_\alpha(x) \, dx, \qquad f, g \in L^2([-1,1], w_\alpha),$$

and satisfy:

$$\langle C_k^{(\alpha)}, C_k^{(\alpha)} \rangle = \frac{\pi 2^{1-2\alpha} \Gamma(k + 2\alpha)}{k!(k+\alpha)(\Gamma(\alpha))^2}.$$

Yet another definition is that the Gegenbauer polynomials $C_k^{(\alpha)}(x)$ for $\alpha > 0$ satisfy the differential equation

$$(1 - x^2) y'' - (2\alpha + 1) x y' + k(k + 2\alpha) y = 0.$$

We also have a direct formula

$$C_k^{(\alpha)}(x) := \sum_{j=0}^{\lfloor k/2 \rfloor} (-1)^j \frac{\Gamma(k - j + \alpha)}{\Gamma(\alpha) j! (k - 2j)!} (2x)^{k-2j},$$

as well as a recursion

$$C_0^{(\alpha)}(x) := 1, \qquad C_1^{(\alpha)}(x) := 2\alpha x,$$
$$C_k^{(\alpha)}(x) := \frac{1}{k}\left(2x(k+\alpha-1)C_{k-1}^{(\alpha)}(x) - (k+2\alpha-2)C_{k-2}^{(\alpha)}(x)\right) \quad \forall k \geq 2.$$

16.7. **Sketch of proof of Theorem 16.20.** Schoenberg's theorem 16.20 has subsequently been studied by many authors, and in a variety of settings over the years. This includes classifying the positive definite functions on different kinds of spaces: locally compact groups, spheres, and products of these. We next give a proof-sketch of this result. In what follows, we use without reference the observation that (akin to Lemma 3.1), the set of functions f, such that $f \circ \cos$ is positive definite on S^{r-1}, also forms a closed convex cone, which is, moreover, closed under entrywise products.

We first outline why (2) $\implies$ (1) in Theorem 16.20. By the above observation, it suffices to show that $C_k^{(\alpha)} \circ \cos$ is positive definite on S^{r-1}. The proof is by induction on r. For the base case $r = 2$, let $\theta_1, \theta_2, \dots, \theta_n \in S^1 = [0, 2\pi)$. Up to the sign, their distance matrix has (i,j) entry $d(\theta_i, \theta_j) = \theta_i - \theta_j$ (or a suitable translate modulo 2π). Now, by Lemma 2.17, the matrix $(\cos(k(\theta_i - \theta_j)))_{i,j=1}^n$ is positive semidefinite. But this is precisely the matrix obtained by applying $C_k^{(0)} \circ \cos$ to the distance matrix above. This proves one implication for $d = 2$. The induction step (for general $r \geq 2$) follows from addition formulas for $C_k^{(\alpha)}$.

For the converse implication, set $\alpha := (r-2)/2$ and note that $f \in L^2([-1,1], w_\alpha)$. Hence, f has a series expansion $\sum_{k=0}^\infty c_k C_k^{(\alpha)}(x)$, with $c_k \in \mathbb{R}$. Now recover the c_k via

$$c_k = \int_{-1}^1 f(x) C_k^{(\alpha)}(x) w_\alpha(x)\, dx,$$

since the $C_k^{(\alpha)}$ form an orthonormal family. Note that $C_k^{(\alpha)}$ and f are both positive definite (upon precomposing with the cosine function), hence so is their product by the Schur product theorem. A result of W.H. Young now shows that $c_k \geq 0$ for all $k \geq 0$. $\square$

16.8. **Entrywise preservers in a fixed dimension.** We conclude by discussing a natural mathematical refinement of Schoenberg's theorem:

"*Which functions entrywise preserve positivity in a **fixed** dimension?*"

This turns out to be a challenging, yet important question from the point of view of applications (see Section 13.1 for more on this.) In particular, note that there exist functions which preserve positivity on $\mathbb{P}_n$ but not on $\mathbb{P}_{n+1}$: the power functions x^α with $\alpha \in (n-3, n-2)$ for $n \geq 3$, by Theorem 9.3. By Vasudeva's theorem 16.4, it follows that these non-integer power functions cannot be absolutely monotonic.

Surprisingly, while Schoenberg's theorem is classical and provides a complete description in the dimension-free case, not much is known about the fixed-dimension case: namely, the classification of functions $f : I \to \mathbb{R}$, such that $f[-] : \mathbb{P}_n(I) \to \mathbb{P}_n(\mathbb{R})$ for a fixed integer $n \geq 1$.

- If $n = 1$, then clearly, any function $f : [0, \infty) \to [0, \infty)$ works.
- For $n = 2$ and $I = (0, \infty)$, these are precisely the functions $f : (0, \infty) \to \mathbb{R}$ that are non-negative, non-decreasing, and multiplicatively mid-convex. This was shown by Vasudeva (see Theorem 12.7), and it implies similar results for $I = [0, \infty)$ and $I = \mathbb{R}$.
- For every integer $n \geq 3$, the question is *open* to date.

Given the scarcity of results in this direction, a promising line of attack has been to study refinements of the problem. These can involve restricting the test set of matrices in fixed dimension (say, under rank or sparsity constraints) or the test set of functions (say, to only the entrywise powers) as was studied in Sections 9, 13, 14, and 15; or to use both restrictions. See Section 13.2 for more on this discussion, as well as the final part of the text, where we study *polynomial* preservers in a fixed dimension.

To conclude: while the general problem in fixed dimension $n \geq 3$ is open to date, there is a known result: a necessary condition satisfied by positivity preservers on $\mathbb{P}_n$, shown by R.A. Horn in his 1969 paper [182] in *Trans. Amer. Math. Soc.* and attributed to his advisor, Loewner. The result is over 50 years old; yet even today, it remains essentially the only known result for general preservers f on $\mathbb{P}_n$. In the next two sections, we will state and prove this result – in fact, a stronger version. We will then show (stronger versions of) Vasudeva's and Schoenberg's theorems, via a different approach than the one by Schoenberg, Rudin, or others: we crucially use the fixed-dimension theory, via the result of Horn and Loewner.

17. Horn's thesis: a determinant calculation. Proof for smooth functions.

As mentioned in the previous section 16, the goal in this part is to prove a stronger form of Schoenberg's theorem 16.2, in the spirit of Rudin's theorem 16.3 but replacing the word "Toeplitz" by "Hankel". In order to do so, we will first prove a stronger version of Vasudeva's theorem 16.4, in which the test set is once again reduced to only low-rank Hankel matrices.

In turn, our proof of this version of Vasudeva's theorem relies on a fixed-dimension result, alluded to at the end of Section 16. Namely, we state and prove a stronger version of a 1969 theorem of Horn (attributed by him to Loewner), in this section and the next (Section 18).

Theorem 17.1 (Horn–Loewner theorem, stronger version)**.** *Let $I = (0, \infty)$, and fix $u_0 \in (0,1)$ and an integer $n \geq 1$. Define $\mathbf{u} := (1, u_0, \dots, u_0^{n-1})^T$. Suppose $f : I \to \mathbb{R}$ is such that $f[-]$ preserves positivity on the set $\{a\mathbf{1}_{n\times n} + b\mathbf{u}\mathbf{u}^T : a, b > 0\}$ as well as on the rank-1 matrices in $\mathbb{P}_2(I)$ and the Toeplitz matrices in $\mathbb{P}_2(I)$. Then*

1. *$f \in C^{n-3}(I)$ and $f, f', \dots, f^{(n-3)}$ are non-negative on I. Moreover, $f^{(n-3)}$ is convex and non-decreasing on I.*
2. *If, moreover, $f \in C^{n-1}(I)$, then $f^{(n-2)}, f^{(n-1)}$ are also non-negative on I.*

All test matrices here are Hankel of rank ≤ 2 – and are, moreover, totally non-negative by Corollary 4.3 since they arise as the truncated moment matrices of the measures $a\delta_1 + b\delta_{u_0}$. This is used later in this part, to prove stronger versions of Vasudeva's and Schoenberg's theorems (see Section 19), with similarly reduced test sets of low-rank Hankel matrices.

Remark 17.2. In the original result by Horn (and Loewner), f was assumed to be continuous and to preserve positivity on all of $\mathbb{P}_N((0,\infty))$. In Theorem 17.1, we have removed the continuity hypothesis, in the spirit of Rudin's work, and also greatly reduced the test set.

Remark 17.3. We also observe that Theorem 17.1 is the "best possible," in that the number of non-zero derivatives that must be positive is sharp. For example, let $n \geq 2$, $I = (0, \infty)$, and $f : I \to \mathbb{R}$ be given by: $f(x) := x^\alpha$, where $\alpha \in (n-2, n-1)$. Using Theorem 9.3, $f[-]$ preserves positivity on the test sets $\{a\mathbf{1}_{n\times n} + b\mathbf{u}\mathbf{u}^T : a, b > 0\}$ and $\mathbb{P}_2(I)$. Moreover, $f \in C^{n-1}(I)$ and $f, f', \dots, f^{(n-1)}$ are strictly positive on I. However, $f^{(n)}$ is negative on I.

This low-rank Hankel example (and more generally, Theorem 9.3) also shows that there exist (power) functions that preserve positivity on $\mathbb{P}_n$ but not on $\mathbb{P}_{n+1}$. In the final Part of these notes, we will show that there also exist polynomial preservers with the same property.

We now proceed toward the proof of Theorem 17.1 for general functions. A major step is the next calculation, which essentially proves the result for smooth functions. In the sequel, define the *Vandermonde determinant* of a vector $\mathbf{u} = (u_1, \dots, u_n)^T$ to be 1 if $n = 1$, and

$$V(\mathbf{u}) := \prod_{1 \leq j < k \leq n} (u_k - u_j) = \det \begin{pmatrix} 1 & u_1 & \cdots & u_1^{n-1} \\ 1 & u_2 & \cdots & u_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & u_n & \cdots & u_n^{n-1} \end{pmatrix}, \qquad \text{if } n > 1. \tag{17.4}$$

Proposition 17.5. *Fix an integer $n > 0$ and define $N := \binom{n}{2}$. Suppose $a \in \mathbb{R}$ and let a function $f : (a - \epsilon, a + \epsilon) \to \mathbb{R}$ be N-times differentiable for some fixed $\epsilon > 0$. Now fix vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, and define $\Delta : (-\epsilon', \epsilon') \to \mathbb{R}$ via $\Delta(t) := \det f[a\mathbf{1}_{n\times n} + t\mathbf{u}\mathbf{v}^T]$ for a sufficiently small $\epsilon' \in (0, \epsilon)$. Then $\Delta(0) = \Delta'(0) = \cdots = \Delta^{(N-1)}(0) = 0$, and*

$$\Delta^{(N)}(0) = \binom{N}{0, 1, \dots, n-1} V(\mathbf{u}) V(\mathbf{v}) \prod_{k=0}^{n-1} f^{(k)}(a),$$

where the first factor on the right is the multinomial coefficient.

This computation was originally due to Loewner. While the result seemingly involves (higher) derivatives, it is in fact a completely algebraic phenomenon, valid over any ground ring. For the interested reader, we isolate this phenomenon in Proposition 17.8; its proof is more or less the same as the one now provided for Proposition 17.5. To gain some feel for the computations, the reader may wish to work out the $N = 3$ case first.

Proof. Let $\mathbf{w}_k$ denote the kth column of $a\mathbf{1}_{n\times n} + t\mathbf{u}\mathbf{v}^T$; thus, $\mathbf{w}_k$ has jth entry $a + tu_jv_k$. To differentiate $\Delta(t)$, we will use the multilinearity of the determinant and the Laplace expansion of $\Delta(t)$ into a linear combination of $n!$ "monomials", each of which is a product of n terms $f(\cdot)$. Using the product rule, taking the derivative yields n terms from each monomial, and we may rearrange all of these terms into n "clusters" of terms (grouping by the column which gets differentiated), and regroup back using the Laplace expansion to obtain

$$\Delta'(t) = \sum_{k=1}^{n} \det(f[\mathbf{w}_1] \mid \cdots \mid f[\mathbf{w}_{k-1}] \mid v_k\mathbf{u} \circ f'[\mathbf{w}_k] \mid f[\mathbf{w}_{k+1}] \mid \cdots \mid f[\mathbf{w}_n]).$$

Now apply the derivative repeatedly, using this principle. Using the chain rule, for $M \geq 0$ the derivative $\Delta^{(M)}(t)$ – evaluated at $t = 0$ – is an integer linear combination of terms of the form

$$\begin{aligned} &\det(v_1^{m_1}\mathbf{u}^{\circ m_1} \circ f^{(m_1)}[a\mathbf{1}] \mid \cdots \mid v_n^{m_n}\mathbf{u}^{\circ m_n} \circ f^{(m_n)}[a\mathbf{1}]) \\ = \; &\det(f^{(m_1)}(a)v_1^{m_1}\mathbf{u}^{\circ m_1} \mid \cdots \mid f^{(m_n)}(a)v_n^{m_n}\mathbf{u}^{\circ m_n}), \qquad m_1 + \cdots + m_n = M, \end{aligned} \tag{17.6}$$

where $\mathbf{1} = (1, \ldots, 1)^T \in \mathbb{R}^n$ and all $m_j \geq 0$. Notice that if any $m_j = m_k$ for $j \neq k$, then the corresponding determinant (17.6) vanishes. Thus, the lowest degree derivative $\Delta^{(M)}(0)$ whose expansion contains a non-vanishing determinant is when $M = 0+1+\cdots+(n-1) = N$. This proves the first part of the result.

To show the second part, consider $\Delta^{(N)}(0)$. Once again, the only determinant terms that do not vanish in its expansion correspond to applying $0, 1, \ldots, n-1$ derivatives to the columns in some order. We first compute the integer multiplicity of each such determinant, noting by symmetry that these multiplicities are all equal. As we are applying N derivatives to Δ (before evaluating at 0), the derivative applied to get f' in some column can be any of $\binom{N}{1}$; now the two derivatives applied to get f'' in a (different) column can be chosen in $\binom{N-1}{2}$ ways; and so on. Thus, the multiplicity is precisely

$$\binom{N}{1}\binom{N-1}{2}\binom{N-3}{3}\cdots\binom{2n-3}{n-2} = \prod_{k=0}^{n-1}\binom{N-\binom{k}{2}}{k} = \frac{N!}{\prod_{k=0}^{n-1} k!} = \binom{N}{0, 1, \ldots, n-1}.$$

We next compute the sum of all determinant terms. Each term corresponds to a unique permutation of the columns $\sigma \in S_n$, with say $\sigma_k - 1$ the order of the derivative applied to the kth column $f[\mathbf{w}_k]$. Using (17.6), the determinant corresponding to σ equals

$$\begin{aligned} &\prod_{k=0}^{n-1} f^{(k)}(a)v_k^{\sigma_k - 1} \cdot (-1)^\sigma \cdot \det(\mathbf{u}^{\circ 0} \mid \mathbf{u}^{\circ 1} \mid \cdots \mid \mathbf{u}^{\circ(n-1)}) \\ = \; &V(\mathbf{u})\prod_{k=0}^{n-1} f^{(k)}(a) \cdot (-1)^\sigma \prod_{k=0}^{n-1} v_k^{\sigma_k - 1}. \end{aligned}$$

Summing this term over all $\sigma \in S_n$ yields precisely

$$V(\mathbf{u}) \prod_{k=0}^{n-1} f^{(k)}(a) \sum_{\sigma \in S_n} (-1)^\sigma \prod_{k=0}^{n-1} v_k^{\sigma_k - 1} = V(\mathbf{u}) \prod_{k=0}^{n-1} f^{(k)}(a) \cdot V(\mathbf{v}).$$

Now, multiply by the (common) integer multiplicity computed above, to finish the proof. $\square$

We next present the promised algebraic formulation of Proposition 17.5. For this, some notation is required. Fix a commutative (unital) ring R and an R-algebra S. The first step is to formalize the notion of the derivative, on a subclass of S-valued functions. This involves more structure than the more common notion of a *derivation*, so we give it a different name.

Definition 17.7. Given a commutative ring R, a commutative R-algebra S (with $R \subset S$), and an R-module X, a *differential calculus* is a pair (A, ∂), where A is an R-subalgebra of functions : $X \to S$ (under pointwise addition and multiplication and R-action) which contains the constant functions, and $\partial : A \to A$ satisfies the following properties:

(1) ∂ is R-linear, i.e., $\partial \sum_j r_j f_j = \sum_j r_j \partial f_j$ for all $r_j \in R,\ f_j \in A$ (and all j).
(2) ∂ is a derivation (product rule): $\partial(fg) = f \cdot (\partial g) + (\partial f) \cdot g$ for $f, g \in A$.
(3) ∂ satisfies a variant of the chain rule for composing with linear functions. Namely, if $x' \in X, r \in R$, and $f \in A$, then the function $g : X \to S,\ g(x) := f(x' + rx)$ also lies in A, and moreover, $(\partial g)(x) = r \cdot (\partial f)(x' + rx)$.

With this definition in hand, we can now state the desired algebraic generalization of Proposition 17.5; the proof is essentially the same.

Proposition 17.8. *Suppose R, S, and X are as in Definition 17.7, with an associated differential calculus (A, ∂). Now, fix an integer $n > 0$, two vectors $\mathbf{u}, \mathbf{v} \in R^n$, a vector $a \in X$, and a function $f \in A$. Define $N \in \mathbb{N}$ and $\Delta : X \to R$ via*

$$N := \binom{n}{2}, \qquad \Delta(t) := \det f[a\mathbf{1}_{n\times n} + t\mathbf{u}\mathbf{v}^T], \quad t \in X.$$

Then $\Delta(0_X) = (\partial\Delta)(0_X) = \cdots = (\partial^{N-1}\Delta)(0_X) = 0_R$, and

$$(\partial^N \Delta)(0_X) = \binom{N}{0, 1, \dots, n-1} V(\mathbf{u}) V(\mathbf{v}) \prod_{k=0}^{n-1} (\partial^k f)(a).$$

Notice that the algebra A is supposed to remind the reader of "smooth functions", and is used here for ease of exposition. One can instead work with an appropriate algebraic notion of "N-times differentiable functions" in order to truly generalize Proposition 17.5; we leave the details to the interested reader.

Remark 17.9. Note that Proposition 17.5 is slightly more general than the original argument of Horn and Loewner, which involved the special case $\mathbf{u} = \mathbf{v}$. As the above proof (and Proposition 17.8) shows, the argument is essentially algebraic, hence holds for any $\mathbf{u}, \mathbf{v}$.

Finally, we use Proposition 17.5 to prove the Horn–Loewner theorem 17.1 for smooth functions. The remainder of the proof – for arbitrary functions – will be discussed in the next section (Section 18).

Proof of Theorem 17.1 for smooth functions. Suppose f is smooth – or more generally, C^N where $N = \binom{n}{2}$. Then the result is shown by induction on n. For $n = 1$ the result says that f is non-negative if it preserves positivity on the given test set, which is obvious. For the induction step, we know that $f, f', \dots, f^{(n-2)} \geq 0$ on I, since the given test set of $(n-1) \times (n-1)$ matrices can be embedded into the test set of $n \times n$ matrices. (Here we do not use the test matrices in $\mathbb{P}_2(I)$.) Now define $f_\epsilon(x) := f(x) + \epsilon x^n$ for each $\epsilon > 0$, and note by the Schur product theorem 3.12 (or Lemma 16.1) that f_ϵ also satisfies the hypotheses.

Given $a, t > 0$ and the vector $\mathbf{u} = (1, u_0, \dots, u_0^{n-1})^T$ as in the theorem, define $\Delta(t) := \det f_\epsilon[a\mathbf{1}_{n\times n} + t\mathbf{u}\mathbf{u}^T]$ as in Proposition 17.5 (but replacing $f, \mathbf{v}$ by $f_\epsilon, \mathbf{u}$ respectively). Then $\Delta(t) \geq 0$ for $t > 0$ by assumption, so

$$0 \leq \lim_{t \to 0^+} \frac{\Delta(t)}{t^N}, \qquad \text{where } N = \binom{n}{2}.$$

On the other hand, by Proposition 17.5 and applying L'Hôpital's rule,

$$\lim_{t\to 0^+} \frac{\Delta(t)}{t^N} = \frac{\Delta^{(N)}(0)}{N!} = \frac{1}{N!}\binom{N}{0,1,\dots,n-1} V(\mathbf{u})^2 \prod_{k=0}^{n-1} f_\epsilon^{(k)}(a) = V(\mathbf{u})^2 \prod_{k=0}^{n-1} \frac{f_\epsilon^{(k)}(a)}{k!}.$$

Thus, the right-hand side here is non-negative. Since $\mathbf{u}$ has distinct coordinates, we can cancel all positive factors to conclude that

$$\prod_{k=0}^{n-1} f_\epsilon^{(k)}(a) \geq 0, \quad \forall \epsilon, a > 0.$$

But $f_\epsilon^{(k)}(a) = f^{(k)}(a) + \epsilon n(n-1)\cdots(n-k+1)a^{n-k}$, and this is positive for $k = 0, \dots, n-2$ by the induction hypothesis. Hence,

$$f_\epsilon^{(n-1)}(a) = f^{(n-1)}(a) + \epsilon\, a\, n! \geq 0, \qquad \forall \epsilon, a > 0.$$

It follows that $f^{(n-1)}(a) \geq 0$, hence $f^{(n-1)}$ is non-negative on $(0, \infty)$, as desired. $\square$

We conclude this line of proof by mentioning that the Horn–Loewner theorem, as well as Proposition 17.5 and its algebraic avatar in Proposition 17.8 afford generalizations; the latter results reveal a surprising and novel application to Schur polynomials and to symmetric function identities. For more details, the reader is referred to the recent paper [214] in *Trans. Amer. Math. Soc.* by the author.

The final remark is that there is a *different*, simpler proof of Theorem 17.1 for smooth functions, essentially by Vasudeva (1979, [353]) and along the lines of FitzGerald–Horn's 1977 argument ([123] and see also the proof of Theorem 9.3). Vasudeva's proof is direct, so does not lead to the connections to Schur polynomials mentioned in the preceding paragraph.

Simpler proof of Theorem 17.1 for smooth functions. This proof in fact works for $f \in C^{n-1}(I)$. Akin to the previous proof, this argument also works more generally than for $\mathbf{u} = (1, u_0, \dots, u_0^{n-1})^T$: choose arbitrary distinct real scalars $v_1, \dots, v_n$ and write $v := (v_1, \dots, v_n)^T$. Then for $a > 0$ and small $t > 0$, $a\mathbf{1}_{n\times n} + tvv^T \in \mathbb{P}_n(I)$. Now, given $0 \leq m \leq n-1$, choose a vector $u \in R^n$ which is orthogonal to the vectors $\mathbf{1}, v, v^{\circ 2}, v^{\circ(m-1)}$ but not to $v^{\circ m}$, and compute using the hypotheses and the Taylor expansion of f at a:

$$0 \leq u^T f[a\mathbf{1}_{n\times n} + tvv^T]u = u^T \left(\sum_{l=0}^{m-1} f^{(l)}(a) \frac{t^l}{l!} v^{\circ l}(v^{\circ l})^T + \frac{t^m}{m!} C \right) u = \frac{t^m}{m!} u^T C u,$$

where $C_{n\times n}$ has (j,k) entry $(v_j v_k)^m f^{(m)}(a + \theta_{jk} t v_j v_k)$ with all $\theta_{j,k} \in (0,1)$. Divide by $t^m/m!$ and let $t \to 0^+$; since f is C^m, we obtain $0 \leq (u^T v^{\circ m})^2 f^{(m)}(a)$. As this holds for all $0 \leq m \leq n-1$, the proof is complete. $\square$

18. The stronger Horn–Loewner theorem. Mollifiers.

We continue with the proof of the Horn–Loewner theorem 17.1. This is in three steps:

(1) Theorem 17.1 holds for smooth functions. This was proved in the previous section.
(2) If Theorem 17.1 holds for smooth functions, then it holds for continuous functions. Here, we need to assume $n \geq 3$.
(3) If f satisfies the hypotheses in Theorem 17.1, then it is continuous. This follows from Vasudeva's 2×2 result – see (the proof of) Theorem 12.7 and Remark 12.10.

To carry out the second step – as well as a similar step in proving Schoenberg's theorem, see Section 22.2 – we will use a standard tool in analysis called *mollifiers*.

18.1. An introduction to (one-variable) mollifiers.

In this subsection, we examine some basic properties of mollifiers of one variable: the theory extends to $\mathbb{R}^n$ for all $n > 1$, but that is not required in what follows.

First, recall that one can construct smooth functions $g : \mathbb{R} \to \mathbb{R}$, such that g and all its derivatives vanish on $(-\infty, 0]$: for instance, $g(x) = \exp(-1/x) \cdot \mathbf{1}(x > 0)$. Indeed, one shows that $g^{(n)}(x) = p_n(1/x) g(x)$ for some polynomial p_n; hence, $g^{(n)}(x) \to 0$ as $x \to 0$. Hence:

Lemma 18.1. *Given scalars $-1 < a < b < 0$, there exists a smooth function ϕ that vanishes outside $[a, b]$, is positive on (a, b), and is a probability distribution on $\mathbb{R}$.*

Of course, the assumption $[a, b] \subset (-1, 0)$ is completely unused in the proof of the lemma, but is included for easy reference since we will require it in what follows.

Proof. The function $\varphi(x) := g(x-a)g(b-x)$ is non-negative, smooth, and supported precisely on (a, b). In particular, $\int_{\mathbb{R}} \varphi > 0$, so the normalization $\phi := \varphi / \int_{\mathbb{R}} \varphi$ has the desired properties. $\square$

We now introduce mollifiers.

Definition 18.2. A *mollifier* is a one-parameter family of functions (in fact probability distributions)

$$\{\phi_\delta(x) := \tfrac{1}{\delta}\phi(\tfrac{x}{\delta}) : \delta > 0\},$$

with real domain and range, corresponding to any function ϕ satisfying Lemma 18.1.

A continuous, real-valued function f (with suitable domain inside $\mathbb{R}$) is said to be *mollified* by convolving with the family ϕ_δ. In this case, we define

$$f_\delta(x) := \int_{\mathbb{R}} f(t)\phi_\delta(x-t)\, dt,$$

where one extends f outside its domain by zero. (This is called *convolution*: $f_\delta = f * \phi_\delta$.)

Remark 18.3. Mollifiers, or *Friedrichs mollifiers*, were used by Horn and Loewner in the late 1960s, as well as previously by Rudin in his 1959 proof of Schoenberg's theorem. They were a relatively modern tool at the time, having been introduced by Friedrichs in his seminal 1944 paper [132] on PDEs in *Trans. Amer. Math. Soc.*, as well as slightly earlier by Sobolev in his famous 1938 paper [343] in *Mat. Sbornik* (which contained the proof of the Sobolev embedding theorem).

Returning to the definition of a mollifier, notice by the change of variables $u = x - t$ and Lemma 18.1 that

$$f_\delta(x) = \frac{1}{\delta}\int_{\mathbb{R}} f(x-u)\phi\left(\frac{u}{\delta}\right) du = \int_{-\delta}^{0} f(x-u)\phi_\delta(u)\, du. \tag{18.4}$$

In particular, f_δ is a "weighted average" of the image set $f([x, x+\delta])$, since ϕ is a probability distribution. Now it is not hard to see that f_δ is continuous and converges to f pointwise as $\delta \to 0^+$. In fact, more is true:

Proposition 18.5. *If $I \subset \mathbb{R}$ is a right open interval and $f : I \to R$ is continuous, then for all $\delta > 0$, the mollified functions f_δ are smooth on $\mathbb{R}$ (where we extend f outside I by zero), and converge uniformly to f on compact subsets of I as $\delta \to 0^+$.*

To prove this result, we show two lemmas in somewhat greater generality. First, some notation: a (Lebesgue measurable) function $f : \mathbb{R} \to \mathbb{R}$ is said to be *locally* L^1 if it is L^1 on each compact subset of $\mathbb{R}$.

Lemma 18.6. *If $f : \mathbb{R} \to \mathbb{R}$ is locally L^1, and $\psi : \mathbb{R} \to \mathbb{R}$ is continuous with compact support, then $f * \psi : \mathbb{R} \to \mathbb{R}$ is also continuous.*

Proof. Suppose $x_n \to x$ in $\mathbb{R}$; without loss of generality $|x_n - x| < 1$ for all $n > 0$. Also choose $r, M > 0$, such that ψ is supported on $[-r, r]$ and $M = \|\psi\|_{L^\infty(\mathbb{R})} = \max_{\mathbb{R}} |\psi(x)|$. Then for each $t \in \mathbb{R}$, we have

$$f(t)\psi(x_n - t) \to f(t)\psi(x - t), \qquad |f(t)\psi(x_n - t)| \leq M|f(t)| \cdot \mathbf{1}(|x - t| \leq r + 1)$$

(the second inequality follows by considering separately the cases $|x-t| \leq r+1$ and $|x-t| > r+1$). Since the right-hand side is integrable, Lebesgue's dominated convergence theorem applies:

$$\begin{aligned}\lim_{n\to\infty} (f * \psi)(x_n) &= \lim_{n\to\infty} \int_{\mathbb{R}} f(t)\psi(x_n - t)\ dt = \int_{\mathbb{R}} \lim_{n\to\infty} f(t)\psi(x_n - t)\ dt \\ &= \int_{\mathbb{R}} f(t)\psi(x - t)\ dt = (f * \psi)(x),\end{aligned}$$

so $f * \psi$ is continuous on $\mathbb{R}$. $\square$

Lemma 18.7. *If $f : \mathbb{R} \to \mathbb{R}$ is locally L^1, and $\psi : \mathbb{R} \to \mathbb{R}$ is C^1 with compact support, then $f * \psi : \mathbb{R} \to \mathbb{R}$ is also C^1 and $(f * \psi)' = f * \psi'$ on $\mathbb{R}$.*

Proof. We compute

$$\begin{aligned}(f * \psi)'(x) &= \lim_{h\to 0} \frac{1}{h} \int_{\mathbb{R}} f(y)\psi(x + h - y)\ dy - \frac{1}{h} \int_{\mathbb{R}} f(y)\psi(x - y)\ dy \\ &= \lim_{h\to 0} \int_{\mathbb{R}} f(y) \frac{\psi(x + h - y) - \psi(x - y)}{h}\ dy \\ &= \lim_{h\to 0} \int_{\mathbb{R}} f(y)\psi'(x - y + c(h, y))\ dy,\end{aligned}$$

where for each $y \in \mathbb{R}$, $c(h, y) \in [0, h]$ is chosen using the Mean Value Theorem. While $c(h, y) \to 0$ as $h \to 0$, the problem is that y is not fixed inside the integral. Thus, to proceed, we argue as follows: suppose ψ is supported inside $[-r, r]$ as above, hence so is ψ'. Choose any sequence $h_n \to 0$. Now **claim** that the last integral above, evaluated at h_n, converges to $(f * \psi')(x)$ as $n \to \infty$ – hence so does the limit of the last integral above. Indeed, we may first assume all $h_n \in (-1, 1)$; and then check for each n that the above integral equals

$$\int_{\mathbb{R}} f(y)\psi'(x - y + c(h_n, y))\ dy = \int_{x-(r+1)}^{x+(r+1)} f(y)\psi'(x - y + c(h_n, y))\ dy$$

by choice of r. Now the integrand on the right-hand side is bounded above by $M_1|f(y)|$ in absolute value, where $M_1 := \|\psi'\|_{L^\infty(\mathbb{R})}$. Hence, by the dominated convergence theorem,

$$\lim_{n\to\infty} \int_{x-(r+1)}^{x+(r+1)} f(y)\psi'(x-y+c(h_n,y))\ dy = \int_{x-(r+1)}^{x+(r+1)} f(y)\psi'(x-y)\ dy = (f*\psi')(x),$$

where the first equality also uses that ψ is C^1 and the second is by choice of r. Since this happens for every sequence $h_n \to 0$, it follows that $(f*\psi)'(x) = (f*\psi')(x)$. Moreover, $f*\psi'$ is continuous by Lemma 18.6, since ψ is C^1. This shows $f*\psi$ is C^1 as claimed. $\square$

Finally, we show the claimed properties of mollified functions.

Proof of Proposition 18.5. Extending f by zero outside I, it follows that f is locally L^1 on $\mathbb{R}$. Repeatedly applying Lemma 18.7 to $\psi = \phi_\delta, \phi'_\delta, \phi''_\delta, \dots$, we conclude that $f_\delta \in C^\infty(\mathbb{R})$.

To prove local uniform convergence, let K be a compact subset of I and $\epsilon > 0$. Denote $b := \sup K$ and $a := \inf K$. Since I is right open, there is a number $l > 0$, such that $J := [a, b+l] \subset I$. Since f is uniformly continuous on J, given $\epsilon > 0$ there exists $\delta \in (0,l)$, such that $|x-y| < \delta,\ x,y \in J \implies |f(x)-f(y)| < \epsilon$.

Now claim that if $0 < \xi < \delta$ then $\|f_\xi - f\|_{L^\infty(K)} \leq \epsilon$; note this proves the uniform convergence of the family f_δ to f on K. To show the claim, compute using (18.4) for $x \in K$:

$$\begin{aligned} |f_\xi(x) - f(x)| &= \left| \int_{-\xi}^0 (f(x-u) - f(x))\phi_\xi(u)\ du \right| \\ &\leq \int_{-\xi}^0 |f(x-u) - f(x)|\phi_\xi(u) du \quad \leq \quad \epsilon \int_{-\xi}^0 \phi_\xi(u)\ du = \epsilon. \end{aligned}$$

This is true for all $x \in K$ by the choice of $\xi < \delta < l$, and hence, proves the claim. $\square$

18.2. **Completing the proof of the Horn–Loewner theorem.** With mollifiers in hand, we finish the proof of Theorem 17.1. As mentioned at the start of this section, the proof can be divided into three steps, and two are now already worked out. It remains to show the second step, that is, if $n \geq 3$ and if the result holds for smooth functions, then it holds for continuous functions.

Thus, suppose $I = (0,\infty)$ and $f : I \to \mathbb{R}$ is continuous. Define the mollified functions $f_\delta,\ \delta > 0$ as above; note each f_δ is smooth. Moreover, given $a, b > 0$, by (18.4) the function f_δ satisfies

$$f_\delta[a\mathbf{1}_{n\times n} + b\mathbf{u}\mathbf{u}^T] = \int_{-\delta}^0 \phi_\delta(y) \cdot f[(a+|y|)\mathbf{1}_{n\times n} + b\mathbf{u}\mathbf{u}^T]\ dy, \tag{18.8}$$

and this is positive semidefinite by the assumptions for f. Thus, $f_\delta[-]$ preserves positivity on the given test set in $\mathbb{P}_n(I)$; a similar argument shows that $f_\delta[-]$ preserves positivity on $\mathbb{P}_2(I)$. Hence, by the proof in the previous section, $f_\delta, f'_\delta, \dots, f_\delta^{(n-1)}$ are non-negative on I.

Observe that the theorem amounts to deducing a similar statement for f; however, as f is a priori known only to be continuous, we can only deduce non-negativity for a discrete version of the derivatives – namely, divided differences:

Definition 18.9. Suppose I is a real interval and a function $f : I \to \mathbb{R}$. Given $h > 0$ and an integer $k \geq 0$, the *kth order forward differences with step size* $h > 0$ are defined as follows:

$$(\Delta_h^0 f)(x) := f(x), \qquad (\Delta_h^k f)(x) := (\Delta_h^{k-1} f)(x+h) - (\Delta_h^{k-1} f)(x) = \sum_{j=0}^k \binom{k}{j} (-1)^{k-j} f(x+jh),$$

whenever $k > 0$ and $x, x + kh \in I$. Similarly, the *kth order divided differences with step size* $h > 0$ are

$$(D_h^k f)(x) := \frac{1}{h^k}(\Delta_h^k f)(x), \qquad \forall k \geq 0,\ x, x + kh \in I.$$

The key point is that if a function is differentiable to some order, and its derivatives of that order are non-negative on an open interval, then using the mean value theorem for divided differences, one shows the corresponding divided differences are also non-negative, hence so are the corresponding forward differences. Remarkably, the converse also holds, *including differentiability*! This is a classical result by Boas and Widder:

Theorem 18.10. *Suppose $I \subset \mathbb{R}$ is an open interval, bounded or not, and $f : I \to \mathbb{R}$.*

(1) *(Cauchy's mean value theorem for divided differences: special case.) If f is k-times differentiable in I for some integer $k > 0$, and $x, x + kh \in I$ for $h > 0$, then there exists $y \in (x, x + kh)$, such that $(D_h^k f)(x) = f^{(k)}(y)/k!$.*

(2) *(Boas–Widder, Duke Math. J., 1940.) Suppose $k \geq 2$ is an integer, and $f : I \to \mathbb{R}$ is continuous and has all forward differences of order k non-negative on I:*

$$(\Delta_h^k f)(x) \geq 0, \quad \text{whenever } h > 0 \text{ and } x, x + kh \in I.$$

Then on all of I, the function $f^{(k-2)}$ exists, is continuous and convex, and has non-decreasing left- and right-hand derivatives.

We make a few remarks on Boas and Widder's result. First, for $k = 2$ the result seems similar to Ostrowski's theorem 12.2, except for the local boundedness being strengthened to continuity. Second, note that while $f_{\pm}^{(k-1)}$ is non-decreasing by the theorem, one cannot claim here that the lower-order derivatives $f, \dots, f^{(k-2)}$ are non-decreasing on I. Indeed, a counterexample for such an assertion for $f^{(l)}$, where $0 \leq l \leq k-2$, is $f(x) = -x^{l+1}$ on $I \subset \mathbb{R}$. Finally, we refer the reader to Section 23.1 for additional related observations and results.

Proof. The second part will be proved in detail in Section 23. For the first, consider the Newton form of the Lagrange interpolation polynomial $P(X)$ for $f(X)$ at $X = x, x+h, \dots, x+kh$. The highest term of $P(X)$ is

$$(D_h^k f)(x) \cdot (X - x_{k-1}) \cdots (X - x_1)(X - x_0), \qquad \text{where } x_j = x_0 + jh\ \forall j \geq 0.$$

Writing $g(X) := f(X) - P(X)$ to be the remainder function, note that g vanishes at $x, x + h, \dots, x + kh$. By successively applying Rolle's theorem to $g, g', \dots, g^{(k-1)}$, it follows that $g^{(k)}$ has a root in $(x, x + kh)$, say y. But then,

$$0 = g^{(k)}(y) = f^{(k)}(y) - (D_h^k f)(x)k!,$$

which concludes the proof. □

Returning to our proof of the stronger Horn–Loewner theorem 17.1, since $f_\delta, f_\delta', \dots, f_\delta^{(n-1)} \geq 0$ on I, by the above theorem the divided differences of f_δ up to order $n-1$ are non-negative on I, hence the same holds for the forward differences of f_δ. Applying Proposition 18.5, the forward differences of f of orders $k = 0, \dots, n-1$ are also non-negative on I. Finally, invoke the Boas–Widder theorem for $k = 2, \dots, n-1$ to conclude the proof of the (stronger) Horn–Loewner theorem – noting for "low orders" that f is non-negative and non-decreasing on I by using forward differences of orders $k = 0, 1$ respectively, and hence $f, f' \geq 0$ on I as well. □

19. The stronger Vasudeva and Schoenberg theorems. Bernstein's theorem. Moment sequence transforms.

19.1. **The theorems of Vasudeva and Bernstein.** Having shown (the stronger form of) the Horn–Loewner theorem 17.1, we use it to prove the following strengthening of Vasudeva's theorem 16.4. In it, recall from Definition 12.18 that $\mathtt{HTN}_n$ denotes the set of $n \times n$ Hankel totally non-negative matrices. (These are automatically positive semidefinite.)

Theorem 19.1 (Vasudeva's theorem, stronger version – also see Remark 19.18)**.** *Suppose $I = (0, \infty)$ and $f : I \to \mathbb{R}$. The following are equivalent:*

1. *The entrywise map $f[-]$ preserves positivity on $\mathbb{P}_n(I)$ for all $n \geq 1$.*
2. *The entrywise map $f[-]$ preserves positivity on all matrices in $\mathtt{HTN}_n$ with positive entries and rank at most 2 for all $n \geq 1$.*
3. *The function f equals a convergent power series $\sum_{k=0}^\infty c_k x^k$ for all $x \in I$, with the Maclaurin coefficients $c_k \geq 0$ for all $k \geq 0$.*

To show the theorem, we require the following well-known classical result by Bernstein:

Definition 19.2. If $I \subset \mathbb{R}$ is open, we say that $f : I \to \mathbb{R}$ is *absolutely monotonic* if f is smooth on I and $f^{(k)} \geq 0$ on I for all $k \geq 0$.

Theorem 19.3 (Bernstein, [44])**.** *Suppose $-\infty < a < b \leq \infty$. If $f : [a, b) \to \mathbb{R}$ is continuous at a and absolutely monotonic on (a, b), then f can be extended analytically to the complex disk $D(a, b - a)$.*

With Bernstein's theorem in hand, the "stronger Vasudeva theorem" follows easily:

Proof of Theorem 19.1. By the Schur product theorem or Lemma 16.1, (3) $\implies$ (1); and clearly (1) $\implies$ (2). Now suppose (2) holds. By the stronger Horn–Loewner theorem 17.1, $f^{(k)} \geq 0$ on I for all $k \geq 0$, i.e., f is absolutely monotonic on I. In particular, f is non-negative and non-decreasing on $I = (0, \infty)$, so it can be continuously extended to the origin via $f(0) := \lim_{x \to 0^+} f(x) \geq 0$. Now apply Bernstein's theorem with $a = 0$ and $b = \infty$ to deduce that f agrees on $[0, \infty)$ with an entire function $\sum_{k=0}^\infty c_k x^k$. Moreover, since $f^{(k)} \geq 0$ on I for all k, it follows that $f^{(k)}(0) \geq 0$, i.e., $c_k \geq 0 \ \forall k \geq 0$. Restricting to I, we obtain (3), as desired. ☐

On a related note, recall Theorems 9.3 and 12.19, which showed that when studying entrywise powers preserving the two closed convex cones $\mathbb{P}_n([0, \infty))$ and $\mathtt{HTN}_n$, the answers were identical. This is perhaps not surprising, given Theorem 4.1. In this vein, we observe that such an equality of preserver sets also holds when classifying the entrywise maps preserving Hankel TN matrices with positive entries:

Corollary 19.4. *With $I = (0, \infty)$ and $f : I \to \mathbb{R}$, the three assertions in Theorem 19.1 are further equivalent to:*

4. *The entrywise map $f[-]$ preserves total non-negativity on all matrices in $\mathtt{HTN}_n$ with positive entries, for all $n \geq 1$.*
5. *The entrywise map $f[-]$ preserves total non-negativity on the matrices in $\mathtt{HTN}_n$ with positive entries and rank at most 2, for all $n \geq 1$.*

Proof. Clearly (4) $\implies$ (5) $\implies$ (2), where (1)–(3) are as in Theorem 19.1. That (3) $\implies$ (4) follows from Lemma 16.1 and Theorem 4.1. ☐

Remark 19.5. Here is one situation where the two sets of preservers – of positivity on $\mathbb{P}_n$ for all n, and of total non-negativity on $\mathtt{HTN}_n$ for all n – differ: if we also allow zero entries, as opposed to only positive entries as in the preceding corollary and Theorem 19.1. In this case, one shows that the preservers of $\mathtt{HTN}_n$ for all n are the functions f, such that $f|_{(0,\infty)}$ is absolutely monotonic, and hence a power series with non-negative Maclaurin coefficients; and, such that $0 \leq f(0) \leq \lim_{x \to 0^+} f(x)$, since the only Hankel TN matrices with a zero entry arise as truncated moment matrices of measures $a\delta_0$. On the other hand, by considering the rank-2 Hankel positive semidefinite matrix $A := \begin{pmatrix} 2 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$, and considering the inequality

$$\lim_{x \to 0^+} \det f[xA] \geq 0,$$

it follows that f is continuous at 0^+. (Note, A is not TN.) In particular, the entrywise preservers of positivity on $\bigcup_{n \geq 1} \mathbb{P}_n([0,\infty))$ are precisely the functions $\sum_{k \geq 0} c_k x^k$, with all $c_k \geq 0$.

Remark 19.6. As a reminder, we recall that if one instead tries to classify the entrywise preservers of total non-negativity on all (possibly symmetric) TN matrices, then one obtains only the constant or linear functions $f(x) = c, cx$ for $c, x \geq 0$. See Theorem 12.11 above.

To complete the proof of the stronger Vasudeva theorem 19.1 as well as its corollary (Corollary 19.4), it remains to show Bernstein's theorem.

Proof of Bernstein's theorem 19.3. First we claim that $f^{(k)}(a^+)$ exists and equals $\lim_{x \to a^+} f^{(k)}(x)$ for all $k \geq 0$. The latter limit here exists because $f^{(k+1)} \geq 0$ on (a,b), so $f^{(k)}(x)$ is non-negative and non-decreasing on $[a,b)$.

It suffices to show the claim for $k = 1$. But here we compute:

$$f'(a^+) = \lim_{h \to 0^+} \frac{f(a+h) - f(a)}{h} = \lim_{h \to 0^+} f'(a + c(h)),$$

where $c(h) \in [0,h]$ exists and goes to zero as $h \to 0^+$, by the Mean Value Theorem. The claim now follows from the previous paragraph. In particular, $f^{(k)}$ exists and is continuous, non-negative, and non-decreasing on $[a,b)$.

Applying Taylor's theorem, we have

$$f(x) = f(a^+) + f'(a^+)(x-a) + \cdots + f^{(n)}(a^+)\frac{(x-a)^n}{n!} + R_n(x),$$

where R_n is the Taylor remainder

$$R_n(x) = \int_a^x \frac{(x-t)^n}{n!} f^{(n+1)}(t)\ dt. \tag{19.7}$$

By the assumption on f, we see that $R_n(x) \geq 0$. Changing variables to $t = a + y(x-a)$, the limits for y change to $0, 1$, and we have

$$R_n(x) = \frac{(x-a)^{n+1}}{n!} \int_0^1 (1-y)^n f^{(n+1)}(a + y(x-a))\ dy.$$

Since $f^{(n+2)} \geq 0$ on $[a,b)$, if $a \leq x \leq c$ for some $c < b$, then uniformly in $[a,c]$ we have $0 \leq f^{(n+1)}(a + y(x-a)) \leq f^{(n+1)}(a + y(c-a))$. Therefore, using Taylor's remainder formula once again, we obtain

$$\begin{aligned}
0 \leq R_n(x) &\leq \frac{(x-a)^{n+1}}{n!} \int_0^1 (1-y)^n f^{(n+1)}(a+y(c-a))\ dy \\
&= R_n(c) \frac{(x-a)^{n+1}}{(c-a)^{n+1}} \\
&= \frac{(x-a)^{n+1}}{(c-a)^{n+1}} \left(f(c) - \sum_{k=0}^n f^{(k)}(a^+) \frac{(c-a)^k}{k!} \right) \\
&\leq f(c) \frac{(x-a)^{n+1}}{(c-a)^{n+1}}.
\end{aligned}$$

From this it follows that $\lim_{n\to\infty} R_n(x) = 0$ for all $x \in [a,c)$. Since this holds for all $c \in (a,b)$, the Taylor series of f converges to f on $[a,b)$. In other words,

$$f(x) = \sum_{k=0}^\infty \frac{f^{(k)}(a^+)}{k!}(x-a)^k, \qquad x \in [a,b).$$

Now if $z \in D(a, b-a)$, then clearly $a+|z-a| < a+(b-a) = b$. Choosing any $c \in (a+|z-a|, b)$, we check that the Taylor series converges (absolutely) at z:

$$\begin{aligned}
\left| \sum_{k=0}^\infty \frac{f^{(k)}(a^+)}{k!}(z-a)^k \right| &\leq \sum_{k=0}^\infty \frac{f^{(k)}(a^+)}{k!}|z-a|^k \\
&\leq \sum_{k=0}^\infty \frac{f^{(k)}(a^+)}{k!}|c-a|^k \\
&= f(c) < \infty.
\end{aligned}$$

This completes the proof of Bernstein's theorem – and with it, the stronger form of Vasudeva's theorem. □

Remark 19.8. We mention for completeness that Bernstein's theorem admits an extension, which was already shown by Bernstein in 1926 [43], and which says that even if $f^{(k)} \geq 0$ in (a,b) only for $k \geq 0$ *even*, then f is necessarily analytic in (a,b). (In fact, Bernstein worked only with divided differences – see Theorem 39.10.) This was further extended by Boas in *Duke Math. J.*, as follows:

Theorem 19.9 (Boas, 1941, [54]). *Let $\{n_p : p \geq 1\}$ be an increasing sequence of positive integers, such that n_{p+1}/n_p is uniformly bounded. Let $(a,b) \subset \mathbb{R}$ and $f : (a,b) \to \mathbb{R}$ be smooth. If for each $p \geq 1$, the derivative $f^{(n_p)}$ does not change sign in (a,b), then f is analytic in (a,b).*

19.2. **The stronger version of Schoenberg's theorem.** We now come to the main result of this part of the text: the promised strengthening of Schoenberg's theorem.

Theorem 19.10 (Schoenberg's theorem, stronger version). *Given $f : \mathbb{R} \to \mathbb{R}$, the following are equivalent:*

(1) The entrywise map $f[-]$ preserves positivity on $\mathbb{P}_n(\mathbb{R})$ for all $n \geq 1$.
(2) The entrywise map $f[-]$ preserves positivity on the Hankel matrices in $\mathbb{P}_n(\mathbb{R})$ of rank at most 3 for all $n \geq 1$.
(3) The function f equals a convergent power series $\sum_{k=0}^\infty c_k x^k$ for all $x \in \mathbb{R}$, with the Maclaurin coefficients $c_k \geq 0$ for all $k \geq 0$.

See also Theorem 19.15 for two a priori weaker, yet equivalent, assertions.

Remark 19.11. Recall the two definitions of positive definite functions, from Definition 16.15 and the discussion preceding Lemma 16.16. The metric-space version for Euclidean and Hilbert spheres was connected by Schoenberg to functions of the form $f \circ \cos$, by requiring that $(f(\langle x_j, x_k \rangle))_{j,k \geq 0}$ be positive semidefinite for all choices of vectors $x_j \in S^{r-1}$ (for $2 \leq r \leq \infty$). A *third* notion of positive definite kernels on Hilbert spaces $\mathcal{H}$ arises from here, and is important in machine learning among other areas: (see, e.g., [281, 345, 352]): one says $f : \mathbb{R} \to \mathbb{R}$ is *positive definite on* $\mathcal{H}$ if, for any choice of finitely many points $x_j, j \geq 0$, the matrix $(f(\langle x_j, x_k \rangle))_{j,k \geq 0}$ is positive semidefinite. Since Gram matrix ranks are bounded above by $\dim \mathcal{H}$, this shows that Rudin's 1959 theorem 16.3 classifies the positive definite kernels/functions on $\mathcal{H}$ for any real Hilbert space of dimension ≥ 3. The stronger Schoenberg theorem 19.10 provides a second proof.

Returning to the stronger Schoenberg theorem 19.10, clearly (1) $\implies$ (2), and (3) $\implies$ (1) by the Pólya–Szegő observation 16.1. Thus, the goal over the next few sections is to prove (2) $\implies$ (3). The proof is simplified when some of the arguments below are formulated in the language of *moment sequences* and their preservers. We begin by defining these and explaining the dictionary between moment sequences and positive-semidefinite Hankel matrices, due to Hamburger (among others).

Definition 19.12. Recall that given an integer $k \geq 0$ and a real measure μ supported on a subset of $\mathbb{R}$, μ has kth moment equal to the following (if it converges):

$$s_k(\mu) := \int_{\mathbb{R}} x^k \, d\mu.$$

Henceforth, we only work with *admissible* measures, i.e., such that μ is non-negative on $\mathbb{R}$ and $s_k(\mu)$ converges for all $k \geq 0$. The *moment sequence* of such a measure μ is the sequence

$$\mathbf{s}(\mu) := (s_0(\mu), s_1(\mu), \dots).$$

We next define *transforms* of moment sequences: a function $f : \mathbb{R} \to \mathbb{R}$ acts entrywise, to take moment sequences to real sequences:

$$f[\mathbf{s}(\mu)] := (f(s_k(\mu)))_{k \geq 0}. \tag{19.13}$$

We are interested in examining when the transformed sequence (19.13) is also the moment sequence of an admissible measure supported on $\mathbb{R}$. This connects to the question of positivity preservers via the following classical result.

Theorem 19.14 (Hamburger, [162])**.** *A real sequence $(s_k)_{k \geq 0}$ is the moment sequence of an admissible measure, if and only if the semi-infinite Hankel matrix $H := (s_{j+k})_{j,k \geq 0}$ is positive semidefinite.*

Recall that the easy half of this result was proved early on, in Lemma 2.22.

Thus, entrywise functions preserving positivity on Hankel matrices are intimately related to moment sequence preservers. Also note that if a measure μ has finite support in the real line, then by examining for example (2.21), the Hankel moment matrix H_μ (i.e., every submatrix) has rank at most the size of the support set. From this and Hamburger's theorem, we deduce all but the last sentence of the following result:

Theorem 19.15. *Theorem 19.10(2) implies the following a priori weaker statement:*

(4) For each measure

$$\mu = a\delta_1 + b\delta_{u_0} + c\delta_{-1}, \quad \textit{with } u_0 \in (0,1),\ a,b,c \geq 0, \tag{19.16}$$

there exists an admissible (non-negative) measure $\sigma = \sigma_\mu$ on $\mathbb{R}$, such that $f(s_k(\mu)) = s_k(\sigma)$ $\forall k \geq 0$.

In turn, this implies the still-weaker statement:

(5) For each measure μ as in (19.16), with semi-infinite Hankel moment matrix H_μ, the matrix $f[H_\mu]$ is positive semidefinite.

In fact, these statements are equivalent to the assertions in Theorem 19.10.

Remark 19.17. In this text, we do not prove Hamburger's theorem; but we have used it to state Theorem 19.15(4) – i.e., in working with the admissible measure $\sigma = \sigma_\mu$. A closer look reveals that the use of Hamburger's theorem and moment sequences is not required to prove Schoenberg's theorem, or even its stronger form in Theorem 19.15, which is explained in (proving) Theorem 19.15(5). Our workaround is explained in the next section (Section 20), via a "positivity-certificate trick" involving limiting sum-of-squares representations of polynomials. That said, moment sequences help simplify the presentation of the proof, and hence we will continue to use them in the proof, in later sections (Sections 21 and 22).

The next three sections are devoted to proving (5) $\implies$ (3) (in Theorems 19.10 and 19.15). Here is an outline of the steps in the proof:

(1) All matrices $A = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \in \mathbb{P}_2((0,\infty))$ with $a \geq c$ occur as leading principal submatrices of the Hankel moment matrices H_μ, where μ is as in (19.16).
(2) Apply the stronger Horn–Loewner theorem 17.1 and Bernstein's theorem to deduce that $f|_{(0,\infty)} = \sum_{k=0}^\infty c_k x^k$ for some $c_k \geq 0$.
(3) If f satisfies assertion (5) in Theorem 19.15, then f is continuous on $\mathbb{R}$.
(4) If, moreover, f is smooth and satisfies assertion (5) in Theorem 19.15, then f is real analytic.
(5) Real analytic functions satisfy the desired implication above: (5) $\implies$ (3).
(6) Using mollifiers and complex analysis, one can go from smooth functions to continuous functions.

Notice that Steps 3, 4–5, and 6 resemble the three steps in the proof of the stronger Horn–Loewner theorem 17.1.

In this section, we complete the first two steps in the proof.

Step 1: For the first step, suppose $A = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \in \mathbb{P}_2((0,\infty))$ with $a \geq c$. There are three cases. First, if $b = \sqrt{ac}$ then use $\mu = a\delta_{b/a}$, since $0 < b/a \leq 1$.

Henceforth, assume $0 < b < \sqrt{ac} \leq a$. (In particular, $2b < 2\sqrt{ac} \leq a + c$.) The second case is if $b > c$; we then find $t > 0$, such that $A - t\mathbf{1}_{2\times 2}$ is singular. This condition amounts to a linear equation in t, with solution (to be verified by the reader)

$$t = \frac{ac - b^2}{a + c - 2b} > 0.$$

Then $c - t = \dfrac{(b-c)^2}{a+c-2b} > 0$, so $a - t, b - t > 0$ and $A = \begin{pmatrix} s_0(\mu) & s_1(\mu) \\ s_1(\mu) & s_2(\mu) \end{pmatrix}$, where

$$\mu = \frac{ac - b^2}{a + c - 2b}\delta_1 + \frac{(a-b)^2}{a+c-2b}\delta_{\frac{b-c}{a-b}}, \quad \text{with } \frac{b-c}{a-b} \in (0,1).$$

The third case is when $0 < b \leq c \leq \sqrt{ac} \leq a$, with $b < \sqrt{ac}$. Now find $t > 0$, such that the matrix $\begin{pmatrix} a-t & b+t \\ b+t & c-t \end{pmatrix} \in \mathbb{P}_2((0,\infty))$ and is singular. To do so requires solving a linear equation, which yields:

$$t = \frac{ac-b^2}{a+c+2b}, \qquad c-t = \frac{(b+c)^2}{a+c+2b}, \qquad a-t = \frac{(a+b)^2}{a+c+2b}, \qquad b+t = \frac{(a+b)(b+c)}{a+c+2b},$$

and all of these are strictly positive. So $a,b,c > 0$ are the first three moments of

$$\mu = \frac{ac-b^2}{a+c+2b}\delta_{-1} + \frac{(a+b)^2}{a+c+2b}\delta_{\frac{b+c}{a+b}}, \quad \text{with } \frac{b+c}{a+b} \in (0,1].$$

Step 2: Observe that the hypotheses of the stronger Horn–Loewner theorem 17.1 (for all n) can be rephrased as saying that $f[-]$ sends the rank-1 matrices in $\mathbb{P}_2(I)$ and the Toeplitz matrices in $\mathbb{P}_2(I)$ to $\mathbb{P}_2(\mathbb{R})$, and that assertion (4) in Theorem 19.15 holds for all measures $a\delta_1 + b\delta_{u_0}$, where $u_0 \in (0,1)$ is fixed and $a,b > 0$. By Step 1 and the hypotheses, we can apply the stronger Horn–Loewner theorem in our setting for each $n \geq 3$, hence $f|_{(0,\infty)}$ is smooth and absolutely monotonic. As in the proof of the stronger Vasudeva theorem 19.1, extend f continuously to the origin, say to a function $\widetilde{f}$, and apply Bernstein's theorem 19.3. It follows that $\widetilde{f}|_{[0,\infty)}$ is a power series with non-negative Maclaurin coefficients, and Step 2 follows by restricting to $\widetilde{f}|_{(0,\infty)} = f|_{(0,\infty)}$. $\square$

Remark 19.18. From Step 2, it follows that assertions (1) and (2) in the stronger Vasudeva theorem 19.1 can be further weakened, to deal only with the rank-1 matrices in $\mathbb{P}_2(I)$, the Toeplitz matrices in $\mathbb{P}_2(I)$, and with the (Hankel TN moment matrices of) measures $a\delta_1 + b\delta_{u_0}$, for a *single* fixed $u_0 \in (0,1)$ and all $a,b \geq 0$ with $a+b > 0$.

20. Proof of stronger Schoenberg Theorem: I. Continuity. The positivity-certificate trick.

We continue with the proof of the stronger Schoenberg theorem 19.10. Previously, we have shown the first two of the six steps in the proof (these are listed following Theorem 19.15).

Step 3: The next step is to show that if assertion (4) (or (5)) in Theorem 19.15 holds, then f is continuous on $\mathbb{R}$. Notice from Steps 1 and 2 of the proof that f is absolutely monotonic, hence continuous, on $(0,\infty)$.

20.1. **Integration trick and proof of continuity.** At this stage, we transition to moment sequence preservers, via Hamburger's theorem 19.14. The following "integration trick" will be used repeatedly in what follows: Suppose $p(t)$ is a real polynomial that takes non-negative values for $t \in [-1,1]$. Write $p(t) = \sum_{k=0}^\infty a_k t^k$ (with only finitely many a_k non-zero, but not necessarily all positive, note). If $\mu \geq 0$ is an admissible measure – in particular, non-negative by Definition 19.12 – then by assumption and Hamburger's theorem we have $f(s_k(\mu)) = s_k(\sigma_\mu)$ $\forall k \geq 0$, for some admissible measure $\sigma_\mu \geq 0$ on $\mathbb{R}$, where $f : \mathbb{R} \to \mathbb{R}$ satisfies Theorem 19.15(4) or (5). Now **assuming** σ_μ is supported on $[-1,1]$ (which is not a priori clear from the hypotheses), we have

$$0 \leq \int_{-1}^1 p(t)\ d\sigma_\mu = \sum_{k=0}^\infty \int_{-1}^1 a_k t^k\ d\sigma_\mu = \sum_{k=0}^\infty a_k s_k(\sigma_\mu) = \sum_{k=0}^\infty a_k f(s_k(\mu)). \tag{20.1}$$

Example 20.2. Suppose $p(t) = 1 - t^d$ on $[-1,1]$, for some integer $d \geq 1$. Then $f(s_0(\mu)) - f(s_d(\mu)) \geq 0$. As a further special case, if $\mu = a\delta_1 + b\delta_{u_0} + c\delta_{-1}$ as in Theorem 19.15(4), if σ_μ is supported on $[-1,1]$ then this would imply

$$f(a+b+c) \geq f(a + bu_0^d + c(-1)^d), \qquad \forall u_0 \in (0,1),\ a,b,c \geq 0.$$

It is not immediately clear how the preceding inequalities can be obtained by considering only the preservation of matrix positivity by $f[-]$ (or more involved such assertions). As we will explain shortly (see Section 20.2), this has connections to real algebraic geometry; in particular, to a well-known program of Hilbert.

Returning to the proof of continuity in Schoenberg's theorem, we suppose without further mention that f satisfies only Theorem 19.15(5) – and hence, is absolutely monotonic on $(0,\infty)$. We begin by showing two preliminary lemmas, which are used in the proof of continuity.

Lemma 20.3. *f is bounded on compact subsets of $\mathbb{R}$.*

Proof. If $K \subset \mathbb{R}$ is compact, say $K \subset [-M,M]$ for some $M > 0$, then note that $f|_{(0,\infty)}$ is non-decreasing, hence $0 \leq |f(x)| \leq f(M)$, $\forall x \in (0,M]$. Now apply $f[-]$ to the matrix $B := \begin{pmatrix} x & -x \\ -x & x \end{pmatrix}$, arising from $\mu = x\delta_{-1}$, with $x > 0$. The positivity of $f[B]$ implies $|f(-x)| \leq f(x) \leq f(M)$. Similarly considering $\mu = \frac{M}{2}\delta_1 + \frac{M}{2}\delta_{-1}$ shows that $|f(0)| \leq f(M)$. $\square$

Now say $\mu = a\delta_1 + b\delta_{u_0} + c\delta_{-1}$ as above, or more generally, μ is any non-negative measure supported in $[-1,1]$. It is easily seen that its moments $s_k(\mu)$, $k \geq 0$ are all uniformly bounded in absolute value – in fact, by the mass $s_0(\mu)$. Our next lemma shows that the converse is also true.

Lemma 20.4. *Given an admissible measure σ on $\mathbb{R}$, the following are equivalent:*

(1) The moments of σ are all uniformly bounded in absolute value.
(2) The measure σ is supported on $[-1,1]$.

Proof. As discussed above, (2) $\implies$ (1). To show the converse, suppose (1) holds but (2) fails. Then σ has positive mass in $(1,\infty)\sqcup(-\infty,-1)$. We obtain a contradiction in the first case; the proof is similar in the other case. Thus, suppose σ has positive mass on

$$(1,\infty) = [1+\tfrac{1}{1}, 1+\tfrac{1}{0}) \sqcup [1+\tfrac{1}{2}, 1+\tfrac{1}{1}) \sqcup \cdots \sqcup [1+\tfrac{1}{n+1}, 1+\tfrac{1}{n}) \sqcup \cdots,$$

where $1/0 := \infty$. Then $\sigma(I_n) > 0$ for some $n \geq 0$, where we denote $I_n := [1+\frac{1}{n+1}, 1+\frac{1}{n})$ for convenience. But now we obtain the desired contradiction:

$$s_{2k}(\sigma) = \int_{\mathbb{R}} x^{2k}\, d\sigma \geq \int_{1+\frac{1}{n+1}}^{1+\frac{1}{n}} x^{2k}\, d\sigma \geq \int_{1+\frac{1}{n+1}}^{1+\frac{1}{n}} (1+\tfrac{1}{n+1})^k\, d\sigma \geq \sigma(I_n)(1+\tfrac{1}{n+1})^k,$$

and this is not uniformly bounded over all $k \geq 0$. $\square$

With these basic lemmas in hand, we have:

Proof of Step 3 for the stronger Schoenberg theorem: continuity. Suppose $f:\mathbb{R}\to\mathbb{R}$ satisfies Theorem 19.15(4). (We explain in Section 20.3, how to weaken the hypotheses to Theorem 19.15(5).) Given a measure $\mu = a\delta_1 + b\delta_{u_0} + c\delta_{-1}$ for $u_0 > 0$ and $a,b,c \geq 0$, note that $|s_k(\mu)| \leq s_0(\mu) = a+b+c$. Hence, by Lemma 20.3, the moments $s_k(\sigma_\mu)$ are uniformly bounded over all k. By Lemma 20.4, it follows that σ_μ must be supported in $[-1,1]$. In particular, we can apply the integration trick (20.1) above.

We use this trick to prove continuity at $-\beta$ for $\beta \geq 0$. (By Step 2, this proves the continuity of f on $\mathbb{R}$.) Thus, fix $\beta \geq 0$, $u_0 \in (0,1)$, and $b > 0$, and define

$$\mu := (\beta + bu_0)\delta_{-1} + b\delta_{u_0}.$$

Let $p_{\pm,1}(t) := (1\pm t)(1-t^2)$; note that these polynomials are non-negative on $[-1,1]$. By the integration trick (20.1),

$$\begin{aligned}
& \int_{-1}^{1} p_{\pm,1}(t)\, d\sigma_\mu(t) \geq 0 \\
\implies\quad & s_0(\sigma_\mu) - s_2(\sigma_\mu) \geq \pm(s_1(\sigma_\mu) - s_3(\sigma_\mu)) \\
\implies\quad & f(s_0(\mu)) - f(s_2(\mu)) \geq |f(s_1(\mu)) - f(s_3(\mu))| \\
\implies\quad & f(\beta + b(1+u_0)) - f(\beta + b(u_0+u_0^2)) \geq |f(-\beta) - f(-\beta - bu_0(1-u_0^2))|.
\end{aligned}$$

Now let $b \to 0^+$. Then the left-hand side goes to zero by Step 2 (in the previous section), hence so does the right-hand side. This implies f is left continuous at $-\beta$ for all $\beta \geq 0$. To show f is right continuous at $-\beta$, use $\mu' := (\beta + bu_0^3)\delta_{-1} + b\delta_{u_0}$ instead of μ. $\square$

Remark 20.5. Akin to its use in proving the continuity of f, the integration trick (20.1) can also be used to prove the boundedness of f on compact sets $[-M,M]$, as in Lemma 20.3. To do so, work with the polynomials $p_{\pm,0}(t) := 1\pm t$, which are also non-negative on $[-1,1]$. Given $0 \leq x < M$, applying (20.1) to $\mu = M\delta_{x/M}$ and $\mu' = x\delta_{-1}$ shows Lemma 20.3.

20.2. **The integration trick explained: semi-algebraic geometry.** In Section 20.1, we used the following integration trick: if $\sigma \geq 0$ is a real measure supported in $[-1,1]$ with all moments finite, i.e., the Hankel moment matrix $H_\sigma := (s_{j+k}(\sigma))_{j,k=0}^\infty$ is positive semidefinite;

and if a polynomial $p(t) = \sum_{k \geq 0} a_k t^k$ is non-negative on $[-1,1]$, then

$$0 \leq \int_{-1}^{1} p(t)\, d\sigma = \sum_{k=0}^{\infty} \int_{-1}^{1} a_k t^k \, d\sigma = \sum_{k=0}^{\infty} a_k s_k(\sigma).$$

This integration trick is at the heart of the link between moment problems and (Hankel) matrix positivity. This trick is now explained; namely, how this integral inequality can be understood purely in terms of the positive semidefiniteness of H_σ. This also has connections to real algebraic geometry and Hilbert's seventeenth problem.

The basic point is as follows: if a d-variate polynomial (in one or several variables) is a sum of squares of real polynomials – also called a *sum-of-squares polynomial* – then it is automatically non-negative on $\mathbb{R}^d$. However, Hilbert showed in his 1888 paper [173] in *Math. Ann.* – following the doctoral dissertation of Hermann Minkowski – that for $d \geq 2$, there exist polynomials that are not sums of squares, yet are non-negative on $\mathbb{R}^d$. The first such example was constructed in 1967, and is the well-known Motzkin polynomial $M(x,y) = x^4y^2 + x^2y^4 - 3x^2y^2 + 1$ [267].[2,3,4] Such phenomena are also studied on polytopes (results of Farkas, Pólya, and Handelman), and on more general "semi-algebraic sets" including compact ones (results of Stengle, Schmüdgen, Putinar, and Vasilescu, among others).

Now given, say, a one-variable polynomial that is non-negative on a semi-algebraic set such as $[-1,1]$, one would like a *positivity certificate* for it, meaning a sum-of-squares representation mentioned above, or more generally, a *limiting sum-of-squares representation.* To make this precise, define the L^1*-norm*, or the *Wiener norm*, of a polynomial $p(t) = \sum_{k \geq 0} a_k t^k$ as:

$$\|p(t)\|_{1,+} := \sum_{k \geq 0} |a_k|. \tag{20.6}$$

One would thus like to find a sequence p_n of sum-of-squares polynomials, such that $\|p_n(t) - p(t)\|_{1,+} \to 0$ as $n \to \infty$. Two simple cases are if there exist polynomials $q_n(t)$, such that (1) $p_n(t) = q_n(t)^2 \ \forall n$ or (2) $p_n(t) = \sum_{k=1}^{n} q_k(t)^2 \ \forall n$.

---

[2]Indeed, as explained, e.g., in [309], by the AM–GM inequality we have $M(x,y) \geq 2t^3 - 3t^2 + 1 = (2t^2 - t - 1)(t-1)$, where $t = |xy|$. Now either $t \in [0,1]$, so both factors on the right are non-positive; or $t > 1$, so both factors are positive. Next, suppose $M(x,y) = \sum_j f_j(x,y)^2$ is a sum of squares; since $M(x,0) = M(0,y) = 1$, it follows that $f_j(x,0), f_j(0,y)$ are constants, and hence the f_j are of the form $f_j(x,y) = a_j + b_j xy + c_j x^2 y + d_j xy^2$. Now equating the coefficient of x^2y^2 in $M = \sum_j f_j^2$ gives: $-3 = \sum_j b_j^2$, a contradiction.

[3]Hilbert then showed in [174] that every non-negative polynomial on $\mathbb{R}^2$ is a sum of four squares of rational functions; e.g., the Motzkin polynomial equals $\frac{x^2y^2(x^2+y^2-2)^2(x^2+y^2+1)+(x^2-y^2)^2}{(x^2+y^2)^2}$. For more on this problem, see, e.g., [309].

[4]The reader may recall the name of Motzkin from Theorem 3.22, in an entirely different context. As a historical digression, we mention several relatively "disconnected" areas of mathematics, in *all* of which, remarkably, Motzkin made fundamental contributions. His thesis [263] was a landmark work in the area of linear inequalities/linear programming, introducing in particular the *Motzkin transposition theorem* and the *Fourier–Motzkin Elimination (FME) algorithm.* Additionally, he proved in the same thesis the fundamental fact in geometric combinatorics that, a convex polyhedral set is the Minkowski sum of a compact (convex) polytope and a convex polyhedral cone. Third, in his thesis Motzkin also characterized the matrices that satisfy the variation diminishing property; see Theorem 3.22.

Then in [264], Motzkin studied what is now called the *Motzkin number* in combinatorics: this is the number of different ways to draw non-intersecting chords between n marked points on a circle. In [265], he provided the first example of a principal ideal domain that is not a Euclidean domain: $\mathbb{Z}[(1+\sqrt{-19})/2]$. In [266], he provided an ideal-free short proof of Hilbert's Nullstellensatz, together with degree bounds. Motzkin also provided in [267] the aforementioned polynomial $M(x,y)$ in connection to Hilbert's seventeenth problem.

How does this connect to matrix positivity? It turns out that in our given situation, what is required is precisely a positivity certificate. For example, say $p(t) = (3-t)^2 = 9-6t+t^2 \geq 0$ on $\mathbb{R}$. Then

$$\begin{aligned}\int_{-1}^{1} p \; d\sigma = 9s_0(\sigma) - 6s_1(\sigma) + s_2(\sigma) &= (3,-1) \begin{pmatrix} s_0(\sigma) & s_1(\sigma) \\ s_1(\sigma) & s_2(\sigma) \end{pmatrix} (3,-1)^T \\ &= (3e_0 - e_1)^T H_\sigma (3e_0 - e_1),\end{aligned} \tag{20.7}$$

where $e_0 = (1,0,0,\dots)^T, e_1 = (0,1,0,0,\dots)^T, \dots$ comprise the standard basis for $\mathbb{R}^{\mathbb{N}\sqcup\{0\}}$, and H_σ is the semi-infinite, positive semidefinite Hankel moment matrix for σ. From this calculation, it follows that $\int_{-1}^{1} p \; d\sigma$ is non-negative – and this holds more generally, whenever there exists a (limiting) sum-of-squares representation for p.

We now prove the existence of such a limiting sum-of-squares representation in two different ways for general polynomials $p(t)$ that are non-negative on $[-1,1]$, and in a constructive third way for the special family of polynomials

$$p_{\pm,n}(t) := (1 \pm t)(1-t^2)^n, \qquad n \geq 0.$$

(Note, we used $p_{\pm,0}$ and $p_{\pm,1}$ to prove the local boundedness and continuity of f on $\mathbb{R}$, respectively; and the next section uses $p_{\pm,n}$ to prove that smoothness implies real analyticity.)

Proof 1: We claim more generally that for any dimension $d \geq 1$, every polynomial that is non-negative on $[-1,1]^d$ has a limiting sum-of-squares representation. This is proved at the end of the 1976 paper [39] of Berg, Christensen, and Ressel in *Math. Ann.*

Proof 2: Here is a constructive proof of a positivity certificate for the polynomials $p_{\pm,n}(t) = (1 \pm t)(1-t^2)^n$, $n \geq 0$. (It turns out, we only need to work with these in order to show the stronger Schoenberg theorem.) First, notice that

$$\begin{aligned} p_{+,0}(t) &= (1+t), & p_{-,0}(t) &= (1-t), \\ p_{+,1}(t) &= (1-t)(1+t)^2, & p_{-,1}(t) &= (1+t)(1-t)^2, \\ p_{+,2}(t) &= (1+t)(1-t^2)^2, & p_{-,2}(t) &= (1-t)(1-t^2)^2, \end{aligned} \tag{20.8}$$

and so on. Thus, if we show that $p_{\pm,0}(t) = 1 \pm t$ are limits of sum-of-squares polynomials, then so are $p_{\pm,n}(t)$ for all $n \geq 0$ (where limits are taken in the Wiener norm). But we have

$$\begin{aligned} \frac{1}{2}(1 \pm t)^2 &= \frac{1}{2} \pm t + \frac{t^2}{2}, \\ \frac{1}{4}(1-t^2)^2 &= \frac{1}{4} - \frac{t^2}{2} + \frac{t^4}{4}, \\ \frac{1}{8}(1-t^4)^2 &= \frac{1}{8} - \frac{t^4}{4} + \frac{t^8}{8}, \end{aligned} \tag{20.9}$$

and so on. Adding the first k of these equations shows that the partial sum

$$p_k^{\pm}(t) := (1 - \frac{1}{2^k}) \pm t + \frac{t^{2^k}}{2^k} = (1 \pm t) + \frac{t^{2^k} - 1}{2^k}$$

is a sum-of-squares polynomial, for every $k \geq 1$. This provides a positivity certificate for $1 \pm t$, as desired. It also implies the sought-for interpretation of the integration trick in Step 3 above:

$$\left| \int_{-1}^{1} [p_k^{\pm}(t) - (1 \pm t)] \; d\sigma \right| \leq \int_{-1}^{1} |p_k^{\pm}(t) - (1 \pm t)| \; d\sigma \leq \int_{-1}^{1} \frac{1}{2^k} \; d\sigma + \int_{-1}^{1} \frac{t^{2^k}}{2^k} \; d\sigma \leq \frac{1}{2^k} \cdot 2s_0(\sigma),$$

which goes to 0 as $k \to \infty$. Hence, using the notation following (20.7),

$$\int_{-1}^{1} (1 \pm t)\, d\sigma = \lim_{k \to \infty} \int_{-1}^{1} p_k^{\pm}(t)\, d\sigma = \frac{1}{2}(e_0 \pm e_1)^T H_\sigma (e_0 \pm e_1) + \sum_{j=2}^{\infty} \frac{1}{2^j} (e_0 - e_{2^{j-1}})^T H_\sigma (e_0 - e_{2^{j-1}}),$$

and this is non-negative because H_σ is positive semidefinite.

Proof 3: If we only want to interpret the integration trick (20.1) in terms of the positivity of the Hankel moment matrix H_σ, then the restriction of using the Wiener norm $\| \cdot \|_{1,+}$ can be relaxed, and one can work instead with the weaker notion of the uniform norm. With this metric, we claim more generally that *every continuous function $f(t_1, \dots, t_d)$ that is non-negative on a compact subset $K \subset \mathbb{R}^d$ has a limiting sum-of-squares representation on K.* (Specialized to $d = 1$ and $K = [-1, 1]$, this proves the integration trick.)

To see the claim, observe that $\sqrt{f(t_1, \dots, t_d)} : K \to [0, \infty)$ is continuous, so by the Stone–Weierstrass theorem, there exists a polynomial sequence q_n converging uniformly to $\sqrt{f}$ in $L^\infty(K)$. Thus, $q_n^2 \to f$ in $L^\infty(K)$, as desired. Explicitly, if $d = 1$ and $q_n(t) = \sum_{k=0}^{\infty} c_{n,k} t^k$, then define the semi-infinite vectors

$$\mathbf{u}_n := (c_{n,0}, c_{n,1}, \dots)^T, \qquad n \geq 1.$$

Now compute for any admissible measure σ supported in K:

$$\int_K f\ d\sigma = \lim_{n \to \infty} \int_K q_n^2(t)\ d\sigma = \lim_{n \to \infty} \mathbf{u}_n^T H_\sigma \mathbf{u}_n \geq 0, \tag{20.10}$$

which is a positivity certificate for all continuous, non-negative functions on compact $K \subset \mathbb{R}$.

This reasoning extends to all dimensions $d \geq 1$ and compact $K \subset \mathbb{R}^d$, by Lemma 2.24.

20.3. **From the integration trick to the positivity-certificate trick.** Proof 2 in Section 20.2 is the key to understanding why Hamburger's theorem is not required to prove the stronger Schoenberg theorem 19.15 (namely, (5) $\implies$ (3)). Specifically, we only need to use the following fact:

For each fixed $n \geq 0$, if $\sum_k a_k t^k$ is the expansion of $p_{\pm,n}(t) = (1 \pm t)(1 - t^2)^n \geq 0$ on $[-1, 1]$, then $\sum_k a_k f(s_k(\mu)) \geq 0$.

This was derived above using the integration trick (20.1) via the auxiliary admissible measure σ_μ, which exists by Theorem 19.15(4). We now explain a workaround via a related "positivity-certificate trick" that requires using only that $f[H_\mu]$ is positive semidefinite, hence allowing us to work with the weaker hypothesis, Theorem 19.15(5) instead. In particular, one can avoid using Hamburger's theorem and requiring the existence of σ_μ.

The positivity-certificate trick is as follows:

Theorem 20.11. *Fix a semi-infinite real Hankel matrix $H = (f_{j+k})_{j,k \geq 0}$ that is positive semidefinite (i.e., its principal minors are positive semidefinite), with all entries f_j uniformly bounded. If a polynomial $p(t) = \sum_{j \geq 0} a_j t^j$ has a positivity certificate – i.e., a Wiener-limiting sum-of-squares representation – then $\sum_{j \geq 0} a_j f_j \geq 0$.*

According to Proof 2 in Section 20.2 (see the discussion around (20.9)), Theorem 20.11 applies to $p = p_{\pm,n}$ for all $n \geq 0$ and $H = f[H_\mu]$, where f, μ are as in Theorem 19.15(5). This implies the continuity of the entrywise positivity preserver f in the above discussion, and also suffices to complete the proof in the next two sections, of the stronger Schoenberg theorem (Theorem 19.10).

Proof of Theorem 20.11. As an illustrative special case, if $p(t)$ is the square of a polynomial $q(t) = \sum_{j \geq 0} c_j t^j$, then as in (20.7),

$$\sum_{j \geq 0} a_j f_j = \sum_{j,k \geq 0} c_j c_k f_{j+k} = \mathbf{u}^T H \mathbf{u} \geq 0, \quad \text{where } \mathbf{u} = (c_0, c_1, \dots)^T.$$

By additivity, the result therefore also holds for a sum of squares of polynomials. The subtlety in working with a limiting sum-of-squares representation is that the degrees of each sum-of-squares polynomial in the limiting sequence need not be uniformly bounded. Nevertheless, suppose in the Wiener norm (20.6) that

$$p(t) = \lim_{n \to \infty} q_n(t), \quad \text{where } q_n(t) = \sum_{k=0}^{K_n} q_{n,k}(t)^2$$

is a sum of squares of polynomials for each n.

Define the linear functional Ψ_H (given the Hankel matrix H) that sends a polynomial $p(t) = \sum_{j \geq 0} a_j t^j$ to the scalar $\Psi_H(p) := \sum_{j \geq 0} a_j f_j$. Now define the vectors $\mathbf{u}_{n,k}$ via

$$q_{n,k}(t) = \sum_{j \geq 0} q_{n,k}^{[j]} t^j \in \mathbb{R}[t], \quad \mathbf{u}_{n,k} := (q_{n,k}^{[0]}, q_{n,k}^{[1]}, \dots)^T.$$

Similarly, define $q_n(t) = \sum_{j \geq 0} q_n^{[j]} t^j$. Then for all $n \geq 1$,

$$\sum_{j \geq 0} q_n^{[j]} f_j = \Psi_H(q_n) = \sum_{k=0}^{K_n} \Psi_H(q_{n,k}^2) = \sum_{k=0}^{K_n} \mathbf{u}_{n,k}^T H \mathbf{u}_{n,k} \geq 0.$$

Finally, taking the limit as $n \to \infty$, and writing $p(t) = \sum_{j \geq 0} a_j t^j$, we claim that

$$\sum_{j \geq 0} a_j f_j = \lim_{n \to \infty} \sum_{j \geq 0} q_n^{[j]} f_j \geq 0.$$

Indeed, the (first) equality holds because if $M \geq \sup_j |f_j|$ is a uniform (and finite) upper bound, then

$$\left| \sum_{j \geq 0} q_n^{[j]} f_j - \sum_{j \geq 0} a_j f_j \right| \leq \sum_{j \geq 0} |q_n^{[j]} - a_j| \, |f_j| \leq M \|q_n - p\|_{1,+},$$

and this goes to zero as $n \to \infty$. $\square$

21. Proof of stronger Schoenberg Theorem: II. Smoothness implies real analyticity.

Having explained the positivity-certificate trick, we return to the proof of the stronger Schoenberg theorem. The present goal is to prove that if a smooth function $f : \mathbb{R} \to \mathbb{R}$ satisfies assertion (5) in Theorem 19.15, then f is real analytic and hence satisfies assertion (3) in Theorem 19.10. (See Steps (4) and (5) in the list following Remark 19.17.) To show these results, we first discuss the basic properties of real analytic functions that are required in the proofs.

21.1. **Preliminaries on real analytic functions.**

Definition 21.1. Suppose $I \subset \mathbb{R}$ is an open interval, and $f : I \to \mathbb{R}$ is smooth, denoted $f \in C^\infty(I)$. Recall that the *Taylor series* of f at a point $x \in I$ is

$$(Tf)_x(y) := \sum_{j=0}^\infty \frac{f^{(j)}(x)}{j!}(y-x)^j, \qquad y \in I,$$

if this sum converges at y. Notice that this sum is not equal to $f(y)$ in general.

Next, we say that f is *real analytic on* I, denoted $f \in C^\omega(I)$, if $f \in C^\infty(I)$ and for all $x \in I$ there exists $\delta_x > 0$, such that the Taylor series of f at x converges to f on $(x - \delta_x, x + \delta_x)$.

Clearly, real analytic functions on I form a real vector space. Less obvious is the following useful property, which is stated without proof:

Proposition 21.2. *Real analytic functions are closed under composition. More precisely, if* $I \xrightarrow{f} J \xrightarrow{g} \mathbb{R}$, *and* f, g *are real analytic on their domains, then so is* $g \circ f$ *on* I.

We also develop a few preliminary results on real analytic functions, which are needed to prove the stronger Schoenberg theorem. We begin with an example of real analytic functions, which depicts what happens in our setting.

Lemma 21.3. *Suppose* $I = (0, R)$ *for* $0 < R \leq \infty$, *and* $f(x) = \sum_{k=0}^\infty c_k x^k$ *on* I, *where* $c_k \geq 0$ $\forall k$. *Then* $f \in C^\omega(I)$, *and* $(Tf)_a(x)$ *converges whenever* $a, x \in I$ *are such that* $|x - a| < R - a$.

In particular, if $R = \infty$ and $a > 0$, then $(Tf)_a(x) \to f(x)$ on the domain of f.

Proof. Note that $\sum_{k=0}^\infty c_k x^k$ converges on $(-R, R)$. Thus, we denote this extension to $(-R, R)$ also by f, and show more generally that $(Tf)_a(x)$ converges to $f(x)$ for $|x - a| < R - a$, $a \in [0, R)$ (whenever f is defined at x). Indeed,

$$f(x) = \sum_{k=0}^\infty c_k((x-a)+a)^k = \sum_{k=0}^\infty \sum_{j=0}^k \binom{k}{j} c_k (x-a)^j a^{k-j}.$$

Notice that this double sum is absolutely convergent, since

$$\sum_{k=0}^\infty \sum_{j=0}^k \binom{k}{j} c_k |x-a|^j a^{k-j} = f(a + |x-a|) < \infty.$$

Hence, we can rearrange the double sum (e.g., by Fubini's theorem), to obtain

$$f(x) = \sum_{j=0}^\infty \left(\sum_{m=0}^\infty \binom{m+j}{j} c_{m+j} a^m \right)(x-a)^j = \sum_{j=0}^\infty \frac{f^{(j)}(a)}{j!}(x-a)^j = (Tf)_a(x)$$

using standard properties of power series. In particular, f is real analytic on I. $\square$

We also require the following well-known result on zeros of real analytic functions:

Theorem 21.4 (Identity theorem)**.** *Suppose $I \subset \mathbb{R}$ is an open interval and $f, g : I \to \mathbb{R}$ are real analytic. If the subset of I where $f = g$ has an accumulation point in I, then $f \equiv g$ on I.*

In other words, the zeros of a non-zero (real) analytic function form a discrete set.

Proof. Without loss of generality, we may suppose $g \equiv 0$. Suppose $c \in I$ is an accumulation point of the zero set of f. Expand f locally at c into its Taylor series, and claim that $f^{(k)}(c) = 0$ for all $k \geq 0$. Indeed, suppose for contradiction that

$$f^{(0)}(c) = \cdots = f^{(k-1)}(c) = 0 \neq f^{(k)}(c)$$

for some $k \geq 0$. Then,

$$\frac{f(x)}{(x-c)^k} = \frac{f^{(k)}(c)}{k!} + o(x-c),$$

so f is non-zero close to c, and this contradicts the hypotheses. Thus, $f^{(k)}(c) = 0 \ \forall k \geq 0$, which in turn implies that $f \equiv 0$ on an open interval around c.

Now consider the set $I_0 := \{x \in I : f^{(k)}(x) = 0 \ \forall k \geq 0\}$. Clearly I_0 is a closed subset of I. Moreover, if $c_0 \in I_0$ then $f \equiv (Tf)_{c_0} \equiv 0$ near c_0, hence the same happens at any point near c_0 as well. Thus, I_0 is also an open subset of I. Since I is connected, $I_0 = I$, and $f \equiv 0$. $\square$

21.2. **Proof of the stronger Schoenberg theorem for smooth functions.** We continue with the proof of the stronger Schoenberg theorem ((5) $\implies$ (2) in Theorems 19.10 and 19.15).

Akin to the proof of the stronger Horn–Loewner theorem 17.1, we have shown that any function satisfying the hypotheses in Theorem 19.15(5) must be continuous. Hence, by the first two steps in the proof – listed after Remark 19.17 – we have that $f(x) = \sum_{k=0}^{\infty} c_k x^k$ on $[0, \infty)$, with all $c_k \geq 0$.

Again, similar to the proof of the stronger Horn–Loewner theorem 17.1, we next prove the stronger Schoenberg theorem for smooth functions. The key step here is:

Theorem 21.5. *Let $f \in C^\infty(\mathbb{R})$ be as in the preceding discussion, and define the family of smooth functions*

$$H_a(x) := f(a + e^x), \qquad a, x \in \mathbb{R}.$$

Then H_a is real analytic on $\mathbb{R}$, for all $a \in \mathbb{R}$.

For ease of exposition, we break the proof into several steps.

Lemma 21.6. *For all $n \geq 1$, we have*

$$H_a^{(n)}(x) = a_{n,1} f'(a + e^x) e^x + a_{n,2} f''(a + e^x) e^{2x} + \cdots + a_{n,n} f^{(n)}(a + e^x) e^{nx},$$

where $a_{n,j}$ is a positive integer for all $1 \leq j \leq n$.

Proof and remarks. One shows by induction on $n \geq 1$ (with the base case of $n = 1$ immediate) that the array $a_{n,j}$ forms a weighted variant of Pascal's triangle, in that

$$a_{n,j} = \begin{cases} 1, & \text{if } j = 1, n, \\ a_{n-1,j-1} + j a_{n-1,j}, & \text{otherwise.} \end{cases}$$

This concludes the proof. Notice that some of the entries of the array $a_{n,j}$ are easy to compute inductively:

$$a_{n,1} = 1, \qquad a_{n,2} = 2^{n-1} - 1, \qquad a_{n,n-1} = \binom{n}{2}, \qquad a_{n,n} = 1.$$

An interesting combinatorial exercise may be to seek a closed-form expression and a combinatorial interpretation for the other entries. □

Lemma 21.7. *We have the following bound:*

$$|H_a^{(n)}(x)| \leq H_{|a|}^{(n)}(x), \qquad \forall a, x \in \mathbb{R},\ n \in \mathbb{Z}^{\geq 0}. \tag{21.8}$$

Proof. By Lemma 21.6 we have that $H_{|a|}^{(n)}(x) \geq 0$ for all a, x, n as in (21.8), so it remains to show the inequality. For this, we assume $a < 0$, and use the positivity-certificate trick from the previous section – i.e., Theorem 20.11, applied to the polynomials

$$p_{\pm,n}(t) := (1 \pm t)(1 - t^2)^n, \qquad n \geq 0$$

and the admissible measure

$$\mu := |a|\delta_{-1} + e^x \delta_{e^{-h}}, \qquad a, h > 0,\ x \in \mathbb{R}.$$

Notice that $p_{\pm,n} \geq 0$ on $[-1, 1]$. Hence, by Theorem 20.11 – and akin to the calculation in the previous section to prove continuity – we get

$$\sum_{k=0}^{n} \binom{n}{k} (-1)^{n-k} f(|a| + e^{x-2kh}) \geq \left| \sum_{k=0}^{n} \binom{n}{k} (-1)^{n-k} f(a + e^{x-(2k+1)h}) \right|.$$

Dividing both sides by $(2h)^n$ and sending $h \to 0^+$, we obtain

$$H_{|a|}^{(n)}(x) \geq |H_a^{(n)}(x)|.$$

□

Remark 21.9. In this computation, we do not need to use the measures $\mu = |a|\delta_{-1} + e^x \delta_{e^{-h}}$ for all $h > 0$. It suffices to fix a single $u_0 \in (0, 1)$ and consider the sequence $h_n := -\log(u_0)/n$, so we work with $\mu = |a|\delta_{-1} + e^x \delta_{u_0^{1/n}}$ (supported at $1, u_0^{1/n}$) for $a > 0, x \in \mathbb{R}, n \geq 1$.

Lemma 21.10. *For all integers $n \geq 0$, the assignment $(a, x) \mapsto H_a^{(n)}(x)$ is non-decreasing in both $a \geq 0$ and $x \in \mathbb{R}$. In particular if $a \geq 0$, then H_a is absolutely monotonic on $\mathbb{R}$, and its Taylor series at $b \in \mathbb{R}$ converges absolutely at all $x \in \mathbb{R}$.*

Proof. The monotonicity in $a \geq 0$ follows from the absolute monotonicity of $f|_{[0,\infty)}$ mentioned at the start of Section 21.2. The monotonicity in x for a fixed $a \geq 0$ follows because $H_a^{(n+1)}(x) \geq 0$ by Lemma 21.6.

To prove the (absolute) convergence of $(TH_a)_b$ at $x \in \mathbb{R}$, notice that

$$|(TH_a)_b(x)| = \left| \sum_{n=0}^{\infty} H_a^{(n)}(b) \frac{(x-b)^n}{n!} \right| \leq \sum_{n=0}^{\infty} H_a^{(n)}(b) \frac{|x-b|^n}{n!} = (TH_a)_b(b + |x - b|).$$

We claim that this final (Taylor) series is bounded above by $H_a(b + |x - b|)$, which would complete the proof. Indeed, by Taylor's theorem, the nth Taylor remainder term for $H_a(b + |x - b|)$ can be written as (see, e.g., (19.7))

$$\int_b^{b+|x-b|} \frac{(b + |x - b| - t)^n}{n!} H_a^{(n+1)}(t)\ dt,$$

which is non-negative, see above. Taking $n \to \infty$ shows the claim and completes the proof. □

Now we can prove the real analyticity of H_a:

Proof of Theorem 21.5. Fix scalars $a, \delta > 0$. We show that for all $b \in [-a, a]$ and $x \in \mathbb{R}$, the nth remainder term for the Taylor series TH_b around the point x converges to zero as $n \to \infty$, uniformly near x. More precisely, define

$$\Psi_n(x) := \sup_{y \in [x-\delta, x+\delta]} |R_n((TH_b)_x)(y)|.$$

We then claim $\Psi_n(x) \to 0$ as $n \to \infty$ for all x. This will imply that at all $x \in \mathbb{R}$, $(TH_b)_x$ converges to H_b on a neighborhood of radius δ. Moreover, this holds for all $\delta > 0$ and at all $b \in [-a, a]$ for all $a > 0$.

Thus, it remains to prove for each $x \in \mathbb{R}$ that $\Psi_n(x) \to 0$ as $n \to \infty$. By the above results, we have:

$$|H_b^{(n)}(y)| \leq H_{|b|}^{(n)}(y) \leq H_a^{(n)}(y) \leq H_a^{(n)}(x+\delta), \qquad \forall b \in [-a, a],\ y \in [x-\delta, x+\delta],\ n \in \mathbb{Z}^{\geq 0}.$$

Using a standard estimate for the Taylor remainder, for all b, y, n as above, it follows that

$$|R_n((TH_b)_x)(y)| \leq H_a^{(n+1)}(x+\delta)\frac{|y-x|^{n+1}}{(n+1)!} \leq H_a^{(n+1)}(x+\delta)\frac{\delta^{n+1}}{(n+1)!}.$$

But the right-hand term goes to zero by the calculation in Lemma 21.10, since

$$0 \leq \sum_{n=-1}^{\infty} H_a^{(n+1)}(x+\delta)\frac{\delta^{n+1}}{(n+1)!} \leq H_a(x+\delta+\delta) = f(a + e^{x+2\delta}) < \infty.$$

Hence, we obtain

$$\lim_{n\to\infty} \sup_{y \in [x-\delta, x+\delta]} |R_n((TH_b)_x)(y)| \to 0, \qquad \forall x \in \mathbb{R},\ \delta > 0,\ b \in [-a, a],\ a > 0.$$

From above, this shows that the Taylor series of H_b converges locally to H_b at all $x \in \mathbb{R}$, for all b as desired. (In fact, the "local" neighborhood of convergence around x is all of $\mathbb{R}$.) $\square$

With the above analysis in hand, we can prove Steps 4 and 5 of the proof of the stronger Schoenberg theorem (see the list after Remark 19.17):

Suppose $f : \mathbb{R} \to \mathbb{R}$ satisfies assertion (5) of Theorem 19.15.

(4) *If f is smooth on $\mathbb{R}$, then f is real analytic on $\mathbb{R}$.*
(5) *If f is real analytic on $\mathbb{R}$, then $f(x) = \sum_{k=0}^{\infty} c_k x^k$ on $\mathbb{R}$, with $c_k \geq 0\ \forall k$.*

Proof of Step 4 for the stronger Schoenberg theorem. Given $x \in \mathbb{R}$, we want to show that the Taylor series $(Tf)_x$ converges to f locally around x. Choose $a > |x|$ and define

$$L_a(y) := \log(a+y) = \log(a) + \log(1 + y/a), \quad y \in (-a, a).$$

This is real analytic on $(-a, a)$ (e.g., akin to Lemma 21.3). Hence, by Proposition 21.2 and Theorem 21.5, the composite

$$y \xrightarrow{L_a} \log(a+y) \xrightarrow{H_{-a}} H_{-a}(L_a(y)) = f(-a + \exp(\log(a+y))) = f(y)$$

is also real analytic on $(-a, a)$, so around $x \in \mathbb{R}$. $\square$

Proof of Step 5 for the stronger Schoenberg theorem. By Step 4, f is real analytic on $\mathbb{R}$; and, by Steps 1 and 2 $f(x) = \sum_{k=0}^{\infty} c_k x^k$ on $(0, \infty)$, with $c_k \geq 0\ \forall k$. Let $g(x) := \sum_{k=0}^{\infty} c_k x^k \in C^{\omega}(\mathbb{R})$. Since $f \equiv g$ on $(0, \infty)$, it follows by the Identity Theorem 21.4 that $f \equiv g$ on $\mathbb{R}$. $\square$

22. Proof of stronger Schoenberg Theorem: III. Complex analysis. Further remarks.

We can now complete the proof of the final **Step 6** (listed after Remark 19.17) of the stronger Schoenberg theorem. Namely, suppose $f : \mathbb{R} \to \mathbb{R}$ is such that for each measure

$$\mu = a\delta_1 + b\delta_{u_0} + c\delta_{-1}, \quad \text{with } u_0 \in (0,1),\ a,b,c \geq 0,$$

with semi-infinite Hankel moment matrix H_μ, the matrix $f[H_\mu]$ is positive semidefinite.

Under these assumptions, we have previously shown (in Steps 1, 2; 3; 4, 5 respectively):

- There exist real scalars $c_0, c_1, \dots \geq 0$ such that $f(x) = \sum_{k=0}^\infty c_k x^k$ for all $x \in (0,\infty)$.
- f is continuous on $\mathbb{R}$.
- If f is smooth, then $f(x) = \sum_{k=0}^\infty c_k x^k$ on $\mathbb{R}$.

We now complete the proof by showing that one can pass from smooth functions to continuous functions. The tools we will use are the "three Ms": Montel, Morera, and Mollifiers. We first discuss some basic results in complex analysis that are required.

22.1. Tools from complex analysis.

Definition 22.1. Suppose $D \subset \mathbb{C}$ is open and $f : D \to \mathbb{C}$ is a continuous function.

1. (Holomorphic.) A function f is *holomorphic at a point* $z \in D$ if the limit $\lim_{y \to z} \frac{f(y)-f(z)}{y-z}$ exists. A function f is *holomorphic on* D if it is holomorphic at every point of D.
2. (Complex analytic.) f is said to be *complex analytic around* $c \in D$ if f can be expressed as a power series locally around c, which converges to $f(z)$ for every z sufficiently close to c. Similarly, f is *analytic on* D if it is so at every point of D.
3. (Normal.) Let $\mathcal{F}$ be a family of holomorphic functions : $D \to \mathbb{C}$. Then $\mathcal{F}$ is *normal* if given any compact $K \subset D$ and a sequence $\{f_n : n \geq 1\} \subset \mathcal{F}$, there exists a subsequence f_{n_k} and a function $f : K \to \mathbb{C}$, such that $f_{n_k} \to f$ uniformly on K.

Remark 22.2. Note that it is not specified that the limit function f be holomorphic. However, this will turn out to be the case, as we shall see presently.

We use without proof the following results (and Cauchy's theorem, which we do not state):

Theorem 22.3. *Let $D \subset \mathbb{C}$ be an open subset.*

1. *A function $f : D \to \mathbb{C}$ is holomorphic if and only if f is complex analytic.*
2. *(Montel.) Let $\mathcal{F}$ be a family of holomorphic functions on D. If $\mathcal{F}$ is uniformly bounded on D, then $\mathcal{F}$ is normal on D.*
3. *(Morera.) Suppose that for every closed oriented piecewise C^1 curve γ in D, we have $\oint_\gamma f\ dz = 0$. Then f is holomorphic on D.*

22.2. Proof of the stronger Schoenberg theorem: conclusion.

Let $f : \mathbb{R} \to \mathbb{R}$ be as described above; in particular, f is continuous on $\mathbb{R}$ and absolutely monotonic on $(0,\infty)$. As discussed in the proof of the stronger Horn–Loewner theorem 17.1, we mollify f with the family $\phi_\delta(u) = \phi(u/\delta)$ for $\delta > 0$ as in Proposition 18.5. As shown in (18.8), f_δ satisfies assertion (5) in Theorem 19.15, so (e.g., by the last bulleted point above, and Steps 4 and 5)

$$f_\delta(x) = \sum_{k=0}^\infty c_{k,\delta} x^k \ \forall x \in \mathbb{R}, \quad \text{with } c_{k,\delta} \geq 0 \ \forall k \geq 0,\ \delta > 0.$$

Since f_δ is a power series with infinite radius of convergence, it extends analytically to an entire function on $\mathbb{C}$ (see, e.g., Lemma 21.3). Let us call this f_δ as well; now define

$$\mathcal{F} := \{f_{1/n} : n \geq 1\}.$$

We claim that for any $0 < r < \infty$, the family $\mathcal{F}$ is uniformly bounded on the complex disk $D(0,r)$. Indeed, since $f_\delta \to f$ uniformly on $[0,r]$ by Proposition 18.5, we have that $|f_{1/n} - f|$ is uniformly bounded over all n and on $[0,r]$, say by $M_r > 0$. Now if $z \in D(0,r)$, then

$$|f_{1/n}(z)| \leq \sum_{k=0}^\infty c_{k,1/n}|z|^k = f_{1/n}(|z|) \leq M_r + f(|z|) \leq M_r + f(r) < \infty,$$

and this bound (uniform over $z \in D(0,r)$) does not depend on n.

By Montel's theorem, the previous claim implies that $\mathcal{F}$ is a normal family on $D(0,r)$ for each $r > 0$. Hence, on the closed disk $\overline{D}(0,r)$, there is a subsequence f_{1/n_l} with n_l increasing, which converges uniformly to some (continuous) $g = g_r$. Since f_{1/n_l} is holomorphic for all $l \geq 1$, by Cauchy's theorem we obtain for every closed oriented piecewise C^1 curve $\gamma \subset D(0,r)$

$$\oint_\gamma g_r \; dz = \oint_\gamma \lim_{l\to\infty} f_{1/n_l} \; dz = \lim_{l\to\infty} \oint_\gamma f_{1/n_l} \; dz = 0.$$

It follows by Morera's theorem that g_r is holomorphic, hence analytic, on $D(0,r)$. Moreover, $g_r \equiv f$ on $(-r,r)$ by the properties of mollifiers; thus, f is real analytic on $(-r,r)$ for every $r > 0$. Now apply the Identity Theorem 21.4 and use the power series for f on $(0,\infty)$. ☐

22.3. **Concluding remarks and variations.** We conclude with several generalizations of the above results. First, the results by Horn–Loewner, Vasudeva, and Schoenberg (more precisely, their stronger versions) that were shown in this part of the text, together with the proofs given above, can be refined to versions with *bounded domains* $(0,\rho)$ or $(-\rho,\rho)$ for $0 < \rho < \infty$. The small change is to use admissible measures with bounded mass

$$\mu = a\delta_1 + b\delta_{u_0} + c\delta_{-1}, \qquad \text{where } u_0 \in (0,1), \; a,b,c \geq 0$$

as above, but moreover, one now imposes the condition that $s_0(\mu) = a + b + c < \rho$.

Second, all of these results, including for bounded domains (i.e., masses of the underlying measures), can be extended to studying functions of several variables. In this case, given a domain $I \subset \mathbb{R}$ and integers $m, n \geq 1$, a function $f : I^m \to \mathbb{R}$ acts entrywise on an *m-tuple of $n \times n$ matrices* $A_1 = (a_{jk}^{(1)}), \dots, A_m = (a_{jk}^{(m)})$ in $I^{m\times m}$, via

$$f[A_1, \dots, A_m] := (f(a_{jk}^{(1)}, \dots, a_{jk}^{(m)}))_{j,k=1}^n. \tag{22.4}$$

One can now ask the multivariable version of the same question as above:

"*Which functions applied entrywise to m-tuples of positive matrices preserve positivity?*"

Observe that the coordinate functions $f(x_1, \dots, x_m) := x_l$ work for all $1 \leq l \leq m$. Hence, by the Schur product theorem and the Pólya–Szegő observation (Lemma 16.1, since $\mathbb{P}_n$ is a closed convex cone for all $n \geq 1$), every convergent multi-power series of the form

$$f(\mathbf{x}) := \sum_{\mathbf{n}\geq\mathbf{0}} c_{\mathbf{n}} \mathbf{x}^{\mathbf{n}}, \qquad \text{with } c_{\mathbf{n}} \geq 0 \; \forall \mathbf{n} \geq \mathbf{0} \tag{22.5}$$

preserves positivity in all dimensions (where $\mathbf{x}^{\mathbf{n}} := x_1^{n_1} \cdots x_m^{n_m}$, etc.). Akin to the Schoenberg–Rudin theorem in the one-variable case, it was shown by FitzGerald, Micchelli, and Pinkus in *Linear Algebra Appl.* (1995) that the functions (22.5) are the only such preservers.

One can ask if the same result holds when one restricts the test set to m-tuples of Hankel matrices of rank at most 3, as in the treatment above. While this does turn out to yield the same classification, the proofs get more involved and now require multivariable machinery. For these stronger multivariate results, we refer the reader to the paper "*Moment-sequence transforms*" by Belton, Guillot, Khare, and Putinar in *J. Eur. Math. Soc.*

23. Appendix A: The Boas–Widder theorem on functions with positive differences.

In this Appendix, we reproduce the complete proof of the theorem by Boas and Widder on functions with non-negative forward differences (*Duke Math. J.*, 1940). This result was stated as Theorem 18.10(2), and we again write its statement here for convenience. In it and throughout this Appendix, recall from just before Theorem 18.10 that given an interval $I \subset \mathbb{R}$ and a function $f : I \to \mathbb{R}$, the *kth order forward differences of f with step size $h > 0$* are defined as follows:

$$(\Delta_h^0 f)(x) := f(x), \qquad (\Delta_h^k f)(x) := (\Delta_h^{k-1} f)(x+h) - (\Delta_h^{k-1} f)(x) = \sum_{j=0}^{k} \binom{k}{j} (-1)^{k-j} f(x+jh),$$

whenever $k > 0$ and $x, x+kh \in I$. It is easily seen that these difference operators commute:

$$\Delta_\delta^m(\Delta_\epsilon^n f(x)) = \Delta_\epsilon^n(\Delta_\delta^m f(x)), \qquad \text{whenever } x, x+m\delta+n\epsilon \in I,$$

and so, we will omit parentheses and possibly permute these operators below, without further reference. Now we (re)state the theorem of interest:

Theorem 23.1 (Boas and Widder)**.** *Suppose $k \geq 2$ is an integer, $I \subset \mathbb{R}$ is an open interval, bounded or not, and $f : I \to \mathbb{R}$ is a function that satisfies the following condition:*

$$(\Delta_h^k f)(x) \geq 0 \text{ whenever } h > 0 \text{ and } x, x+kh \in I, \qquad \text{and } f \text{ is continuous on } I. \qquad (H_k)$$

(In other words, f is continuous and has all forward differences of order k non-negative on I.) Then on all of I, the function $f^{(k-2)}$ exists, is continuous and convex, and has non-decreasing left- and right-hand derivatives.

This is a "finite-order" result; for completeness, an order-free result can be found in Bernstein's theorem 39.10 below.

23.1. **Further remarks and results.** Before writing down Boas and Widder's proof of Theorem 23.1, we make several additional observations beyond the result and its proof. The first observation (which was previously mentioned following Theorem 18.10(2)) is that while $f_\pm^{(k-1)}$ is non-decreasing by the above theorem, it is not always true that any other lower-order derivatives $f, \dots, f^{(k-2)}$ are non-decreasing on I. For example, let $0 \leq l \leq k-2$ and consider $f(x) := -x^{l+1}$ on $I \subset \mathbb{R}$; then $f^{(l)}$ is strictly decreasing on I.

Second, it is natural to seek examples of *non-smooth* functions satisfying the differentiability conditions of Theorem 23.1, but no more – in other words, to explore if Theorem 23.1 is indeed "sharp." This is now verified to be true:

Example 23.2. Let $I = (a, b) \subset \mathbb{R}$ be an open interval, where $-\infty \leq a < b \leq \infty$. Consider any function $g : I \to \mathbb{R}$ that is non-decreasing, hence Lebesgue integrable. For any interior point $c \in I$, the function $f_2(x) := \int_c^x g(t)\ dt$ satisfies (H_2):

$$\begin{aligned}
\Delta_h^2 f_2(x) &= \int_c^x g(t)\ dt - 2\int_c^{x+h} g(t)\ dt + \int_c^{x+2h} g(t)\ dt \\
&= \int_{x+h}^{x+2h} g(t)\ dt - \int_x^{x+h} g(t)\ dt \\
&= \int_x^{x+h} (\Delta_h g)(t)\ dt \geq 0.
\end{aligned}$$

However, not every monotone g gives rise to an antiderivative that is differentiable on all of I.

Finally, to see that the condition (H_k) is sharp for all $k > 2$ as well, define f to be the $(k-1)$-fold indefinite integral of g. We claim that f satisfies (H_k). Continuity is obvious; and to study the kth order divided differences of f, first note by the fundamental theorem of calculus that f is $(k-2)$-times differentiable, with $f^{(k-2)}(x) \equiv f_2(x) = \int_c^x g(t)\ dt$. In particular, $\Delta_h^2 f \in C^{k-2}(a, b-kh)$ whenever $a < x < x + kh < b$ as in (H_k).

Now given such x, h, we compute using the Cauchy mean value theorem 18.10(1) for divided differences (and its notation)

$$\Delta_h^k f(x) = \Delta_h^{k-2}(\Delta_h^2 f)(x) = h^{k-2} D_h^{k-2}(\Delta_h^2 f)(x) = \frac{h^{k-2}}{(k-2)!}(\Delta_h^2 f)^{(k-2)}(y),$$

for some $y \in (a, b-2h)$. But this is easily seen to equal

$$= \frac{h^{k-2}}{(k-2)!}(\Delta_h^2 f^{(k-2)})(y) = \frac{h^{k-2}}{(k-2)!}\Delta_h^2 f_2(y),$$

and we just showed that this is non-negative. $\square$

The final observation in this subsection is that there are natural analogues for $k = 0, 1$ of the Boas–Widder theorem (which is stated for $k \geq 2$). For this, we make the natural definition: *for $k < 0$, $f^{(k)}$ will denote the $|k|$-fold antiderivative of f.* Since f is assumed to be continuous, this is just the iterated indefinite Riemann integral starting at an(y) interior point of I. With this notation at hand:

Proposition 23.3. *The Boas–Widder theorem 23.1 also holds for $k = 0, 1$.*

Proof. In both cases, the continuity of $f^{(k-2)}$ is immediate by the fundamental theorem of calculus. Next, suppose $k = 1$ and choose $c \in I$. Now claim that if f is continuous and non-decreasing (i.e., (H_1)), then $f^{(-1)}(x) := \int_c^x f(t)\ dt$ is convex on I. Indeed, given $x_0 < x_1 \in I$, define $x_\lambda := (1-\lambda)x_0 + \lambda x_1$ for $\lambda \in [0, 1]$, and compute

$$\begin{aligned}
&(1-\lambda)f^{(-1)}(x_0) + \lambda f^{(-1)}(x_1) - f^{(-1)}(x_\lambda)\\
= &\ (1-\lambda)\int_c^{x_1} \mathbf{1}(t \leq x_0) f(t)\ dt + \lambda \int_c^{x_1} \mathbf{1}(t \leq x_1) f(t)\ dt - \int_c^{x_1} \mathbf{1}(t \leq x_\lambda) f(t)\ dt\\
= &\ -(1-\lambda)\int_{x_0}^{x_\lambda} f(t)\ dt + \lambda \int_{x_\lambda}^{x_1} f(t)\ dt.
\end{aligned}$$

But since f is non-decreasing, each integral – together with the accompanying sign – is bounded below by the corresponding expression where $f(t)$ is replaced by $f(x_\lambda)$. An easy computation now yields

$$(1-\lambda)f^{(-1)}(x_0) + \lambda f^{(-1)}(x_1) - f^{(-1)}(x_\lambda) \geq f(x_\lambda)\left(\lambda(x_1 - x_\lambda) - (1-\lambda)(x_\lambda - x_0)\right) = 0;$$

therefore, $f^{(-1)}$ is convex, as desired.

This shows the result for $k = 1$. Next, if $k = 0$ then f is continuous and non-negative on I, hence $f^{(-1)}$ is non-decreasing on I. Now the above computation shows that $f^{(-2)}$ is convex; the remaining assertions are obvious. $\square$

23.2. **Proof of the main result.** In this subsection, we reproduce Boas and Widder's proof of Theorem 23.1. We first make a few clarifying remarks about this proof.

(1) As Boas and Widder mention, Theorem 23.1 was shown earlier by T. Popoviciu (*Mathematica*, 1934) via an alternate argument using divided differences involving unequally spaced points. Here we will only explain Boas and Widder's proof.
(2) There is a minor error in the arguments of Boas and Widder, which is resolved by adding one word. See Remark 23.11 and the proof of Lemma 23.13 for more details. (There are a few other minor typos in the writing of Lemmas 23.6 and 23.10 and in some of the proofs; these are corrected without elaboration in the exposition in this Appendix.)
(3) Boas and Widder do not explicitly write out a proof of the convexity of f (in the case $k = 2$). This is addressed below as well – see the paragraph following Proposition 23.12.

Notice that Theorem 23.1 follows for the case of unbounded domain I from that for bounded domains, so *we assume henceforth that*

$$I = (a, b), \qquad \text{with } -\infty < a < b < \infty.$$

We now reproduce a sequence of 14 lemmas shown by Boas and Widder, which culminate in the above theorem. These lemmas are numbered "Lemma 23.1", ..., "Lemma 23.14" and will be referred to only in this Appendix. The rest of the results, equations, and remarks – starting from Theorem 23.1 and ending with Proposition 23.12 – are numbered using the default counter in this text. None of the results in this Appendix are cited elsewhere in the text.

The first of the 14 lemmas by Boas and Widder says that if the kth order "equispaced" forward differences are non-negative, then so are the kth order "possibly non equispaced" differences (the converse is immediate):

Lemma 23.1. *If $f(x)$ satisfies (H_k) in (a, b) for some $k \geq 2$, then for any k positive numbers $\delta_1, \dots, \delta_k > 0$,*

$$\Delta_{\delta_1} \Delta_{\delta_2} \cdots \Delta_{\delta_k} f(x) \geq 0, \qquad \textit{whenever } a < x < x + \delta_1 + \delta_2 + \cdots + \delta_k < b.$$

Proof. The key step is to prove using (H_k) that

$$\Delta_h^{k-1} \Delta_{\delta_1} f(x) \geq 0, \qquad \text{whenever } a < x < x + (k-1)h + \delta_1 < b. \tag{23.4}$$

After this, the lemma is proved using induction on $k \geq 2$. Indeed, (23.4) is precisely the assertion in the base case $k = 2$; and using (23.4) we can show the induction step as follows: for a fixed $\delta_1 \in (0, b - a)$, it follows that $\Delta_{\delta_1} f$ satisfies (H_{k-1}) in the interval $(a, b - \delta_1)$. Therefore,

$$\Delta_{\delta_2} \cdots \Delta_{\delta_k} (\Delta_{\delta_1} f(x)) \geq 0, \qquad \textit{whenever } a < x < x + \delta_1 + \cdots + \delta_k < b.$$

Since the Δ_{δ_j} commute, and since δ_1 was arbitrary, the induction step follows.

Thus, it remains to show (23.4). Let $h > 0$ and $n \in \mathbb{N}$ be such that $a < x < x + h/n + (k-1)h < b$. One can check using an easy telescoping computation that

$$\Delta_h f(x) = \sum_{i=0}^{n-1} \Delta_{h/n} f(x + ih/n);$$

and iterating this procedure, we obtain

$$\Delta_h^{k-1} f(x) = \sum_{i_1=0}^{n-1} \cdots \sum_{i_{k-1}=0}^{n-1} \Delta_{h/n}^{k-1} f(x + [i_1 + \cdots + i_{k-1}]h/n). \tag{23.5}$$

(This works by induction on $k \geq 2$: the previous telescoping identity is the base case for $k = 2$, and for the induction step we evaluate the innermost sum using the base case.)

From the above computations, it further follows that

$$\Delta_{h/n} \Delta_h^{k-1} f(x) = \sum_{i_1=0}^{n-1} \cdots \sum_{i_{k-1}=0}^{n-1} \Delta_{h/n}^{k} f(x + [i_1 + \cdots + i_{k-1}]h/n) \geq 0,$$

where the final inequality uses the assumption (H_k). From this it follows that $\Delta_h^{k-1} f(x) \leq \Delta_h^{k-1} f(x + h/n)$.

Now suppose x is such that $a < x < x + mh/n + (k-1)h < b$. Applying the preceding inequality to $x, x + h/n, \ldots, x + (m-1)h/n$, we obtain

$$\Delta_h^{k-1} f(x) \leq \Delta_h^{k-1} f(x + h/n) \leq \cdots \leq \Delta_h^{k-1} f(x + mh/n). \tag{23.6}$$

We can now prove (23.4). As in it, choose $\delta_1 > 0$, such that $a < x < x + \delta_1 + (k-1)h < b$; and choose sequences m_j, n_j of positive integers, such that $m_j/n_j \to \delta_1/h$ and $x + m_j h/n_j + (k-1)h < b$ for all $j \geq 1$.

Since $f(x)$ is continuous, $f(x + m_j h/n_j)$ converges to $f(x + \delta_1)$, and $\Delta_h^{k-1} f(x + m_j h/n_j)$ to $\Delta_h^{k-1} f(x + \delta_1)$, as $j \to \infty$. Hence, using (23.6) with m_j, n_j in place of m, n respectively, we obtain by taking limits

$$\Delta_h^{k-1} f(x) \leq \Delta_h^{k-1} f(x + \delta_1).$$

But this is equivalent to (23.4), as desired. □

Lemma 23.2. *If $f(x)$ satisfies (H_k) in (a, b) for some $k \geq 2$, then $\Delta_\epsilon^{k-1} f(x)$ and $\Delta_\epsilon^{k-1} f(x-\epsilon)$ are non-decreasing functions of x in $(a, b - (k-1)\epsilon)$ and $(a + \epsilon, b - (k-2)\epsilon)$, respectively.*

Proof. For the first part, suppose $y < z$ are points in $(a, b - (k-1)\epsilon)$, and set

$$\delta_1 := z - y, \qquad \delta_2 = \cdots = \delta_k := \epsilon.$$

Then by Lemma 23.1 – or simply (23.4) – it follows that

$$\Delta_\epsilon^{k-1} f(z) - \Delta_\epsilon^{k-1} f(y) = \Delta_{\delta_1} \Delta_\epsilon^{k-1} f(y) \geq 0,$$

which is what was asserted.

Similarly, for the second part we suppose $y < z$ are points in $(a + \epsilon, b - (k-2)\epsilon)$. Then $y - \epsilon < z - \epsilon$ are points in $(a, b - (k-1)\epsilon)$, so we are done by the first part. (Remark: Boas and Widder repeat the computations of the first part in this second part; but this is not required.) □

We **assume for the next four lemmas** that f satisfies (H_2) in the interval $x \in (a, b)$.

Lemma 23.3. *Suppose f satisfies (H_2) in (a, b) and $x \in (a, b)$. Then $h^{-1} \Delta_h f(x)$ is a non-decreasing function of h in $(a - x, b - x)$.*

Remark 23.7. Notice that $h = 0$ lies in $(a - x, b - x)$, and at this point the expression $h^{-1} \Delta_h f(x)$ is not defined. Hence, the statement of Lemma 23.3 actually says that $h \mapsto h^{-1} \Delta_h f(x)$ is non-decreasing for h in $(0, b - x)$ and separately for h in $(a - x, 0)$. The latter can be reformulated as follows: since $\Delta_{-h} f(x) = -\Delta_h f(x - h)$, Lemma 23.3 asserts that the map $h \mapsto h^{-1} \Delta_h f(x - h)$ is a non-increasing function of h in $(0, x - a)$.

Proof of Lemma 23.3. We first prove the result for $h \in (0, b-x)$. Thus, suppose $0 < \epsilon < \delta < b - x$. By condition (H_2), for all integers $n \geq 2$ we have

$$\begin{aligned} &\Delta^2_{\delta/n} f(x) \geq 0, \quad \Delta^2_{\delta/n} f(x+\delta/n) \geq 0, \quad \dots, \quad \Delta^2_{\delta/n} f(x+(n-2)\delta/n) \geq 0 \\ \implies &\Delta_{\delta/n} f(x) \leq \Delta_{\delta/n} f(x+\delta/n) \leq \cdots \leq \Delta_{\delta/n} f(x+(n-1)\delta/n). \end{aligned}$$

If $0 < m < n$, then the average of the first m terms here cannot exceed the average of all n terms. Therefore,

$$\frac{f(x+m\delta/n) - f(x)}{m\delta/n} \leq \frac{f(x+\delta) - f(x)}{\delta}.$$

Now since $\epsilon \in (0, \delta)$, choose integer sequences $0 < m_j < n_j$, such that $m_j/n_j \to \epsilon/\delta$ as $j \to \infty$. Applying the preceding inequality (with m, n replaced respectively by m_j, n_j) and taking limits, it follows that $\epsilon^{-1} \Delta_\epsilon f(x) \leq \delta^{-1} \Delta_\delta f(x)$, since f is continuous. This proves the first part of the lemma, for positive h.

The proof for negative $h \in (a-x, 0)$ is similar, and is shown using the reformulation of the assertion in Remark 23.7. Given $0 < \epsilon < \delta < x - a$, by condition (H_2) it follows for all integers $0 < m < n$ that

$$\begin{aligned} &\Delta_{\delta/n} f(x-\delta) \leq \Delta_{\delta/n} f(x-(n-1)\delta/n) \leq \cdots \leq \Delta_{\delta/n} f(x-\delta/n) \\ \implies &\frac{f(x) - f(x-\delta)}{\delta} \leq \frac{f(x) - f(x-m\delta/n)}{m\delta/n}, \end{aligned}$$

this time using the last m terms instead of the first. Now work as above: using integer sequences $0 < m_j < n_j$, such that $m_j/n_j \to \epsilon/\delta$, it follows from the continuity of f that $\delta^{-1} \Delta_\delta f(x-\delta) \leq \epsilon^{-1} \Delta_\epsilon f(x-\epsilon)$, as desired. □

We next define the one-sided derivatives of functions.

Definition 23.8. Let f be a real-valued function on (a, b). Define

$$f'_+(x) := \lim_{\delta \to 0^+} \frac{\Delta_\delta f(x)}{\delta}, \qquad f'_-(x) := \lim_{\delta \to 0^-} \frac{\Delta_\delta f(x)}{\delta} = \lim_{\delta \to 0^+} \frac{\Delta_\delta f(x-\delta)}{\delta}.$$

Lemma 23.4. *Suppose f satisfies (H_2) in (a, b). Then f'_+, f'_- exist and are finite and non-decreasing on all of (a, b).*

Proof. That $f'_\pm$ exist on (a, b) follows from Lemma 23.3, though the limits may possibly be infinite. Now fix scalars $\delta, \epsilon, x, y, z$ satisfying

$$0 < \delta < \epsilon \qquad \text{and} \qquad a < z - \epsilon < x - \epsilon < x < x + \epsilon < y + \epsilon < b,$$

which implies that $a < z < x < y < b$. Then we have

$$\frac{\Delta_\epsilon f(z-\epsilon)}{\epsilon} \leq \frac{\Delta_\epsilon f(x-\epsilon)}{\epsilon} \leq \frac{\Delta_\delta f(x-\delta)}{\delta} \leq \frac{\Delta_\delta f(x)}{\delta} \leq \frac{\Delta_\epsilon f(x)}{\epsilon} \leq \frac{\Delta_\epsilon f(y)}{\epsilon},$$

where the five inequalities follow respectively using Lemma 23.2, Remark 23.7, Lemma 23.2, Lemma 23.3, and Lemma 23.2.

Now let $\delta \to 0^+$ keeping ϵ, x, y, z fixed; this yields

$$\frac{\Delta_\epsilon f(z-\epsilon)}{\epsilon} \leq f'_-(x) \leq f'_+(x) \leq \frac{\Delta_\epsilon f(y)}{\epsilon},$$

which implies that $f'_\pm(x)$ are finite on (a, b). In turn, letting $\epsilon \to 0^+$ yields:

$$f'_-(z) \leq f'_-(x) \leq f'_+(x) \leq f'_+(y),$$

which shows that $f'_\pm$ are non-decreasing on (a, b). □

Lemma 23.5. *If f satisfies (H_2) in (a,b) then f approaches a limit in $(-\infty,+\infty]$ as x goes to a^+ and x goes to b^-.*

Proof. Note by Lemma 23.2 that $\Delta_\delta f(x)$ is non-decreasing in $x \in (a, b-\delta)$. Hence, $\lim_{x\to a^+} \Delta_\delta f(x)$ exists and is finite, or equals $-\infty$. (The key point is that it is not $+\infty$.) Therefore, since f is continuous,

$$+\infty > \lim_{x\to a^+} \Delta_\delta f(x) = \lim_{x\to a^+} (f(x+\delta) - f(x)) = f(a+\delta) - f(a^+).$$

It follows that $f(a^+)$ exists and cannot equal $-\infty$.

By the same reasoning, the limit $\lim_{x\to(b-\delta)^-} \Delta_\delta f(x)$ exists and is finite, or equals $+\infty$, in which case

$$-\infty < \lim_{x\to(b-\delta)^-} \Delta_\delta f(x) = f(b^-) - f(b-\delta).$$

It follows that $f(b^-)$ exists and cannot equal $-\infty$. $\square$

Lemma 23.6. *Suppose $f(x)$ satisfies (H_2) in (a,b).*

(1) If $f(a^+) < +\infty$, define $f(a) := f(a^+)$. Then $f'_+(a)$ exists and is finite or $-\infty$.
(2) If $f(b^-) < +\infty$, define $f(b) := f(b^-)$. Then $f'_-(b)$ exists and is finite or $+\infty$.

Proof. First, if $f(a^+)$ or $f(b^-)$ are not $+\infty$ then they are finite by Lemma 23.5. To show (1), by Lemma 23.3, for $h \in (0, b-a)$ the map $h \mapsto h^{-1}\Delta_h f(x)$ is non-decreasing. Therefore, $h \mapsto h^{-1}\Delta_h f(a)$ is the limit of a set of non-decreasing functions in h, so it too is non-decreasing in h. This proves (1).

The second part is proved similarly, using that $h \mapsto h^{-1}\Delta_h f(b-h)$ is a non-increasing function in h. $\square$

Common hypothesis for Lemmas 7–14: f satisfies (H_k) in (a,b), for some $k \geq 3$. (We use this hypothesis below without mention.)

Lemma 23.7. *For any $a < x < b$, the map $h \mapsto h^{-k+1}\Delta_h^{k-1} f(x)$ is a non-decreasing function of h in $(0, (b-x)/(k-1))$.*

Proof. First, note that the given map is indeed well-defined. Now we prove the result by induction on $k \geq 2$; the following argument is similar in spirit to (for instance) computing by induction the derivative of x^{k-1}.

For $k = 2$ the result follows from Lemma 23.3. To show the induction step, given fixed $0 < h < (b-a)/(k-2)$ and $\delta \in (0, b-a)$, it is clear by Lemma 23.1 that if f satisfies (H_k) in (a,b), then we have, respectively:

$$\begin{gathered} \Delta_h^{k-2} f \text{ satisfies } (H_2) \text{ in } (a, b-(k-2)h), \\ \Delta_\delta f \text{ satisfies } (H_{k-1}) \text{ in } (a, b-\delta). \end{gathered} \tag{23.9}$$

In particular, if $0 < \delta < \epsilon < (b-x)/(k-1)$, then we have

$$\frac{\Delta_\epsilon \Delta_\epsilon^{k-2} f(x)}{\epsilon^{k-2}\ \epsilon} \geq \frac{\Delta_\delta \Delta_\epsilon^{k-2} f(x)}{\epsilon^{k-2}\ \delta} = \frac{\Delta_\epsilon^{k-2} \Delta_\delta f(x)}{\epsilon^{k-2}\ \delta} \geq \frac{\Delta_\delta^{k-2} \Delta_\delta f(x)}{\delta^{k-2}\ \delta}.$$

Indeed, the first inequality is by the assertion for $k = 2$, which follows via Lemma 23.3 from the first condition in (23.9); and the second inequality is by the induction hypothesis (i.e., the assertion for $k-1$) applied using the second condition in (23.9).

We saw in the preceding calculation that $\epsilon^{-k+1}\Delta_\epsilon^{k-1} f(x) \geq \delta^{-k+1}\Delta_\delta^{k-1} f(x)$. But this is precisely the induction step. $\square$

Lemma 23.8. *There is a point $c \in [a,b]$, such that $f(x)$ satisfies (H_{k-1}) in (c,b) and $-f(x)$ satisfies (H_{k-1}) in (a,c).*

Proof. Define subsets $A, B \subset (a,b)$ via

$$\begin{aligned} A &:= \{x \in (a,b) : \Delta_\delta^{k-1} f(x) \geq 0 \text{ for all } \delta \in (0, (b-x)/(k-1))\}, \\ B &:= (a,b) \setminus A. \end{aligned}$$

If both A, B are non-empty, and $z \in A, y \in B$, then we claim that $y < z$. Indeed, since $y \notin A$, there exists $0 < \epsilon < (b-y)/(k-1)$, such that $\Delta_\epsilon^{k-1} f(y) < 0$. By Lemma 23.2, if $z' \in (a,y]$, then $\Delta_\epsilon^{k-1} f(z') < 0$, and hence $z' \notin A$. Now conclude that $z > y$.

The above analysis implies the existence of $c \in [a,b]$, such that $(a,c) \subset B \subset (a,c]$ and $(c,b) \subset A \subset [c,b)$. It is also clear that f satisfies (H_{k-1}) in (c,b).

It remains to show that if $a < c$, then $-f$ satisfies (H_{k-1}) in (a,c). Begin by defining a map $\varepsilon : (a,c) \to (0,\infty)$ as follows: for $x \in (a,c)$, there exists $\epsilon \in (0,(c-x)/(k-1))$, such that $\Delta_\epsilon^{k-1} f(x) < 0$. By Lemmas 23.2 and 23.7, this implies that

$$\Delta_\delta^{k-1} f(y) < 0, \qquad \forall a < y \leq x, \ 0 < \delta \leq \epsilon.$$

Now define $\varepsilon : (a,c) \to (0,\infty)$ by setting

$$\varepsilon(x) := \sup\{\epsilon \in (0, \tfrac{c-x}{k-1}) : \Delta_\epsilon^{k-1} f(x) < 0\}.$$

By the reasoning just described, ε is a non-increasing function on (a,c).

With the function ε in hand, we now complete the proof by showing that $-f(x)$ satisfies (H_{k-1}) in (a,c). Let $x \in (a,c)$ and let $h > 0$ be such that $x + (k-1)h < c$. Choose any $y \in (x+(k-1)h, c)$ as well as an integer $n > h/\varepsilon(y)$. It follows that $\Delta_{h/n}^{k-1} f(y) < 0$.

Now recall from Equation (23.5) that

$$\Delta_h^{k-1} f(x) = \sum_{i_1=0}^{n-1} \cdots \sum_{i_{k-1}=0}^{n-1} \Delta_{h/n}^{k-1} f(x + [i_1 + \cdots + i_{k-1}]h/n).$$

But in each summand, the argument $x + [i_1 + \cdots + i_{k-1}]h/n < y$, so by Lemmas 23.2 and 23.7, the previous paragraph implies that each summand is negative. It follows that $\Delta_h^{k-1} f(x) < 0$. This shows that $-f(x)$ satisfies (H_{k-1}) in (a,c), as desired, and concludes the proof. $\square$

Lemma 23.9. *There are points*

$$a = x_0 < x_1 < \cdots < x_p = b, \qquad \text{with } 1 \leq p \leq 2^{k-1},$$

such that in each interval $x_j < x < x_{j+1}$, either $f(x)$ or $-f(x)$ satisfies (H_2).

This follows immediately from Lemma 23.8 by induction on $k \geq 2$.

Lemma 23.10. *The derivatives $f'_\pm$ both exist and are finite on all of (a,b).*

We remark here that $f'_\pm$ are both needed in what follows, yet Boas and Widder completely avoid discussing f'_- in this lemma or its proof (or in the sequel). For completeness, the proof for f'_- is also now described.

Proof. By Lemmas 23.9, 23.4, and 23.6, the functions $f'_\pm$ exist on all of (a,b), and are finite, possibly except at the points $x_1, \dots, x_{p-1}$ in Lemma 23.9. We now show that $f'_\pm$ are finite at each of these points x_j.

First, suppose $f'_+(x_j)$ or $f'_-(x_j)$ equals $+\infty$. Choose $\delta > 0$ small enough, such that

$$x_{j-1} < x_j - (k-2)\delta < x_j < x_j + \delta < x_{j+1}.$$

Now if $f'_+(x_j) = +\infty$, then

$$\begin{aligned}
&\Delta_\delta^{k-1} f'_+(x_j - (k-2)\delta) = -\infty \\
\implies &\lim_{h\to 0^+} \frac{1}{h} \Delta_\delta^{k-1} \Delta_h f(x_j - (k-2)\delta) = -\infty \\
\implies &\Delta_\delta^{k-1} \Delta_h f(x_j - (k-2)\delta) < 0 \qquad \text{for all small positive } h.
\end{aligned}$$

But this contradicts Lemma 23.1. Similarly, if $f'_-(x_j) = +\infty$, then

$$\begin{aligned}
&\Delta_\delta^{k-1} f'_-(x_j - (k-2)\delta) = -\infty \\
\implies &\lim_{h\to 0^+} \frac{1}{h} \Delta_\delta^{k-1} \Delta_h f(x_j - (k-2)\delta - h) = -\infty \\
\implies &\Delta_\delta^{k-1} \Delta_h f(x_j - (k-2)\delta - h) < 0 \qquad \text{for all small positive } h,
\end{aligned}$$

which again contradicts Lemma 23.1.

The other case is if $f'_+(x_j)$ or $f'_-(x_j)$ equals $-\infty$. The first of these subcases is now treated; the subcase $f'_-(x_j) = -\infty$ is similar. Begin as above by choosing $\delta > 0$, such that

$$x_{j-1} < x_j - (k-1)\delta < x_j < x_{j+1}.$$

Now if $f'_+(x_j) = +\infty$, then a similar computation to above yields

$$\begin{aligned}
&\Delta_\delta^{k-1} f'_+(x_j - (k-1)\delta) = -\infty \\
\implies &\lim_{h\to 0^+} \frac{1}{h} \Delta_\delta^{k-1} \Delta_h f(x_j - (k-1)\delta) = -\infty \\
\implies &\Delta_\delta^{k-1} \Delta_h f(x_j - (k-1)\delta) < 0 \qquad \text{for all small positive } h,
\end{aligned}$$

which contradicts Lemma 23.1. □

The above trick of studying $\Delta_\delta^n g(y - p\delta)$ where $p = k-1$ or $k-2$ (and $n = k-1$, $g = f'_\pm$ so that we deal with the kth order divided differences/derivatives of f) is a powerful one. Boas and Widder now use the same trick to further study the derivative of f, and show its existence, finiteness, and continuity in Lemmas 23.11 and 23.13.

Lemma 23.11. *f' exists and is finite on (a, b).*

Proof. We fix $x \in (a, b)$, and work with $\delta > 0$ small, such that $a < a + k\delta < x < b - 2\delta < b$. Let $p \in \{0, 1, \dots, k\}$; then

$$0 \le \frac{1}{\delta} \Delta_\delta^k f(x - p\delta) = \frac{1}{\delta} \sum_{i=0}^{k} \binom{k}{i} (-1)^{k-i} f(x + (i-p)\delta).$$

Subtract from this the identity $0 = \delta^{-1} f(x)(1-1)^k = \delta^{-1} f(x) \sum_{i=0}^k \binom{k}{i} (-1)^{k-i}$, so that the $i = p$ term cancels, and multiply and divide the remaining terms by $(i-p)$ to obtain

$$0 \le \frac{1}{\delta} \Delta_\delta^k f(x - p\delta) = \sum_{\substack{i=0, \\ i\neq p}}^{i=k} \binom{k}{i} (-1)^{k-i} \frac{f(x + (i-p)\delta) - f(x)}{(i-p)\delta} (i-p).$$

Letting $\delta \to 0^+$, it follows that

$$A_p f'_-(x) + B_p f'_+(x) \geq 0, \qquad \text{where } A_p := \sum_{i=0}^{p-1} \binom{k}{i} (-1)^{k-i}(i-p), \qquad B_p := \sum_{i=p+1}^{k} \binom{k}{i} (-1)^{k-i}(i-p); \tag{23.10}$$

note here that

$$A_p + B_p = \sum_{i=0}^{k} \binom{k}{i} (-1)^{k-i}(i-p) = k \sum_{i=1}^{k-1} \binom{k-1}{i-1} (-1)^{k-i} - p \sum_{i=0}^{k} \binom{k}{i} (-1)^{k-i} = 0.$$

Now specialize p to be $k-1$ and $k-2$. In the former case $B_p = 1$, so $A_p = -1$, and by (23.10) we obtain $f'_+(x) \geq f'_-(x)$. In the latter case $p = k-2$ (with $k \geq 3$), we have $B_p = 2-k < 0$. Thus, $A_p = k-2 > 0$, and by (23.10) we obtain $f'_-(x) \geq f'_+(x)$. Therefore, $f'(x)$ exists and by Lemma 23.10 it is finite. ☐

Lemma 23.12. *If $a < x < x+(k-1)h < b$, then $\Delta_h^{k-1} f'(x) \geq 0$.*

Proof. $\Delta_h^{k-1} f'(x) = \lim_{\delta \to 0^+} \dfrac{\Delta_\delta \Delta_h^{k-1} f(x)}{\delta}$, and this is non-negative by Lemma 23.1. ☐

Lemma 23.13. *f' is continuous on (a,b).*

Remark 23.11. We record here a minor typo in the Boas–Widder paper [55]. Namely, the authors begin the proof of Lemma 23.13 by claiming that f' is monotonic. However, this is not true as stated: for any $k \geq 3$, the function $f(x) = x^3$ satisfies (H_k) on $I = (-1,1)$ but f' is not monotone on I. The first paragraph of the following proof addresses this issue, using that f' is *piecewise* monotone on (a,b).

Proof of Lemma 23.13. By Lemmas 23.9 and 23.4, there are finitely many points x_j, $0 \leq j \leq p \leq 2^{k-1}$, such that on each (x_j, x_{j+1}), $f'_\pm = f'$ is monotone (where this last equality follows from Lemma 23.11). Thus, f' is piecewise monotone on (a,b).

Now define the limits

$$f'(x^\pm) := \lim_{h \to 0^+} f'(x \pm h), \qquad x \in (a,b).$$

It is clear that $f'(x^\pm)$ exists on (a,b), including at each $x_j \neq a,b$. Note that $f'(x_j^\pm) \in [-\infty, +\infty]$, while $f'(x^\pm) \in \mathbb{R}$ for all other points $x \neq x_j$. First, claim that $f'(x^+) = f'(x^-)$ – where this common limit is possibly infinite – and then that $f'(x^+) = f'(x)$, which will rule out the infinitude using Lemma 23.11, and complete the proof.

For each of the two steps, we proceed as in the proof of Lemma 23.11. Begin by fixing $x \in (a,b)$, and let $\delta > 0$ be such that $a < x - k\delta < x < x + 2\delta < b$. Let $p \in \{0, 1, \dots, k\}$; then by Lemma 23.12,

$$0 \leq \Delta_\delta^{k-1} f'(x - (p - \tfrac{1}{2})\delta) = \sum_{i=0}^{k-1} \binom{k-1}{i} (-1)^{k-1-i} f'(x + (i - p + \tfrac{1}{2})\delta).$$

Let $\delta \to 0^+$; then,

$$A_p f'(x^-) - A_p f'(x^+) \geq 0, \quad \text{where } A_p := \sum_{i=0}^{p-1} \binom{k-1}{i} (-1)^{k-1-i} = -\sum_{i=p}^{k-1} \binom{k-1}{i} (-1)^{k-1-i}.$$

Now specialize p to be $k-1$ and $k-2$. In the former case $A_p = -1$, hence $f'(x^-) \leq f'(x^+)$; whereas if $p = k-2$, then $A_p = k-2 > 0$, hence $f'(x^-) \geq f'(x^+)$. These inequalities and the trichotomy of the extended real line $[-\infty, +\infty]$ imply that $f'(x^-) = f'(x^+)$.

Using the same $\delta \in ((x-a)/k, (b-x)/2)$ and $p \in \{0, 1, \dots, k\}$, Lemma 23.12 also implies

$$0 \leq \Delta_\delta^{k-1} f'(x - p\delta).$$

Taking $\delta \to 0^+$ and using that $f'(x^-) = f'(x^+)$ yields

$$B_p f'(x) - B_p f'(x^+) \geq 0, \quad \text{where } B_p := \binom{k-1}{p}(-1)^{k-1-p} = -\sum_{\substack{i=0,\\ i \neq p}}^{k-1} \binom{k-1}{i}(-1)^{k-1-i}.$$

Now specialize p to be $k-1$ and $k-2$. In the former case $B_p = 1$, hence $f'(x) \geq f'(x^+)$; whereas if $p = k-2$, then $B_p = 1-k < 0$, hence $f'(x) \leq f'(x^+)$. These inequalities imply that $f'(x^+) = f'(x^-)$ equals $f(x)$, and in particular is finite, for all $x \in (a,b)$. $\square$

The final lemma simply combines the preceding two:

Lemma 23.14. *f' satisfies the condition (H_{k-1}) in (a,b).*

Proof. This follows immediately from Lemmas 23.12 and 23.13. $\square$

Having shown the 14 lemmas above, we conclude with:

Proof of the Boas–Widder Theorem 23.1. The proof is by induction on $k \geq 2$. The induction step is clear: use Lemma 23.14. We now show the base case of $k = 2$. By Lemma 23.4, the functions $f'_\pm$ exist and are non-decreasing on (a,b). Moreover, f is continuous by assumption. To prove its convexity, we make use of the following basic result from one-variable calculus:

Proposition 23.12. *Let $f : [p,q] \to \mathbb{R}$ be a continuous function whose right-hand derivative f'_+ exists on $[p,q)$ and is Lebesgue integrable. Then,*

$$f(y) = f(p) + \int_p^y f'_+(t)\ dt, \qquad \forall y \in [p,q].$$

Proposition 23.12 applies to our function f satisfying (H_2), since f'_+ is non-decreasing by Lemma 23.4, and hence Lebesgue integrable. Therefore, $f(y) - f(x) = \int_x^y f'_+(t)\ dt$ for $a < x < y < b$. Now repeat the proof of Proposition 23.3 to show that f is convex on (a,b). This completes the base case of $k = 2$ and concludes the proof. $\square$

24. Appendix B: Functions acting outside forbidden diagonal blocks. Dimension-free non-absolutely-monotonic preservers.

In this section, we explore a variant of the question of classifying the dimension-free preservers. Recall that Schoenberg's original motivation in proving his result was to classify the entrywise positivity preservers $f[-]$ on correlation/Gram matrices – with or without rank constraints – since these are the matrices that arise as distance matrices on Euclidean spheres (after applying $\cos(\cdot)$ entrywise). In a sense, this is equivalent to applying $f/f(1)$ to the off-diagonal entries of correlation matrices and preserving positivity.

In a similar vein, and motivated by modern applications via high-dimensional covariance estimation, Guillot and Rajaratnam in *Trans. Amer. Math. Soc.* (2015) classified entrywise maps that operate only on off-diagonal entries, and preserve positivity in all dimensions.

Theorem 24.1 (Guillot–Rajaratnam)**.** *Let $0 < \rho \leq \infty$ and $f : (-\rho, \rho) \to \mathbb{R}$. Given a square matrix $A \in \mathbb{P}_n((-\rho, \rho))$, define $f^*[A] \in \mathbb{R}^{n \times n}$ to be the matrix with (j,k)-entry $f(a_{jk})$ if $j \neq k$, and a_{jj} otherwise. Then the following are equivalent:*

(1) $f^*[-]$ preserves positivity on $\mathbb{P}_n((-\rho, \rho))$ for all $n \geq 1$.

(2) There exist scalars $c_k \geq 0$, such that $f(x) = \sum_{k \geq 0} c_k x^k$ and $|f(x)| \leq |x|$ on all of $(-\rho, \rho)$. (Thus, if $\rho = \infty$, then $f(x) \equiv cx$ on $\mathbb{R}$, for some $c \in [0, 1]$.)

Once again, the robust characterization of absolute monotonicity emerges out of this variant of entrywise operations.

The main result of this section provides – in a closely related setting – an example of a dimension-free preserver that is *not* absolutely monotonic. To elaborate: Theorem 24.1 was recently strengthened by Vishwakarma in *Trans. Amer. Math. Soc.*, where he introduced the more general model in which a different function $g(x)$ acts on the diagonal entries. Even more generally, Vishwakarma allowed $g[-]$ to act on prescribed principal submatrices/diagonal blocks and $f[-]$ to act on the remaining entries. To explain his results, we adopt the following notation throughout this section:

Definition 24.2. Fix $0 < \rho \leq \infty$, $I = (-\rho, \rho)$, and $f, g : I \to \mathbb{R}$. Also fix families of subsets $T_n \subset (2^{[n]}, \subset)$ for each $n \geq 1$, such that all elements in a fixed family T_n are pairwise incomparable. Now given $n \geq 1$ and a matrix $A \in I^{n \times n}$, define $(g, f)_{T_n}[A] \in \mathbb{R}^{n \times n}$ to be the matrix with (j,k)-entry $g(a_{jk})$ if there is some $E \in T_n$ containing j, k (here, j may equal k), and $f(a_{jk})$ otherwise.

Adopting this notation, Vishwakarma classifies the pairs (g, f) which preserve positivity according to a given sequence $\{T_n : n \geq 1\}$. Notice that if $T_n = \{[n]\}$ for $n > n_0$ and T_n is empty for $n \leq n_0$, this implies from Section 16.1 that $g(x)$ is absolutely monotonic as in Schoenberg–Rudin's results; and that $f[-]$ preserves positivity on $\mathbb{P}_{n_0}((-\rho, \rho))$. Such functions f do not admit a known characterization for $n_0 \geq 3$; and the following result will also not consider them. Thus, below we require $T_n \neq \{[n]\}$ for infinitely many $n \geq 1$.

Theorem 24.3 (Vishwakarma)**.** *Notation as in Definition 24.2. Suppose $\{T_n\}$ is such that $T_n \neq \{[n]\}$ for infinitely many $n \geq 1$. Then $(g, f)_{T_n}[-]$ preserves positivity on $\mathbb{P}_n(I)$ for all $n \geq 1$, if and only if exactly one of the following occurs:*

(1) If T_n is the empty collection, i.e., $(g, f)_{T_n}[-] = f[-]$ for all $n \geq 1$, then $f(x) = \sum_{k \geq 0} c_k x^k$ on I, where $c_k \geq 0$ for all $k \geq 0$.

(2) If some T_n contains two non-disjoint subsets of $[n]$, then $g(x) = f(x)$, and $f(x)$ is a power series as in the preceding subcase.

(3) If $T_n \subset \{\{1\}, \dots, \{n\}\}$ for all $n \geq 2$, and some T_n is non-empty, then f is as in (1), and $0 \leq f \leq g$ on $[0, \rho)$.

(4) If $T_2 = \{\{1, 2\}\}$ and $T_n \subset \{\{1\}, \dots, \{n\}\}$ for all $n \geq 3$, then f is as in (1), $g(x)$ is non-negative, non-decreasing, and multiplicatively mid-convex on $[0, \rho)$, and $|g(x)| \leq g(|x|)$ for all x. If, moreover, some T_n is non-empty for $n \geq 3$, then $0 \leq f \leq g$ on $[0, \rho)$.

(5) Otherwise $T_n \not\subseteq \{\{1\}, \dots, \{n\}\}$ for some $n \geq 3$; and T_n is a partition of some subset of $[n]$ for each $n \geq 1$. In this case, with the additional assumption that $g(x) = \alpha x^k$ for some $\alpha \geq 0$ and $k \in \mathbb{Z}^{\geq 0}$:

(a) If for all $n \geq 3$ we have $T_n = \{[n]\}$ or $\{\{1\}, \dots, \{n\}\}$, then f is as in (1) and $0 \leq f \leq g$ on $[0, \rho)$.

(b) If T_n is not a partition of $[n]$ for some $n \geq 3$, then $f(x) = cg(x)$ for some $c \in [0, 1]$.

(c) If neither (a) nor (b) holds, then $f(x) = cg(x)$ for some $c \in [-1/(K-1), 1]$, where

$$K := \max_{n \geq 1} |T_n| \in [2, +\infty].$$

In fact, the assertions in the above cases are equivalent to the weaker assertion (than above) that $(g, f)_{T_n}[-]$ preserves positivity on the rank ≤ 3 matrices in $\bigcup_{n \geq 1} \mathbb{P}_n(I)$.

We refer the reader to Vishwakarma's work for similar results with the domain I replaced by $(0, \rho)$, $[0, \rho)$, or even the complex disk $D(0, \rho)$. As mentioned above, one interesting feature here is that in the final assertion (5)(c), we find the first example of a function that is not absolutely monotonic, yet is a dimension-free preserver, in this setting.

To prove Theorem 24.3, we require two well-known preliminaries, and a couple of additional results, shown below:

Proposition 24.4. *Given a Hermitian matrix $A_{n \times n}$, denote its largest and smallest eigenvalues by $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$, respectively.*

(1) (Rayleigh–Ritz theorem.) If $A \in \mathbb{C}^{n \times n}$ is Hermitian, then the ratio $v^* A v / v^* v$, as v runs over $\mathbb{C}^n \setminus \{0\}$, attains its maximum and minimum values, which equal $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$, respectively.

(2) (Weyl's inequality, special case.) If $A, B \in \mathbb{C}^{n \times n}$ are Hermitian, then

$$\lambda_{\min}(A) + \lambda_{\min}(B) \leq \lambda_{\min}(A + B) \leq \lambda_{\min}(A) + \lambda_{\max}(B). \tag{24.5}$$

The second assertion holds more generally; we do not state/prove/require it below.

Proof. For the first part, it suffices to show the minimum bound, since $\lambda_{\max}(A) = -\lambda_{\min}(-A)$. (That the bound is attained follows from the compactness of the unit complex sphere.) The matrix $A - \lambda_{\min}(A) \operatorname{Id}_{n \times n}$ is Hermitian with smallest eigenvalue zero, hence, is positive semidefinite. Thus, we compute for non-zero $v \in \mathbb{C}^n$:

$$0 \leq \frac{v^*(A - \lambda_{\min}(A) \operatorname{Id}_{n \times n}) v}{v^* v} = \frac{v^* A v}{v^* v} - \lambda_{\min}(A).$$

This shows the first assertion. For the second, let $v \in \ker(A - \lambda_{\min}(A) \operatorname{Id}_{n \times n})$ be non-zero. Applying the previous part twice,

$$\lambda_{\min}(A + B) \leq \frac{v^*(A + B) v}{v^* v} = \frac{v^* A v}{v^* v} + \frac{v^* B v}{v^* v} \leq \lambda_{\min}(A) + \lambda_{\max}(B).$$

Similarly, if v is a non-zero eigenvector for $A+B$ with eigenvalue $\lambda_{\min}(A+B)$, then by the previous part applied twice,

$$\lambda_{\min}(A+B) = \frac{v^*(A+B)v}{v^*v} = \frac{v^*Av}{v^*v} + \frac{v^*Bv}{v^*v} \geq \lambda_{\min}(A) + \lambda_{\min}(B). \qquad \square$$

We also require the following special case of the main result:

Lemma 24.6. *Let $0 < \rho \leq \infty$, $I = (-\rho, \rho)$, and $f : I \to \mathbb{R}$. Let $g(x) = \alpha x^k$ for $\alpha \geq 0$ and $k \in \mathbb{Z}^{\geq 0}$. Finally, let $T_3 = \{\{1,2\}\}$ and $c_0 = 0$. The following are equivalent:*

(1) $(g, f)_{T_3}[A] \in \mathbb{P}_3$ for all matrices $A \in \mathbb{P}_3(I)$.
(2) $(g, f)_{T_3}[A] \in \mathbb{P}_3$ for all rank-1 matrices $A \in \mathbb{P}_3(I)$.
(3) $f(x) \equiv cg(x)$ on I for some $c \in [c_0, 1]$.

The same equivalence holds if $T_3 = \{\{1,2\},\{3\}\}$ and $c_0 = -1$.

Proof. First suppose $T_3 = \{\{1,2\}\}$ and $c_0 = 0$. Clearly (1) $\implies$ (2). Now suppose (2) holds. If $f \equiv 0$ or $g \equiv 0$, then the result is immediate, so suppose $f, g \not\equiv 0$ (hence, $\alpha > 0$). Now given $z, w \in (-\rho, \rho)$ such that

$$0 \leq |z| \leq w, \quad 0 < w < \rho,$$

define

$$A(w,z) := \begin{pmatrix} z^2/w & z & z \\ z & w & w \\ z & w & w \end{pmatrix} = \frac{1}{w} uu^T, \quad \text{where} \quad u = (z, w, w)^T. \tag{24.7}$$

By choice of w, z, we have $A(w,z) \in \mathbb{P}_3(I)$, so $\det(g, f)_{T_3}[A(w,z)] \geq 0$.

There are now two cases. First, if $k = 0$, then $g(x) \equiv \alpha > 0$. Let $w > 0$ and expand the above determinant along the third row to compute

$$0 \leq \det(g, f)_{T_3}[A(w,z)] = -\alpha (f(z) - f(w))^2.$$

Using an increasing sequence $0 < w_n \to \rho^-$, this shows that f is constant on I, and by considering $(g, f)_{T_3}[0_{3\times 3}]$, we have $f(x) \equiv c\alpha$ for some $c \in [0,1]$.

The other case is if $k > 0$, so that $g(0) = 0$. Now $f(0) = 0$ by considering $(g, f)_{T_3}[0_{3\times 3}]$. Again expand the above determinant along the third row, to obtain

$$0 \leq \det(g, f)_{T_3}[A(w,z)] = -\frac{\alpha}{w^k} \left(w^k f(z) - z^k f(w) \right)^2. \tag{24.8}$$

Thus, we have $f(z)/z^k = f(w)/w^k$ whenever $0 < |z| \leq w < \rho$. By using an increasing sequence $0 < w_n \uparrow \rho^-$, this shows $f(z)/z^k$ is constant on $I \setminus \{0\}$, say $c \in \mathbb{R}$. By considering $A(w,w) = w\mathbf{1}_{3\times 3}$ for $w > 0$, it is not hard to see that $c \in [0,1]$, which proves (3).

Finally, if (3) holds, then $(g, f)_{T_3}[A]$ is the sum of $cg[A]$ and $(1-c)g[B]$ (padded by a zero row and column at the end), where B is the leading principal 2×2 submatrix of A. This shows (1) by the Schur product theorem.

The proof is similar if $T_3 = \{\{1,2\},\{3\}\}$ and $c_0 = -1$. Clearly, (1) $\implies$ (2); similarly, the proof of (2) $\implies$ (3) is unchanged (including the computation (24.8)) until the very last steps for both $k = 0$ and $k > 0$, at which points we can only conclude $c \in [-1,1]$. Finally, we assume (3) holds and show (1). The point is that for any scalar $c \in [-1,1]$ and any matrix $A \in \mathbb{P}_3(I)$, the principal minors of $(g, cg)_{T_3}[A]$ equal those of $(g, |c|g)_{T_3}[A]$, so that we may work with $|c| \in [0,1]$ instead of $c \in [-1,1]$. Now one shows (1) similarly as the previous case. $\square$

A final preliminary result – the second part easily follows from the first, and in turn strengthens Lemma 24.6.

Proposition 24.9. *Suppose for an integer $n \geq 3$ that $T_n \subset 2^{[n]}$ is a partition of $[n]$ into $k \geq 2$ non-empty subsets.*

(1) Let $g(0) = 1$ and $f(0) = c$. Then $(g, f)_{T_n}[0_{n\times n}]$ is positive semidefinite if and only if $c \in [-1/(k-1), 1]$.

(2) Suppose $0 < \rho \leq \infty$, $I = (-\rho, \rho)$, and $f : I \to \mathbb{R}$. Also suppose $g : I \to \mathbb{R}$ is multiplicative and preserves positivity on $\mathbb{P}_n(I)$. If $T_n \neq \{\{1\}, \dots, \{n\}\}$, then the following are equivalent:
   (a) $(g, f)_{T_n}[-]$ preserves positivity on $\mathbb{P}_n(I)$.
   (b) $(g, f)_{T_n}[-]$ preserves positivity on the rank-1 matrices in $\mathbb{P}_n(I)$.
   (c) $f(x) \equiv cg(x)$ on I, for some $c \in [-1/(k-1), 1]$.

The non-zero functions in part (2) include the powers $x^k, k \in \mathbb{Z}^{\geq 0}$ by the Schur product theorem; but also – as studied by Hiai in *Linear Algebra Appl.* (2009) – the "powers"

$$\phi_\alpha(x) := |x|^\alpha, \qquad \psi_\alpha(x) := \operatorname{sgn}(x)|x|^\alpha, \qquad \alpha \geq n-2.$$

Proof. Let $T_n = \{J_1, \dots, J_k\}$ with $\sqcup_j J_j = [n]$.

(1) Choose elements $j_1, \dots, j_k$ with $j_i \in J_i$. By possibly relabeling the rows and columns, we may assume without loss of generality that $1 \leq j_1 < \cdots < j_k \leq n$. Now if $(g, f)_{T_n}[0_{n\times n}] \in \mathbb{P}_n$, then by considering the principal $k \times k$ submatrix corresponding to the indices $\{j_1, \dots, j_k\}$, we obtain

$$C := c\mathbf{1}_{k\times k} + (1-c)\operatorname{Id}_{k\times k} \in \mathbb{P}_k. \tag{24.10}$$

Since this matrix has eigenvalues $(1-c)$ and $1 + (k-1)c$, we get $c \in [-1/(k-1), 1]$, as desired.

For the converse, define the "decompression" of C, given by

$$\widetilde{C} := c\mathbf{1}_{n\times n} + (1-c)\sum_{j=1}^k \mathbf{1}_{J_j \times J_j} = (g, f)_{T_n}[0_{n\times n}] \in \mathbb{C}^{n\times n}. \tag{24.11}$$

We now show that if $c \in [-1/(k-1), 1]$, then $\widetilde{C} \in \mathbb{P}_n$. Indeed, given a vector $u \in \mathbb{C}^n$, define $u_{T_n} \in \mathbb{C}^k$ to have jth coordinate $\sum_{i \in J_j} u_i$. Then,

$$u^*\widetilde{C}u = u_{T_n}^* C u_{T_n} \geq 0, \qquad \forall u \in \mathbb{C}^n,$$

because the matrix C as in (24.10) is positive semidefinite as above.

(2) If $g \equiv 0$, then the result is easy to prove, so we suppose henceforth that $g \not\equiv 0$. Clearly, (a) implies (b). Next if (b) holds, then one can restrict to a suitable 3×3 submatrix – without loss of generality indexed by $1, 2, 3$, such that $T_n \cap \{1, 2, 3\} = \{\{1, 2\}, \{3\}\}$ by a slight abuse of notation. Hence, $f(x) \equiv cg(x)$ on I for some $c \in [-1, 1]$, by Lemma 24.6. Now if $g(x_0) \neq 0$, then $(g, f)_{T_n}[x_0\mathbf{1}_{n\times n}]$ has as a principal submatrix, $g(x_0)C$, where C is as in (24.10). Hence, $c \in [-1/(k-1), 1]$ by the previous part, proving (c). Finally, given any matrix $A \in \mathbb{P}_n(I)$, we have

$$(g, cg)_{T_n}[A] = g[A] \circ \widetilde{C},$$

where $\widetilde{C}$ is as in (24.11). Now if (c) holds, then $\widetilde{C} \in \mathbb{P}_n$ by the previous part, and this shows (a) by the assumptions on g, f as well as the Schur product theorem. $\square$

With these results in hand, we are ready to proceed.

Proof of Theorem 24.3. Clearly if $(g,f)_{T_n}[-]$ preserves positivity on $\mathbb{P}_n(I)$, then it does so on the rank ≤ 3 matrices in $\mathbb{P}_n(I)$. Thus, we will prove that this latter assertion implies the conclusions on (g,f) in the various cases; and that these conclusions imply in turn that $(g,f)_{T_n}[-]$ preserves positivity on $\mathbb{P}_n(I)$. This is done in each of the subcases (which place constraints on the family T_n). First if (1) all T_n are empty sets, then the result follows from the stronger Schoenberg–Rudin Theorem 16.3 (which holds over $(-\rho,\rho)$ instead of $\mathbb{R}$, as remarked in Section 22.3).

Next, suppose from (2) some T_n contains subsets $I_1, I_2 \subset [n]$ that are not disjoint. Clearly, if $g = f$ and f is as in (1), then $(g,f)_{T_n}[-] = f[-]$ preserves positivity by the Schur product theorem. Conversely, if $(g,f)_{T_n}[-]$ preserves positivity even on the rank-1 matrices in $\mathbb{P}_n((-\rho,\rho))$ for all $n \geq 3$, then there exist integers $n \geq 3$ and $a,b,c \in [n]$, such that

$$a,b \in I_1, \quad c \notin I_1, \qquad b,c \in I_2, \quad a \notin I_2.$$

By relabeling indices if needed, we will assume without loss of generality that $a=1, b=2$, and $c=3$. Now let $x \in (-\rho,\rho)$ and define $A := \begin{pmatrix} |x| & x & x \\ x & |x| & |x| \\ x & |x| & |x| \end{pmatrix} \oplus 0_{(n-3)\times(n-3)} \in \mathbb{P}_n(I)$. If B denotes the leading principal 3×3 submatrix of $(g,f)_{T_n}[A]$, then

$$0 \leq \det B = \det \begin{pmatrix} g(|x|) & g(x) & f(x) \\ g(x) & g(|x|) & g(|x|) \\ f(x) & g(|x|) & g(|x|) \end{pmatrix} = -g(|x|)(f(x)-g(x))^2.$$

If $g(|x|) = 0$, then by considering the 2×2 submatrices of B, we see that $f(x) = g(x) = 0$. If $g(|x|) \neq 0$, then it is positive, so we obtain $f(x) = g(x)$. This implies $f \equiv g$ on $(-\rho,\rho)$. Hence, $(g,f)_{T_n}[-] = f[-]$, and we reduced to case (1). This proves the equivalence for case (2).

Next suppose (3) holds. First assume f is as in (1) and $0 \leq f \leq g$ on $[0,\rho)$. If $A \in \mathbb{P}_n((-\rho,\rho))$, then $(g,f)_{T_n}[A]$ is the sum of $f[A]$ and a diagonal matrix with non-negative entries. Hence, $(g,f)_{T_n}[A]$ is positive semidefinite by the Schur product theorem. The converse has two subcases. Let $s_n := \# \cup_{E\in T_n} E$, so $0 \leq s_n \leq n$, and hence, either $n - s_n$ or s_n is an unbounded sequence. If the former, then by restricting to the corresponding principal submatrices (padded by zeros), we are done by case (1) – considering the 2×2 matrix $\begin{pmatrix} g(x) & f(x) \\ f(x) & g(x) \end{pmatrix}$ or $\begin{pmatrix} g(x) & f(x) \\ f(x) & f(x) \end{pmatrix}$ for $x \in [0,\rho)$, we obtain $f(x) \leq g(x)$, as desired.

Thus, we henceforth assume the latter holds, i.e., s_n is unbounded; restricting to these principal submatrices, we may assume without loss of generality that $T_n = \{\{1\},\dots,\{n\}\}$ for all $n \geq 1$. We claim that $f[-]$ preserves positivity on rank ≤ 3 matrices in $\mathbb{P}_n(I)$ for all n. This would finish the proof in case (3), since now f is as in (1), and as above, this implies $0 \leq f(x) \leq g(x)$ for $x \in [0,\rho)$.

To prove the claim, let $A \in \mathbb{P}_n((-\rho,\rho))$, and let D_A be the diagonal matrix with (j,j)-entry $g(a_{jj}) - f(a_{jj})$. If $\mathbf{1}_{m\times m}$ denotes the all-ones $m\times m$ matrix, then $\mathbf{1}_{m\times m} \otimes A = \begin{pmatrix} A & \cdots & A \\ \vdots & \ddots & \vdots \\ A & \cdots & A \end{pmatrix}$, a matrix in $\mathbb{P}_{mn}(I)$. Also note that if A has rank ≤ 3, then by (3.13), so does $\mathbf{1}_{m\times m} \otimes A$. Now applying $(g,f)_{T_{mn}}[-]$ yields

$$(g,f)_{T_{mn}}[\mathbf{1}_{m\times m} \otimes A] = \mathbf{1}_{m\times m} \otimes f[A] + \mathrm{Id}_{m\times m} \otimes D_A \geq 0.$$

Hence, by (24.5),

$$\begin{aligned} 0 \leq \lambda_{\min}((g,f)_{T_{mn}}[\mathbf{1}_{m\times m}\otimes A]) &\leq \lambda_{\min}(\mathbf{1}_{m\times m}\otimes f[A]) + \lambda_{\max}(\mathrm{Id}_{m\times m}\otimes D_A)\\ &= m\lambda_{\min}(f[A]) + \max_{1\leq j\leq n}\{g(a_{jj}) - f(a_{jj})\}, \end{aligned}$$

where the equality holds because of (3.13) and since the eigenvalues of $\mathbf{1}_{m\times m}$ are $0, m$. From this it follows that $\lambda_{\min}(f[A]) \geq -\max_j(g(a_{jj}) - f(a_{jj}))/m$ for all $m \geq 1$. This shows $f[A]$ is positive semidefinite, and concludes the proof in case (3).

If (4) holds, the proof in case (3) shows f is as in (1); and using $(g,f)_{T_2}[-] = g[-]$ via an argument similar to Theorem 12.7 shows the desired constraints on g. (This is left to the reader to work out.) The converse is shown using (variations of) the same proofs.

It remains to prove the equivalence in case (5); here we are also given that $g(x) = \alpha x^k$ for $\alpha, k \geq 0$ (and k an integer). If $\alpha = 0$, then the result is easy, so we suppose henceforth without loss of generality that $\alpha = 1$. In subcase (a), since $T_n = \{\{1\},\dots,\{n\}\}$ for infinitely many n by assumption, we can repeat the proof for case (3) to show that any preserver-pair (g,f) must satisfy $0 \leq f \leq g$ on $[0,\rho)$ and f is as in (1). Conversely, given such (g,f), if $T_n = \{[n]\}$, then $(g,f)_{T_n}[-] = g[-]$, which preserves positivity by the Schur product theorem. Otherwise, for $A \in \mathbb{P}_n(I)$, we compute

$$(g,f)_{T_n}[A] = f[A] + \mathrm{diag}(g(a_{jj}) - f(a_{jj}))_{j=1}^n,$$

and both matrices are positive, hence so is $(g,f)_{T_n}[A]$, as desired.

Next for (b), we fix $n_1 \geq 3$ such that $T_{n_1} \not\subseteq \{\{1\},\dots,\{n_1\}\}$; also fix $n_0 \geq 3$ such that T_{n_0} is not a partition of $[n_0]$. If $f(x) = cg(x)$ for $c \in [0,1]$, then $(g,f)_{T_{n_0}}[A]$ is the sum of $cA^{\circ k}$ and matrices of the form $(1-c)B^{\circ k}$, where B is a principal submatrix of $A \in \mathbb{P}_{n_0}$, hence positive semidefinite. It follows by the Schur product theorem that $(g,f)_{T_{n_0}}[-]$ preserves positivity. Conversely, suppose $(g,f)_{T_n}[-]$ preserves positivity for all $n \geq 1$, on rank ≤ 3 matrices in $\mathbb{P}_n(I)$. At $n = n_1$, we can find three indices – labeled $1, 2$, and 3 without loss of generality – such that for all $A \in \mathbb{P}_{n_1}(I)$, the leading 3×3 submatrix of $(g,f)_{T_{n_1}}[A]$ equals $(g,f)_{\{\{1,2\}\}}[A_{[3]\times[3]}]$ or $(g,f)_{\{\{1,2\},\{3\}\}}[A_{[3]\times[3]}]$. Now using rank-1 matrices via Lemma 24.6 shows $f(x) = cg(x)$ for $c \in [-1,1]$. Finally, considering matrices in $\mathbb{P}_{n_0}(I)$ yields $c \geq 0$, as desired.

The remaining subcase is (5)(c), in which case every T_n is a partition of $[n]$. Also note by the hypotheses that $K > 1$; and there exists $n_1 \geq 3$ and three indices – labeled $1, 2$, and 3 without loss of generality – such that for all $A \in \mathbb{P}_{n_1}(I)$, the leading 3×3 submatrix of $(g,f)_{T_{n_1}}[A]$ equals $(g,f)_{\{\{1,2\},\{3\}\}}[A_{[3]\times[3]}]$. Now using rank-1 matrices via Lemma 24.6 or Proposition 24.9 implies $f \equiv cg$, with $c \in [-1/(K-1), 1]$. Conversely, if f and g are as specified and $T_n = \{[n]\}$ then $(g,f)_{T_n}[-] = g[-]$, which preserves positivity by the Schur product theorem. Else we are done by Proposition 24.9, since $k = |T_n| \leq K$. $\square$

25. Appendix C. Preservers of positivity on kernels.

We now present two Appendices on the transforms that preserve positive (semi)definiteness, and Loewner monotonicity and convexity, on kernels on infinite domains. We begin with preservers of positive semidefinite and positive definite kernels.

Definition 25.1. Let X, Y be non-empty sets, and $K : X \times Y \to \mathbb{R}$ a kernel.

(1) Given $\mathbf{x} \in X^m$ and $\mathbf{y} \in Y^n$ for integers $m, n \geq 1$, define $K[\mathbf{x}; \mathbf{y}]$ to be the $m \times n$ real matrix, with (j, k) entry $K(x_j, y_k)$.
(2) Given an integer $n \geq 1$, define $X^{n,\neq}$ to be the set of all n-tuples in X with pairwise distinct coordinates.
(3) A kernel $K : X \times X \to \mathbb{R}$ is said to be *positive semidefinite* (respectively, *positive definite*) if K is symmetric – i.e., $K(x, y) = K(y, x)\ \forall x, y \in X$ – and for all $n \geq 1$ and tuples $\mathbf{x} \in X^{n,\neq}$, the matrix $K[\mathbf{x}; \mathbf{x}]$ is positive semidefinite (respectively, positive definite).
(4) Given an integer $n \geq 1$, and a totally ordered set X, define $X^{n,\uparrow}$ to be the set of all n-tuples $\mathbf{x} = (x_1, \ldots, x_n) \in X$ with strictly increasing coordinates: $x_1 < \cdots < x_n$. (Karlin calls this the open simplex $\Delta_n(X)$ in his book [200].)

By "padding principal submatrices by the identity kernel," it is easily seen that given subsets $X \subset Y$ and a positive (semi)definite kernel K on $X \times X$, we can embed K into a kernel $\widetilde{K} : Y \times Y \to \mathbb{R}$ that is also positive (semi)definite: define $\widetilde{K}(x, y)$ to be $\mathbf{1}_{x=y}$ if $(x, y) \notin X \times X$, and $K(x, y)$ otherwise.

Now given a set X and a domain $I \subset \mathbb{R}$, we will study the inner transforms

$$\mathscr{F}_X^{\rm psd}(I) := \{F : I \to \mathbb{R} \mid \text{if } K : X \times X \to I \text{ is positive semidefinite, so is } F \circ K\},$$
$$\mathscr{F}_X^{\rm pd}(I) := \{F : I \to \mathbb{R} \mid \text{if } K : X \times X \to I \text{ is positive definite, so is } F \circ K\}.$$

(see the beginning of this text). Here, $F \circ K$ sends $X \times X$ to $\mathbb{R}$.

Notice that if X is finite then $\mathscr{F}_X^{\rm psd}(I)$ is precisely the set of entrywise maps preserving positivity on $\mathbb{P}_{|X|}(I)$; as mentioned in Section 16.8, this question remains open for all $|X| \geq 3$. If instead X is infinite, then the answer follows from Schoenberg and Rudin's results:

Theorem 25.2. *Fix $0 < \rho \leq \infty$, and suppose I is any of $(0, \rho)$, $[0, \rho)$, or $(-\rho, \rho)$. If X is an infinite set, then $\mathscr{F}_X^{\rm psd}(I)$ consists of all power series with non-negative Maclaurin coefficients, which are convergent on I.*

This observation is useful in the study of positive definite kernels in computer science.

Proof. For $I = (0, \rho)$ or $(-\rho, \rho)$ with $\rho = \infty$, the result follows by embedding every positive semidefinite matrix into a kernel on $X \times X$, and applying Theorems 16.4 and 16.3, respectively. If $I = [0, \rho)$, then from above we have the desired power series expansion on $(0, \infty)$, and it remains to show that any preserver F is right continuous at 0. To see why, first note that $F(0) \geq 0$, and F is non-decreasing and non-negative on $(0, \infty)$, so $F(0^+) := \lim_{x \to 0^+} F(x)$ exists. Now consider a three-point subset $\{x_1, x_2, x_3\}$ of X, with complement X', and define

$$K_0(x, y) := \begin{cases} 3, & \text{if } x = y, \\ 1, & \text{if } (x, y) = (x_1, x_2), (x_2, x_3), (x_3, x_2), (x_2, x_1), \\ 0, & \text{otherwise.} \end{cases} \tag{25.3}$$

Thus, K_0 is the padding by the identity of a positive definite 3×3 matrix. It follows that $F \circ (cK_0)$ is positive semidefinite for $c > 0$, so its principal submatrix $\begin{pmatrix} F(3c) & F(0) \\ F(0) & F(3c) \end{pmatrix} \in \mathbb{P}_2$. It follows by taking determinants and then $c \to 0^+$ that $F(0^+) \geq F(0) \geq 0$. Finally,

$$0 \leq \lim_{c \to 0^+} \det F[cK_0[\mathbf{x}; \mathbf{x}]] = -F(0^+)(F(0^+) - F(0))^2, \qquad \mathbf{x} := (x_1, x_2, x_3).$$

Thus, either $F(0^+) > 0$ and so $F(0^+) = F(0)$; or else $F(0^+) = 0$, hence $F(0) = 0 = F(0^+)$ as well. This ends the proof for $\rho = \infty$; for $\rho < \infty$, use the remarks in Section 22.3. $\square$

We next classify the preservers of positive definite kernels. As above, if X is finite, then the fixed-dimension case remains open; but for infinite X we have:

Theorem 25.4. *Fix $0 < \rho \leq \infty$, and suppose I is any of $(0, \rho)$, $[0, \rho)$, or $(-\rho, \rho)$. If X is an infinite set, then $\mathscr{F}_X^{\rm pd}(I)$ consists of all non-constant power series with non-negative Maclaurin coefficients, which are convergent on I.*

Proof. By the Schur product theorem, every monomial x^k for $k \geq 1$ preserves positive definiteness. This observation shows one implication. Conversely, first say $F \in \mathscr{F}_X^{\rm pd}(I)$ is continuous. Now every positive semidefinite kernel $K : X \times X \to \mathbb{R}$ is the pointwise limit as $\epsilon \to 0^+$ of the family $K_\epsilon : X \times X \to \mathbb{R}$, given by $K_\epsilon(x, y) := K(x, y) + \epsilon \mathbf{1}_{x=y}$. It follows that F preserves positive semidefinite kernels, reducing the problem to the preceding result. Moreover, F is not constant, e.g., by considering its action on the identity kernel.

The rest of the proof is devoted to showing that F is continuous on I. First suppose $I = (0, \rho)$ and $A \in \mathbb{P}_2(I)$ is positive definite. Then there exists $\epsilon \in (0, \rho/2)$ such that $A' := A - \epsilon \, \mathrm{Id}_{2 \times 2}$ is still positive definite. Choose $x_1, x_2 \in X$ and define the kernel

$$K : X \times X \to \mathbb{R}, \qquad (x, y) \mapsto \begin{cases} a_{jk}, & \text{if } x = x_j, y = x_k, \quad 1 \leq j, k \leq 2; \\ \rho/2, & \text{if } x = y \notin \{x_1, x_2\}; \\ \epsilon, & \text{otherwise.} \end{cases}$$

Clearly,

$$K = \epsilon \mathbf{1}_{X \times X} + (A' \oplus (\rho/2 - \epsilon) \, \mathrm{Id}_{X \setminus \{x_1, x_2\}}),$$

and so K is positive definite on X with all values in $I = (0, \rho)$. Hence, $F \circ K$ is also positive definite. It follows that the entrywise map $F[-]$ preserves positive definiteness on 2×2 matrices. Now invoke Lemma 12.14 to conclude that F is continuous on $(0, \rho)$.

This concludes the proof for $I = (0, \rho)$. Next, suppose $I = [0, \rho)$; by the preceding case, F is given by a non-constant power series as asserted, and we just need to show F is right continuous at 0. Since F is increasing on $(0, \rho)$, the limit $F(0^+) := \lim_{x \to 0^+} F(x)$ exists and $F(0^+) \geq F(0) \geq 0$. Now use the kernel K_0 from (25.3) and repeat the subsequent arguments.

The final case is if $I = (-\rho, \rho)$. In this case we fix $u_0 \in (0, 1)$ and a countable subset $Y := \{x_0, x_1, \dots\} \subset X$. Denote $Y^c := X \setminus Y$. Given $a, b > 0$ such that $a + b < \rho$, let

$$\mu = \mu_{a,b} := a\delta_{-1} + b\delta_{u_0}.$$

The corresponding Hankel moment matrix is H_μ, with (j, k) entry $a(-1)^{j+k} + bu_0^{j+k}$, and this is positive semidefinite of rank 2. Now for each $\epsilon > 0$, define $K_\epsilon : X \times X \to \mathbb{R}$, via

$$K_\epsilon(x, y) := \begin{cases} H_\mu(j, j) + \epsilon, & \text{if } x = y = x_j, \ j \geq 0; \\ H_\mu(j, k), & \text{if } (x, y) = (x_j, x_k), \ j \neq k; \\ \epsilon, & \text{if } x = y \in Y^c; \\ 0, & \text{otherwise.} \end{cases}$$

Clearly, K_ϵ is positive definite, with entries in $I = (-\rho, \rho)$ for sufficiently small $\epsilon > 0$. It follows that $F \circ K_\epsilon$ is positive definite. Since F is continuous on $[0, \rho)$ by the previous cases, $\lim_{\epsilon \to 0^+} F \circ K_\epsilon = F[H_\mu \oplus 0_{Y^c \times Y^c}]$ is positive semidefinite, and so $F[-]$ preserves positivity on the Hankel moment matrices H_μ for all $\mu = \mu_{a,b}$ as above. It follows by the proof of Step 3 for the stronger Schoenberg theorem above (see the computations following Lemma 20.4) that F is continuous on $(-\rho, \rho)$, as desired. This concludes the proof in all cases. $\square$

26. Appendix D. Preservers of Loewner monotonicity and convexity on kernels.

Thus far, we have studied the preservers of (total) positivity and related variants, with a brief look in Section 15 at entrywise powers preserving other Loewner properties. In this Appendix, we return to these properties. Specifically, we classify all composition operators preserving Loewner monotonicity and convexity, on kernels on infinite domains. (The case of finite domains remains open, as for positivity preservers.)

The results for infinite domains will crucially use the finite versions; thus, we begin by reminding the reader of the definitions. Roughly speaking, a function is *Loewner monotone* (see Definition 14.7) if $f[A] \geq f[B]$ whenever $A \geq B \geq 0_{n\times n}$. Similarly, a function is *Loewner convex* (see Definition 15.7) if $f[\lambda A + (1-\lambda)B] \leq \lambda f[A] + (1-\lambda) f[B]$ whenever $A \geq B \geq 0$ and $\lambda \in [0,1]$.

As explained in Remark 14.8, for $n = 1$ the usual notion of a monotonically non-decreasing function coincides with Loewner monotonicity. The same holds for convex functions vis-à-vis Loewner convex functions, for $n = 1$. Now for $n = 1$, a differentiable function $f : (0,\infty) \to \mathbb{R}$ is monotone (respectively, convex) if and only if f' is non-negative, i.e., has image in $[0,\infty)$ (respectively, monotone). The following result by Hiai in *Linear Algebra Appl.* (2009) extends this to the corresponding Loewner properties, in every dimension:

Theorem 26.1 (Hiai, fixed dimension)**.** *Suppose $0 < \rho \leq \infty$, $I = (-\rho,\rho)$, and $f : I \to \mathbb{R}$.*

1. *Given $n \geq 2$, the function f is Loewner convex on $\mathbb{P}_n(I)$ if and only if f is differentiable on I and f' is Loewner monotone on $\mathbb{P}_n(I)$. This result also holds if we restrict both test sets to rank $\leq k$ matrices in $\mathbb{P}_n(I)$ for every $2 \leq k \leq n$.*
2. *Given $n \geq 3$, the function f is Loewner monotone on $\mathbb{P}_n(I)$ if and only if f is differentiable on I and f' is Loewner positive on $\mathbb{P}_n(I)$.*

Recall the related but somewhat weaker variant in Proposition 15.9.

Here we show the first part and a weaker version of the second part of Theorem 26.1 – see Hiai's 2009 paper for the complete proof. (Note: Hiai showed the first part only for $k = n$; also, we do not use the second part in the present text.) First, as a consequence of the first part and the previous results, we obtain the following Schoenberg-type classification of the corresponding "dimension-free" entrywise preservers:

Theorem 26.2 (Dimension-free preservers of monotonicity and convexity)**.** *Suppose $0 < \rho \leq \infty$, $I = (-\rho,\rho)$, and $f : I \to \mathbb{R}$. The following are equivalent:*

1. *f is Loewner monotone on $\mathbb{P}_n(I)$ for all n.*
2. *f is Loewner monotone on the rank ≤ 3 Hankel matrices in $\mathbb{P}_n(I)$ for all n.*
3. *$f(x) = \sum_{k=0}^\infty c_k x^k$ on I, with $c_1, c_2, \dots \geq 0$.*

Similarly, the following are equivalent conditions characterizing Loewner convexity:

1. *f is Loewner convex on $\mathbb{P}_n(I)$ for all n.*
2. *f is Loewner convex on the rank ≤ 3 matrices in $\mathbb{P}_n(I)$ for all n.*
3. *$f(x) = \sum_{k=0}^\infty c_k x^k$ on I, with $c_2, c_3, \dots \geq 0$.*

Proof. We begin with the dimension-free Loewner monotone maps. Clearly, $(1) \implies (2)$. To show $(3) \implies (1)$, note that $f(x) - c_0$ is also Loewner monotone for any $c_0 \in \mathbb{R}$ if $f(x)$ is, so it suffices to consider $f(x) = x^k$ for $k \geq 1$. But such a function is clearly monotone, by the Schur product theorem. This is an easy exercise, or see e.g., the proof of Theorem 14.9. Finally, note from the definition of Loewner monotonicity that $f - f(0)$ entrywise preserves positivity if f is Loewner monotone – on $\mathbb{P}_n(I)$ or on subsets of these that contain the zero matrix. In particular, if (2) holds then $f - f(0)$ is a dimension-free positivity preserver, hence of the form $\sum_{k\geq 0} c_k x^k$ with all $c_k \geq 0$ by Theorem 19.10 – or more precisely, its variant for restricted domains $(-\rho, \rho)$ as in Section 22.3. Since $f - f(0)$ also vanishes at the origin, we have $c_0 = 0$, proving (3).

We next come to convexity preservers. Clearly, (1) $\implies$ (2). To show (3) $\implies$ (1), note that $f(x) - c_0 - c_1 x$ is also Loewner convex for any $c_0, c_1 \in \mathbb{R}$ if $f(x)$ is, so it again suffices to consider $f(x) = x^k$ for integers $k \geq 2$. In fact, we claim by induction that x^k is Loewner convex for all $k \geq 0$. The convexity of $1, x$ is immediate, and for the induction step, if x^k is convex, then for any integer $n \geq 1$, scalar $\lambda \in [0,1]$, and matrices $A \geq B \geq 0_{n\times n}$,

$$\begin{aligned}(\lambda A + (1-\lambda)B)^{\circ(k+1)} &\leq (\lambda A + (1-\lambda)B) \circ (\lambda A^{\circ k} + (1-\lambda)B^{\circ k})\\ &= \lambda A^{\circ(k+1)} + (1-\lambda)B^{\circ(k+1)} - \lambda(1-\lambda)(A-B)\circ(A^{\circ k} - B^{\circ k})\\ &\leq \lambda A^{\circ(k+1)} + (1-\lambda)B^{\circ(k+1)},\end{aligned}$$

where the final inequality follows from the Loewner monotonicity of x^k and the Schur product theorem. Finally, if (2) holds, then by Theorem 26.1(1) for $k = 3$, f' exists and is Loewner monotone on rank ≤ 3 matrices in $\mathbb{P}_n(I)$ for all n, hence a power series as in the preceding set of equivalent statements. This immediately implies (3). $\square$

The remainder of this section is devoted to proving Theorem 26.1(1), beginning with some elementary properties of convex functions:

Lemma 26.3 (Convex functions)**.** *Suppose $I \subset \mathbb{R}$ is an interval and $f : I \to \mathbb{R}$ is convex.*

(1) The function $(s,t) \mapsto \dfrac{f(t) - f(s)}{t - s}$, where $t > s$, is non-decreasing in both $t, s \in I$.
(2) If I is open, then $f'_{\pm}$ exist on I. In particular, f is continuous on I.
(3) If I is open and $z_1 < x < z_2$ in I, then $f'_+(z_1) \leq f'_-(x) \leq f'_+(x) \leq f'_-(z_2)$. In particular, $f'_{\pm}$ are non-decreasing in I, hence each continuous except at countably many points of jump discontinuity.
(4) If I is open, then for all $x \in I$,

$$f'_+(x) = \lim_{z \to x^+} f'_{\pm}(z), \qquad f'_-(x) = \lim_{z\to x^-} f'_{\pm}(z).$$

(5) If I is open, there exists a co-countable (therefore dense) subset $D \subset I$ on which f' exists. Moreover, f' is continuous and non-decreasing on D.

Note that the assertions involving open intervals I may be carried over to the interiors of arbitrary intervals I on which f is convex.

Proof.

(1) Suppose $s < t < u$ lie in I. One needs to show

$$\frac{f(t)-f(s)}{t-s} \leq \frac{f(u)-f(s)}{u-s} \leq \frac{f(u)-f(t)}{u-t}.$$

But both inequalities can be reformulated to say

$$f(t) \leq \frac{u-t}{u-s} f(s) + \frac{t-s}{u-s} f(u),$$

which holds as f is convex.

(2) Given $t \in I$, choose $s < t < u$ in I, and note by the previous part that the ratio
$$\frac{f(x) - f(t)}{x - t}, \qquad x \in (t, u)$$
is non-increasing in x as $x \to t^+$ and bounded below by $\dfrac{f(t) - f(s)}{t - s}$. Thus, $f'_+(t)$ exists; a similar argument works to show $f'_-(t)$ exists. In particular, the two limits $\lim_{x \to t^\pm} f(x) - f(t)$ are both zero, proving f is continuous at $t \in I$.

(3) The second sentence follows from the first, which in turn follows from the first part by taking limits and is left to the reader.

(4) The preceding part implies $f'_\pm(z)$ are non-decreasing as $z \to x^-$ and non-increasing as $z \to x^+$, and shows "half" of the desired inequalities. We now show $f'_+(x) \geq \lim_{z \to x^+} f'_+(z)$; the remaining similar inequalities are shown similarly, and again left to the reader. Let $y \in I,\ y > x$; then the first part implies
$$\frac{f(y) - f(z)}{y - z} \geq \frac{f(y') - f(z)}{y' - z}, \qquad \forall x < z < y' < y.$$
Taking $y' \to z^+$, we have $f'_+(z) \leq \frac{f(y)-f(z)}{y-z}$. From above, f is continuous on I, so
$$\frac{f(y) - f(x)}{y - x} = \lim_{z \to x^+} \frac{f(y) - f(z)}{y - z} \geq \lim_{z \to x^+} f'_+(z).$$
Finally, taking $y \to x^+$ concludes the proof.

(5) Let $D \subset I$ be the subset where f' exists, which is if and only if f'_+ is continuous (by the preceding part). In particular, D is co-countable from a previous part, and $f' = f'_+$ is continuous and non-decreasing on D by the same part. □

The next preliminary result shows the continuity (respectively, differentiability) of monotone (respectively, convex) functions on 2×2 matrices:

Proposition 26.4. *Suppose* $0 < \rho \leq \infty$, $I = (-\rho, \rho)$, *and* $g : I \to \mathbb{R}$.

(1) If $g[-]$ *is monotone on* $\mathbb{P}_2(I)$, *then* g *is continuous on* I.

(2) If $g[-]$ *is convex on* $\mathbb{P}_2(I)$, *then* g *is differentiable on* I.

Proof. We begin with the first assertion. It is easily verified that if g is monotone on $\mathbb{P}_2(I)$, then $g - g(0)$, when applied entrywise to $\mathbb{P}_2(I)$, preserves positivity. Hence, by (the bounded domain-variant of) Theorem 12.7, g is continuous on $(0, \rho)$. Moreover, we may assume without loss of generality that $g(0) = 0$.

Now let $0 < a < \rho$ and $0 < \epsilon < \rho - a$; then the monotonicity of g implies
$$\begin{pmatrix} a+\epsilon & a \\ a & a \end{pmatrix} \geq \begin{pmatrix} \epsilon & 0 \\ 0 & 0 \end{pmatrix} \geq 0_{2\times 2} \qquad \Longrightarrow \qquad \begin{pmatrix} g(a+\epsilon) - g(\epsilon) & g(a) \\ g(a) & g(a) \end{pmatrix} \geq 0_{2\times 2}.$$
Pre- and post-multiplying this last matrix by $(1, -1)$ and $(1, -1)^T$ respectively, we have $g(a + \epsilon) - g(a) \geq g(\epsilon)$, and by the monotonicity of g (applied to $\epsilon \mathbf{1}_{2\times 2} \geq \epsilon' \mathbf{1}_{2\times 2}$ for $0 \leq \epsilon' < \epsilon < \rho$), it follows that g is non-decreasing on $[0, \rho)$. Now taking the limit as $\epsilon \to 0^+$, we have
$$0 = g(a^+) - g(a) \geq g(0^+) \geq 0,$$
where the first equality follows from the continuity of g. Hence, g is right continuous at 0.

Next, for the continuity of g on $(-\rho, 0)$, let
$$a \in (0, \rho), \quad 0 < \epsilon < \min(a, \rho - a), \quad A = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix},$$

and deduce from the monotonicity of g:

$$(a+\epsilon)A \geq aA \geq (a-\epsilon)A \geq 0 \quad \implies \quad g[(a+\epsilon)A] \geq g[aA] \geq g[(a-\epsilon)A].$$

The positivity of the difference matrices implies, upon taking determinants:

$$|g(a \pm \epsilon) - g(a)| \geq |g(-a \mp \epsilon) - g(-a)|.$$

Let $\epsilon \to 0^+$; then the continuity of g at a implies that at $-a$, as desired. A similar (one-sided) argument shows the left continuity of g at 0, via the step $g(\epsilon) - g(0) \geq |g(-\epsilon) - g(0)|$.

Next, we come to the second assertion. If g is convex on $\mathbb{P}_2(I)$, then restricting to the matrices $a\mathbf{1}_{2\times 2}$ for $a \in [0,\rho)$, it follows that g is convex on $[0,\rho)$. Hence, g'_+ exists on $(0,\rho)$ by Lemma 26.3. Now suppose $0 < s < t < \rho$ and $0 < \epsilon < \rho - t$. Then by the convexity of g,

$$\begin{pmatrix} t+\epsilon & t \\ t & t \end{pmatrix} \geq \begin{pmatrix} s+\epsilon & s \\ s & s \end{pmatrix} \geq 0_{2\times 2} \implies \begin{pmatrix} g(\lambda(t+\epsilon)+(1-\lambda)(s+\epsilon)) & g(\lambda t+(1-\lambda)s) \\ g(\lambda t+(1-\lambda)s) & g(\lambda t+(1-\lambda)s) \end{pmatrix}$$
$$\leq \lambda \begin{pmatrix} g(t+\epsilon) & g(t) \\ g(t) & g(t) \end{pmatrix} + (1-\lambda) \begin{pmatrix} g(s+\epsilon) & g(s) \\ g(s) & g(s) \end{pmatrix},$$

for all $\lambda \in [0,1]$. Write this inequality in the following form: $\begin{pmatrix} \alpha & \beta \\ \beta & \beta \end{pmatrix} \geq 0_{2\times 2}$. As above, pre- and post-multiplying this last matrix by $(1,-1)$ and $(1,-1)^T$ respectively yields

$$g(\lambda t+(1-\lambda)s+\epsilon) - g(\lambda t+(1-\lambda)s) \leq \lambda(g(t+\epsilon)-g(t)) + (1-\lambda)(g(s+\epsilon)-g(s)).$$

Divide by ϵ and let $\epsilon \to 0^+$; this shows g'_+ is convex, hence continuous by Lemma 26.3, on $(0,\rho)$.

Next, denote by g_0, g_1 the even and odd parts of g, respectively:

$$g_0(t) := \frac{1}{2}(g(t)+g(-t)), \qquad g_1(t) := \frac{1}{2}(g(t)-g(-t)).$$

We claim that g_0, g_1 are convex on $[0,\rho)$. Indeed, by the convexity of g we deduce for $0 \leq s \leq t < \rho$ and $\lambda \in [0,1]$:

$$\begin{pmatrix} t & -t \\ -t & t \end{pmatrix} \geq \begin{pmatrix} s & -s \\ -s & s \end{pmatrix} \geq 0$$
$$\implies \begin{pmatrix} g(c_\lambda) & g(-c_\lambda) \\ g(-c_\lambda) & g(c_\lambda) \end{pmatrix} \leq \lambda \begin{pmatrix} g(t) & g(-t) \\ g(-t) & g(t) \end{pmatrix} + (1-\lambda) \begin{pmatrix} g(s) & g(-s) \\ g(-s) & g(s) \end{pmatrix},$$

where $c_\lambda = \lambda t + (1-\lambda)s$. Pre- and post-multiplying this last inequality by $(1,\pm 1)$ and $(1,\pm 1)^T$ respectively, yields

$$g(\lambda t+(1-\lambda)s) \pm g(-(\lambda t+(1-\lambda)s)) \leq \lambda(g(t) \pm g(-t)) + (1-\lambda)(g(s) \pm g(-s)).$$

This yields: g_0, g_1 are convex on $[0,\rho)$. Next, note that if $0 \leq s < t < \rho$, and $0 < \epsilon \leq \min(t-s, \rho-t)$, then

$$\frac{g(s+\epsilon)-g(s)}{\epsilon} \leq \frac{g(t)-g(t-\epsilon)}{\epsilon},$$

by Lemma 26.3(1). Taking $\epsilon \to 0^+$ shows that $g'_+(s) \leq g'_-(t)$ if $0 \leq s < t < \rho$ and g is convex. Similarly, $g'_-(t) \leq g'_+(t)$; therefore,

$$g'_+(s) \leq g'_-(t) = (g_0)'_-(t) + (g_1)'_-(t) \leq (g_0)'_+(t) + (g_1)'_+(t) = g'_+(t).$$

Since g'_+ is continuous, letting $s \to t^-$ shows $(g_j)'_-(t) = (g_j)'_+(t)$ for $j = 0, 1$. Thus, g_j is differentiable on $(0,\rho)$. Since g_0 is even and g_1 is odd, they are also differentiable on $(-\rho, 0)$. Hence, g_0, g_1, g are differentiable on $I \setminus \{0\}$. Finally, let $I' := (-2\rho/3, 2\rho/3)$ and define

$h(x) := g(x + \rho/3)$. It is easy to check that h is convex on $\mathbb{P}_2(I')$, so it is differentiable at $-\rho/3$ by the above analysis, and hence g is differentiable at 0, as desired. $\square$

With these preliminary results in hand, we now complete the remaining proof:

Proof of Theorem 26.1. We begin by showing the first assertion. First, suppose f is differentiable on I and f' is monotone on the rank $\leq k$ matrices in $\mathbb{P}_n(I)$. Also assume $A \geq B \geq 0_{n\times n}$ are matrices of rank $\leq k$. Now follow the proof of Proposition 15.9(3) to show that $f[-]$ is Loewner convex on rank $\leq k$ matrices in $\mathbb{P}_n(I)$. Here we use the fact that since $A \geq B \geq 0$, we have the chain of Loewner inequalities

$$A \geq \lambda A + (1-\lambda)B \geq \lambda \frac{A+B}{2} + (1-\lambda)B \geq B, \tag{26.5}$$

and hence the ranks of all matrices here are at most $\operatorname{rk}(A) \leq k$.

The converse is shown in two steps; in fact, we will also prove that f is continuously differentiable on I. The first step is to show the result for $n = k = 2$. Note by Proposition 26.4 that f is differentiable on I. Now say $A \geq B \geq 0_{2\times 2}$ with $A \neq B$ in $\mathbb{P}_2(I)$. Writing $A = \begin{pmatrix} a_1 & a \\ a & a_2 \end{pmatrix}$ and $B = \begin{pmatrix} b_1 & b \\ b & b_2 \end{pmatrix}$, we have $a_j \geq b_j \geq 0$ for $j = 1,2$ and $(a-b)^2 \leq (a_1-b_1)(a_2-b_2)$. Define $\delta \in [0, a_1 - b_1]$ and the matrix $C_{2\times 2}$ via

$$(a-b)^2 = (a_1 - b_1 - \delta)(a_2 - b_2), \qquad C := \begin{pmatrix} b_1 + \delta & b \\ b & b_2 \end{pmatrix} \in \mathbb{P}_2(I).$$

Clearly, $A \geq C \geq B$, all matrices are in $\mathbb{P}_2(I)$, and $A - C, C - B$ have rank at most 1. Thus, we may assume without loss of generality that $A - B$ has rank 1; write $A - B = \begin{pmatrix} a & \sqrt{ab} \\ \sqrt{ab} & b \end{pmatrix} \in \mathbb{P}_2$. First, if $ab = 0$, then $f'[A] - f'[B]$ is essentially a scalar on the main diagonal. Now since f is convex on $[0,\rho)$ by considering $a\mathbf{1}_{2\times 2}$ for $a \in [0,\rho)$, we have f' is non-decreasing on $(0,\rho)$, and hence $f'[A] \geq f'[B]$.

The other case is $a, b > 0$. In this case $A \geq B \geq 0$ and $A - B$ is rank-1 with no zero entries. Now follow the proof of Proposition 15.9(3) to infer $f'[A] \geq f'[B]$. Together, both cases show that f' is monotone on $\mathbb{P}_2(I)$, so f' is continuous on I by Proposition 26.4(1).

This shows the result for $n = k = 2$. Now suppose $n > 2$. First, f is convex on $\mathbb{P}_2(I)$, so f' is monotone on $\mathbb{P}_2(I)$ and hence continuous on I by the previous case. Second, to show that f' is monotone as asserted, suppose $A \geq B \geq 0_{n\times n}$ are matrices in $\mathbb{P}_n(I)$ of rank $\leq k$. Now claim that there is a chain of Loewner matrix inequalities

$$A = A_n \geq A_{n-1} \geq \cdots \geq A_0 = B,$$

satisfying: (1) $A_j \in \mathbb{P}_n(I)$ for all $0 \leq j \leq n$, and (2) $A_j - A_{j-1}$ has rank at most 1 for each $1 \leq j \leq n$. Note that such a chain of inequalities would already imply the reverse inclusions for the corresponding null spaces, so each A_j has rank at most k.

To show the claim, spectrally decompose $A - B = UDU^T$, where U is orthogonal and $D = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$ with $\lambda_j \geq 0$, and write

$$A_j := B + U \operatorname{diag}(\lambda_1, \dots, \lambda_j, 0, \dots, 0) U^T, \qquad 0 \leq j \leq n.$$

Note that $A_j \leq A$, so the same applies to each of their corresponding (non-negative) diagonal entries. Thus, $0 \leq (A_j)_{ll} \leq a_{ll}$ for $1 \leq l \leq n$. Thus, the diagonal entries of each A_j lie in $I = (-\rho, \rho)$, hence so do the off-diagonal entries. This shows the claim.

Thus, to show $f'[A] = f'[A_n] \geq f'[A_0] = f'[B]$, it suffices to assume, as in the previous case of $n = k = 2$, that $A - B$ has rank 1. First, if $A - B$ has no zero entries, then $f'[A] \geq f'[B]$ by

Proposition 15.9(3). Otherwise, suppose $A-B = uu^T$, with $u \in \mathbb{R}^n$ a non-zero vector having zero entries. Without loss of generality, write $u = \begin{pmatrix} v \\ 0 \end{pmatrix}$, with $v \in \mathbb{R}^l$ having no zero entries for some $1 \leq l \leq n-1$. Accordingly, write $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$, and similarly for B; it follows that $A_{ij} = B_{ij}$ for all $(i,j) \neq (1,1)$, and $A_{11} = B_{11} + vv^T$. Now since f is Loewner convex on $[B, A]$, it is so on $[B_{11}, A_{11}]$, where all matrices are positive semidefinite and also have rank $\leq k$. Moreover, f' exists and is continuous on I from above. Again, by Proposition 15.9(3), it follows that f' is Loewner monotone on $[B_{11}, A_{11}]$ (if $k = 1$ then this assertion is true by one-variable calculus). But then,

$$f'[A] - f'[B] = \begin{pmatrix} f'[A_{11}] - f'[B_{11}] & 0 \\ 0 & 0 \end{pmatrix} \geq 0.$$

This proves the first assertion; we turn to the second. First, suppose $A \geq B \geq 0_{n \times n}$, and f' is Loewner positive on $\mathbb{P}_n(I)$. Then follow the proof of Proposition 15.9(2) to infer $f[A] \geq f[B]$. Conversely, we prove the result under a stronger hypothesis: namely, f is differentiable. Now the proof of Proposition 15.9(2) again applies: given $A \in \mathbb{P}_n(I)$, we have $A + \epsilon \mathbf{1}_{n \times n} \in \mathbb{P}_n(I)$ for small $\epsilon > 0$. By monotonicity, it follows that

$$\tfrac{1}{\epsilon}(f[A + \epsilon \mathbf{1}_{n \times n}] - f[A]) \in \mathbb{P}_n.$$

Taking $\epsilon \to 0^+$ proves $f'[A] \in \mathbb{P}_n$, as desired. □

We now move to kernels.

Definition 26.6. Suppose X is a non-empty set, $I \subset \mathbb{R}$ a domain, and $\mathcal{V}$ is a set of (real symmetric) positive semidefinite kernels on $X \times X$, with values in I.

(1) The *Loewner order* on kernels on $X \times X$ is: $K \succeq L$ for K, L kernels on $X \times X$, if $K - L$ is a positive semidefinite kernel. (Note, if X is finite, this specializes to the usual Loewner ordering on real $|X| \times |X|$ matrices.)
(2) A function $F : I \to \mathbb{R}$ is *Loewner monotone on* $\mathcal{V}$ if $F \circ K \succeq F \circ L$ whenever $K \succeq L \succeq 0$ are kernels in $\mathcal{V}$.
(3) A function $F : I \to \mathbb{R}$ is *Loewner convex on* $\mathcal{V}$ (here I is assumed to be convex) if whenever $K \succeq L \succeq 0$ are kernels in $\mathcal{V}$, we have
$$\lambda F \circ K + (1-\lambda) F \circ L \succeq F \circ (\lambda K + (1-\lambda) L), \qquad \forall \lambda \in [0,1].$$

The above results for matrices immediately yield the results for kernels:

Theorem 26.7. *Suppose $0 < \rho \leq \infty$, $I = (-\rho, \rho)$, and $F : I \to \mathbb{R}$, and X is an infinite set. The composition map $F \circ -$ is Loewner monotone (respectively, Loewner convex) on positive kernels on $X \times X$, if and only if F satisfies the respective equivalent conditions on matrices of all sizes, in Theorem 26.2.*

Proof. First, if F is Loewner monotone or convex on kernels on $X \times X$, then by restricting the defining inequalities to kernels on $Y \times Y$ (padded by zeros) for finite sets $Y \subset X$, it follows that F is, respectively, Loewner monotone or convex on $\mathbb{P}_n(I)$ for all $n \geq 1$.

To show the converse, suppose first that $F(y) = \sum_{k=0}^\infty c_k y^k$ on I, with $c_1, c_2, \dots \geq 0$. To show that $F \circ -$ is Loewner monotone on kernels on $X \times X$, it suffices to do so on every "principal submatrix" of such kernels – i.e., for every finite indexing subset of X. But this is indeed true for F, by Theorem 26.2. A similar proof holds for Loewner convex maps. □

27. Appendix E. Menger's results and Euclidean distance geometry.

We conclude this part of the text with a brief detour into the same area where we started this part of the text: metric geometry, specifically, that of Euclidean spaces $\mathbb{R}^n$ – and of their closure, Hilbert space ℓ^2. This is a beautiful area of mathematical discovery, which has featured work by several prominent mathematicians, including Birkhoff, Cauchy, Cayley, Gödel, Menger, Schoenberg, and von Neumann, among others. See [237] for a modern exposition of some of the gems of distance geometry (which begins, interestingly, with Heron's formula for the area of a triangle, from two millennia ago).

The main result of this section is a 1928 theorem of Menger:

Theorem 27.1 (Menger, [258]; see also [315]). *A metric space (X, d) can be isometrically embedded in Hilbert space ℓ^2 if and only if X is separable and every subset of X of size $n+1$ can be isometrically embedded in $\mathbb{R}^n$ (equivalently, in ℓ^2) for $n \geq 2$.*

This result, together with Schoenberg's theorems 16.10 and 16.17 on Hilbert space embeddings of finite metric spaces X, immediately yields those theorems for all separable X:

Theorem 27.2 (Schoenberg). *Suppose (X, d) is a separable metric space.*

(1) X embeds isometrically into Hilbert space ℓ^2 if and only if for every integer $n \geq 2$ and $(n+1)$-tuple of points $Y := (x_0, \dots, x_n)$ in X, the "alternate Cayley–Menger matrix" $CM'(Y) := (d(x_0, x_j)^2 + d(x_0, x_k)^2 - d(x_j, x_k)^2)_{j,k=1}^n$ is positive semidefinite.

(2) X embeds isometrically into Hilbert space ℓ^2 if and only if the Gaussian kernel $\exp(-\sigma(\cdot)^2)$ is a positive definite function on X for all $\sigma > 0$ (equivalently, for some sequence σ_m of positive numbers decreasing to 0^+).

Here, we explore some simple, yet beautiful observations in Euclidean distance geometry, which help prove Theorem 27.1, and also provide connections to Cayley–Menger matrices [78, 259] and to n-point homogeneous spaces (see Remark 16.22). We begin with the latter.

27.1. n-point homogeneity of Euclidean and Hilbert spaces.

As early as 1944, in his influential work [50] Birkhoff defines a metric space (X, d) to be *n-point homogeneous* if given two equinumerous subsets of X of size at most n, an isometry between them extends to a self-isometry of X. The heart of the present proof of Theorem 27.1 is to show that Euclidean space $\mathbb{R}^k$ is n-point homogeneous for all $k, n \geq 1$:

Theorem 27.3. *Fix an integer $k \geq 1$.*

(1) The Euclidean space $\mathbb{R}^k$ with the Euclidean metric is n-point homogeneous for all n. More strongly: any isometry between two subsets $M, N \subset \mathbb{R}^k$ is, up to a translation, the restriction of an orthogonal linear transformation of $\mathbb{R}^k$.

(2) Hilbert space ℓ^2 is n-point homogeneous for all n.

The first step in proving Theorem 27.3 is the following observation about Gram matrices:

Lemma 27.4. *Given vectors $y_0, \dots, y_n \in \ell^2$ for some $n \geq 0$, the Gram matrix $(\langle y_j, y_k \rangle)_{j,k=0}^n$ is invertible if and only if the y_j are linearly independent.*

Proof. We prove the contrapositive. If $\sum_{k=0}^n c_k y_k = 0$ is a nontrivial linear combination, then applying $\langle y_j, - \rangle$ for all j yields $\mathrm{Gram}((y_k)_k)\mathbf{c} = 0$, where $\mathbf{c} = (c_0, \dots, c_n)^T \neq 0$. Conversely, if $\mathrm{Gram}((y_k)_k)\mathbf{c} = 0$ and $\mathbf{c} \neq 0$, then

$$0 = \mathbf{c}^T \, \mathrm{Gram}((y_k)_k)\mathbf{c} = \left\| \sum_{k=0}^n c_k y_k \right\|^2,$$

so that some nontrivial linear combination of the y_k vanishes, as desired. □

This simple lemma leads to striking consequences. We will presently mention two, the first of which involves a classical concept (which already featured in Theorem 16.10):

Definition 27.5. Given a metric space $X = (\{x_0, x_1, \dots, x_n\}, d)$, write $d_{jk} := d(x_j, x_k)$ for $0 \le j, k \le n$. The associated *Cayley–Menger matrix* is

$$CM(X)_{(n+2)\times(n+2)} := \begin{pmatrix} 0 & d_{01}^2 & d_{02}^2 & \cdots & d_{0n}^2 & 1 \\ d_{10}^2 & 0 & d_{12}^2 & \cdots & d_{1n}^2 & 1 \\ d_{20}^2 & d_{21}^2 & 0 & \cdots & d_{2n}^2 & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ d_{n0}^2 & d_{n1}^2 & d_{n2}^2 & \cdots & 0 & 1 \\ 1 & 1 & 1 & \cdots & 1 & 0 \end{pmatrix}, \tag{27.6}$$

Similarly, the "alternate form" of the Cayley–Menger matrix here is

$$CM'(X)_{n\times n} := \begin{pmatrix} 2d_{01}^2 & d_{01}^2 + d_{02}^2 - d_{12}^2 & \cdots & d_{01}^2 + d_{0n}^2 - d_{1n}^2 \\ d_{01}^2 + d_{02}^2 - d_{12}^2 & 2d_{02}^2 & \cdots & d_{01}^2 + d_{0n}^2 - d_{1n}^2 \\ \vdots & \vdots & \ddots & \vdots \\ d_{01}^2 + d_{0n}^2 - d_{1n}^2 & d_{02}^2 + d_{0n}^2 - d_{2n}^2 & \cdots & 2d_{0n}^2 \end{pmatrix} \tag{27.7}$$

Recall that the positive semidefiniteness of the second matrix features in Schoenberg's recasting of Menger and Fréchet's results on Hilbert space embeddings of finite metric spaces. We now write down a preliminary observation that relates the two determinants above:

Lemma 27.8. *For all finite metric spaces X with at least two points,*

$$\det CM(X) = (-1)^{|X|} \det CM'(X). \tag{27.9}$$

Proof. Starting with the matrix $CM(X)$, perform elementary row and column operations, leaving the determinant unchanged. First, subtract the first row from all non-extremal rows. Then subtract the first column and d_{0j}^2 times the last column each from the non-extremal columns. This yields

$$\begin{pmatrix} 0 & \mathbf{0}_n^T & 1 \\ (d_{i0}^2)_i & -CM'(X) & \mathbf{0}_n \\ 1 & \mathbf{0}_n^T & 0 \end{pmatrix},$$

a bordered matrix with determinant $(-1)^{n+1} \det CM'(X)$, as desired. □

We can now state and prove the two consequences of Lemma 27.4 promised above. The first of these is a well-known result, proved in 1841 by Cayley during his undergraduate days [78]. The second is the underlying principle behind the Global Positioning System, or GPS – *trilateration* (also referred to more colloquially as "triangulation"): every point in the plane (or on the surface of a sphere "like" the Earth) is uniquely determined by intersecting three circles that denote distances from three non-collinear points (or four spheres centered at four non-coplanar points).

Proposition 27.10.

(1) *(Cayley, [78]). Suppose an isometry Ψ sends a finite metric space $(X = \{x_0, \dots, x_n\}, d)$ into Hilbert space ℓ^2. Then the vectors $\Psi(x_0), \dots, \Psi(x_n)$ are affine linearly dependent (i.e., lie on an $(n-1)$-dimensional subspace) in ℓ^2, if and only if the Cayley–Menger determinant of X vanishes.*

(2) *Fix vectors $y_0 = 0, y_1, \dots, y_n \in \ell^2$. Given any $y \in \ell^2$, the following are equivalent:*

(a) *y is* (uniquely) *determined by the tuple of Euclidean distances* $(\|y\|, \|y-y_1\|, \dots, \|y-y_n\|) \in \mathbb{R}^{n+1}$.

(b) *y is in the span of $y_1, \dots, y_n$.*

Proof.

(1) Denote $y_j := \Psi(x_j)$ for $0 \leq j \leq n$. Now compute, as in Equation (16.12) in the proof of Theorem 16.10:

$$d(y_0, y_j)^2 + d(y_0, y_k)^2 - d(y_j, y_k)^2 = \langle y_0 - y_j, y_0 - y_k \rangle,$$

so that $CM'(X) = \mathrm{Gram}((y_0 - y_j)_{j=1}^n)$. Now $CM(X)$ is singular if and only if so is $CM'(X)$. From above and by Lemma 27.4, this happens if and only if the vectors $y_0 - y_j$, $1 \leq j \leq n$ are linearly dependent. This completes the proof.

(2) For this part, let $V \subset \ell^2$ denote the span of the y_j. First, suppose $y \notin V$. Write $y = y_V \oplus y_{V^\perp}$ as the orthogonal decomposition of y. One verifies that for any unit vector $v \in V^\perp$ (for instance, $v = \pm y_{V^\perp}/\|y_{V^\perp}\|$), both y as well as the vector

$$y_V \oplus \|y_{V^\perp}\| v$$

have the same distances from every vector in V – in particular, from each of $0, y_1, \dots, y_n$. This shows (the contrapositive of) one implication.

Conversely, suppose $y \in V$. We show that y is uniquely determined by the distances to the y_j and to 0 – in fact, it suffices to consider the distances to a basis of V. Thus, suppose without loss of generality that the y_j are linearly independent. Let $y := \sum_{j=1}^n c_j y_j$, and let $d_0 := \|y\|, d_j := \|y - y_j\|$. We show that the d_j uniquely determine the c_j, and hence y. Indeed, a straightforward computation yields

$$d_0^2 - d_j^2 = \left\| \sum_{k=1}^n c_k y_k \right\|^2 - \left\| \sum_{k=1}^n c_k y_k - y_j \right\|^2 = -\|y_j\|^2 + 2 \sum_{k=1}^n c_k \langle y_j, y_k \rangle, \qquad 1 \leq j \leq n.$$

Rewriting this system of linear equations (in $\mathbf{c} = (c_1, \dots, c_n)$) yields

$$\mathrm{Gram}((y_j)_j)\mathbf{c} = \frac{1}{2}(\|y\|^2 + \|y_j\|^2 - \|y - y_j\|^2)_{j=1}^n. \tag{27.11}$$

Hence, $\mathbf{c}$ is unique, by Lemma 27.4. $\square$

Equipped with these preliminaries, we are now ready to proceed toward proving Menger's result. We first show:

Proof of Theorem 27.3.

(1) First suppose that both M, N contain the origin, and $\Psi : M \to N$ sends 0 to itself. This is not really a constraint: if here we can show that $\Psi = T|_M$ for some orthogonal matrix $T \in O_k(\mathbb{R})$, then for a general isometry Ψ and an arbitrary (base)point $m_\circ \in M$, the isometry

$$\Phi : M - m_\circ \to \Psi(M) - \Psi(m_\circ), \qquad v \mapsto \Psi(m_\circ + v) - \Psi(m_\circ)$$

sends 0 to 0, hence equals the restriction to M of some $T \in O_k(\mathbb{R})$. Thus,

$$\Psi(m) = T(m) + (\Psi(m_\circ) - T(m_\circ)), \qquad \forall m \in M.$$

Therefore, we may assume without loss of generality that $m_\circ := 0 \in M \cap N$ and $\Psi(0) = 0$. In this case, we need to show that Ψ is the restriction to M of an orthogonal matrix.

Begin by isolating a basis $m_1, \dots, m_r \in M$ of the $\mathbb{R}$-span of M. We assume $r > 0$, else M is a singleton, hence so is N, and then the result is immediate. Let $y_j := \Psi(m_j)$; then

$$\begin{aligned} 2\langle y_j, y_k \rangle &= \|\Psi(m_j) - \Psi(0)\|^2 + \|\Psi(m_k) - \Psi(0)\|^2 - \|\Psi(m_j) - \Psi(m_k)\|^2 \\ &= \|m_j\|^2 + \|m_k\|^2 - \|m_j - m_k\|^2 = 2\langle m_j, m_k \rangle, \qquad \forall 1 \leq j,k \leq r. \end{aligned} \tag{27.12}$$

Hence, by Lemma 27.4, the y_j are also linearly independent. By the same reasoning, for any $m \in M \setminus \{m_1, \dots, m_r\}$ we have that the Gram matrix of $m_1, \dots, m_r, m$ is singular, hence so is the Gram matrix of $y_1, \dots, y_r, \Psi(m)$. Again, using Lemma 27.4, the image $\Psi(M) = N$ is contained in the $\mathbb{R}$-span of $\{y_j = \Psi(m_j) : 1 \leq j \leq r\}$.

At this point, we define the linear map

$$T : \operatorname{span}_{\mathbb{R}}(M) \to \operatorname{span}_{\mathbb{R}}(N), \qquad m_j \mapsto y_j \ \forall 1 \leq j \leq r.$$

The next claim is that $T \equiv \Psi$ on M. Indeed, given $m \in M$, write

$$m = \sum_{j=1}^r c_j(m) m_j, \qquad \Psi(m) = \sum_{j=1}^r c'_j(m) y_j.$$

We now apply Proposition 27.10 to both m and $\Psi(m)$. Since $\operatorname{Gram}((m_j)_j) = \operatorname{Gram}((y_j)_j)$ by (27.12), the computation in Equation (27.11) for both y_j and m_j reveals that $c_j(m) = c'_j(m)$ for all $m \in M$, $1 \leq j \leq r$. In particular,

$$T(m) = \sum_{j=1}^r c_j(m) T(m_j) = \sum_{j=1}^r c'_j(m) y_j = \Psi(m), \qquad \forall m \in M$$

which proves the claim. Finally, we assert that T preserves lengths on $\operatorname{span}_{\mathbb{R}}(M)$. But this is clear by (27.12): if $v := \sum_j c_j m_j$ with all $c_j \in \mathbb{R}$, then

$$\langle Tv, Tv \rangle = \sum_{j,k=1}^r c_j c_k \langle y_j, y_k \rangle = \sum_{j,k=1}^r c_j c_k \langle m_j m_k \rangle = \langle v, v \rangle.$$

To conclude the proof of this part, choose orthonormal bases of the ortho-complements in $\mathbb{R}^k$ of $\operatorname{span}_{\mathbb{R}}(M)$ and $\operatorname{span}_{\mathbb{R}}(N)$, and map one basis to another to extend T to an orthogonal linear map on all of $\mathbb{R}^k$.

(2) This is clear by the previous part: given $x_j, y_j \in \ell^2$ with $1 \leq j \leq n$, such that

$$\|x_j - x_k\| = \|y_j - y_k\|, \qquad \forall 1 \leq j,k \leq n,$$

choose a finite-dimensional subspace of ℓ^2 which contains all x_j, y_j. Apply the previous part to this subspace; modulo the translation, one has an orthogonal transformation of this subspace, which we augment by the identity map on its ortho-complement. □

Remark 27.13. The proof of Theorem 27.3 is reminiscent of the well-known "lurking isometry" method – so named by J. Ball – in bounded analytic interpolation. This involves using Hilbert space realizations, and has numerous applications, including to the problems of Pick–Nevanlinna and Carathéodory–Fejer, among others (see, e.g., [2, 3]), and also indirectly in H_∞ methods in control theory (see [16] and the references therein).

Finally, we use Theorem 27.3 to prove Menger's result:

Proof of Theorem 27.1. The "only if" part is immediate, modulo a translation in order to map one of the points to the origin. Conversely, if X is finite, then the result is again easy. Thus, we now assume that X is both infinite and separable. Let $D := \{x_n : n \geq 0\}$ denote a countably infinite dense subset of X and define $D_n := \{x_0, \dots, x_n\}$ for $n \geq 0$. We are given isometric embeddings $\Psi_n : D_n \hookrightarrow \ell^2$ for each $n \geq 2$, where we assume without loss of generality that $\Psi_n(x_0) = 0 \; \forall n \geq 0$. We now construct an isometric embedding $: D \hookrightarrow \ell^2$, once again sending x_0 to 0.

To do so, fix and start at any integer $n_0 \geq 2$, say. Given $n \geq n_0$, we have $\Psi_n : D_n \hookrightarrow \ell^2$ (sending x_0 to 0). Now

$$\Psi_n \circ \Psi_{n+1}^{-1} : \Psi_{n+1}(D_n) \to \Psi_n(D_n)$$

is an isometry of an $(n+1)$-point set in ℓ^2, sending 0 to 0. Extend this to an orthogonal linear transformation on ℓ^2 by (the proof of) Theorem 27.3, say T_{n+1}. Thus, we have "increased" $\Psi_n(D_n)$ to an isometric image of D_{n+1}, namely $T_{n+1}(\Psi_{n+1}(D_{n+1}))$, while not changing the images of $x_0, x_1, \dots, x_n$.

We now repeatedly compose the T_n, to obtain the increasing family of sets

$$S_n := (T_{n_0+1} \circ \cdots \circ T_{n-1} \circ T_n)(\Psi_n(D_n)), \quad n \geq n_0$$

which satisfy

$$0 \in S_{n_0} = \Psi_{n_0}(D_{n_0}) \subset S_{n_0+1} \subset S_{n_0+2} \subset \cdots .$$

Moreover, each $S_n = \{y_0 = 0, y_1, \dots, y_n\}$ for $n \geq n_0$, together with an isometry $: D_n \to S_n$, sending $x_j \mapsto y_j$ for $0 \leq j \leq n$. The union of these sets provides the desired isometric embedding $\Phi : D \hookrightarrow \bigcup_n S_n = \lim_{n \to \infty} S_n$.

The final step is to apply the following standard fact from analysis, with $Y = \ell^2$:

Suppose $(X, d), (Y, d')$ are metric spaces, with Y complete. If $D \subset X$ is dense, any isometric embedding $\Phi : D \to Y$ extends uniquely to an isometric embedding $\widetilde{\Phi} : X \to Y$. □

27.2. **Cayley–Menger determinants, simplex volumes, and Heron's formula.** It is impossible to discuss Cayley–Menger matrices $CM(X)$ without explaining their true content: their connection to the squared volume of the simplex with vertices the elements of X.

Theorem 27.14. *Suppose $n \geq 1$ and $X = \{x_0, \dots, x_n\} \subset \mathbb{R}^n$. Then the volume $V_n(X)$ of the $(n+1)$-simplex with vertices x_j satisfies*

$$V_n(X)^2 = \frac{(-1)^{n+1} \det CM(X)}{2^n (n!)^2} = \frac{\det CM'(X)}{2^n (n!)^2}.$$

As a special case, if the points x_j are affine linearly dependent, then the volume of the corresponding simplex is zero, as is the determinant by Cayley's proposition 27.10.

Corollary 27.15. *For all finite subsets X of Euclidean (or Hilbert) space,* $\det CM'(X) \geq 0$.

Remarkably, Theorem 27.14 can be proved using only determinants (and a bit of visual geometry). Variants of the following proof can be found in several sources, including online.

Proof of Theorem 27.14. This proof is split into two parts. In the **first part**, we show the "usual/Cartesian" description of the volume of the simplex via determinants. Recall that the n-volume of a simplex in $\mathbb{R}^n$ having $n+1$ vertices is obtained inductively, by integrating the area of cross section as one goes from the base (which is a simplex in $\mathbb{R}^{n-1}$ with n vertices) to the apex/remaining vertex along an "altitude" of height h_n. An easy undergraduate calculus exercise reveals that if the base has $(n-1)$-volume V_{n-1}, then

$$V_n = \frac{h_n V_{n-1}}{n}.$$

One can now proceed inductively. Thus, let h_1 denote the length $\|x_0 - x_1\|$, let h_2 denote the "height" of x_2 "above" the segment joining x_0, x_1 (so it can be written as the norm of a suitable orthogonal complement), and so on. Then,

$$V_n(X) = \frac{1}{n!} h_n h_{n-1} \cdots h_1.$$

We now show that this product expression equals (up to sign) a determinant. Write

$$x_j := (x_j^{(1)}, \dots, x_j^{(n)})^T \in \mathbb{R}^n, \qquad 0 \leq j \leq n$$

and **claim** that up to a sign,

$$\frac{1}{n!} h_n h_{n-1} \cdots h_1 = \pm \frac{1}{n!} \det(A), \quad \text{where} \quad A := (x_j^{(k)} - x_0^{(k)})_{j,k=1}^n. \tag{27.16}$$

To show the claim, note that working with A essentially amounts to assuming $x_0 = 0$. Choosing a suitable orthonormal basis (i.e., by applying a suitable orthogonal transformation), we may further assume that $x_1, \dots, x_{n-1} \in \mathbb{R}^{n-1}$ – i.e., the final column of A has all entries zero except at most the (n,n) entry. Now the final row of A, which denotes the vector $x_n - x_0$, may be replaced by its orthogonal complement to the span of $\{x_j - x_0 : j < n\}$ without changing the determinant, and so we obtain (up to a sign) the height h_n – and in the nth coordinate since $x_j - x_0 \in \mathbb{R}^{n-1}$ for $j < n$. This scalar can be taken out of the determinant and we are now left with the determinant of an $(n-1) \times (n-1)$ matrix.

Applying the same arguments for $x_j - x_0$ with $j \leq n-2$ now, we obtain h_{n-1}, and so on. Proceeding by downward induction (and taking the absolute value), we obtain (27.16).

The **second step** is now easy: square the identity in the first step, to get:

$$V_n(X)^2 = \frac{1}{(n!)^2} (\det A)(\det A^T) = \frac{\det AA^T}{(n!)^2}.$$

But AA^T is precisely the Gram matrix of the vectors $\{x_j - x_0 : 1 \leq j \leq n\}$ – which has (j,k)-entry

$$\langle x_j - x_0, x_k - x_0 \rangle = \frac{1}{2}(d_{j0}^2 + d_{k0}^2 - d_{jk}^2)$$

as in Equation (16.12). Hence $AA^T = 2^{-1} CM'(X)$, so that

$$V_n(x)^2 = \frac{\det CM'(X)}{2^n (n!)^2}.$$

The proof is complete by Equation (27.9). $\square$

As a special case, this result leads to a well-known formula from two thousand years ago:

Corollary 27.17 (Heron's formula)**.** *A (Euclidean) triangle with edge lengths a, b, c and semi-perimeter $s = \frac{1}{2}(a+b+c)$ has area $\sqrt{s(s-a)(s-b)(s-c)}$.*

Proof. Explicitly expand the determinant in Theorem 27.14 for $n = 2$, to obtain:

$$V_2^2 = \frac{-1}{16} \det \begin{pmatrix} 0 & a^2 & b^2 & 1 \\ a^2 & 0 & c^2 & 1 \\ b^2 & c^2 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} = \frac{-1}{16} \left(-(a+b+c)(a+b-c)(b+c-a)(c+a-b) \right),$$

and this is precisely $s(s-a)(s-b)(s-c)$. $\square$

27.3. **Complements: completely monotone functions and distance transforms.** In parallel to the use of absolutely monotone functions earlier in this text – to characterize positivity preservers of kernels on infinite domains (or all finite domains) – we present here a related result by Ressel that features completely monotone functions. We then provide a sampling of early results in metric geometry, again by Schoenberg, that feature such functions.

Definition 27.18. A function $f : (0, \infty) \to \mathbb{R}$ is *completely monotone* if f is smooth and $(-1)^k f^{(k)}$ is non-negative on $(0, \infty)$ for all $k \geq 0$. A continuous function $f : [0, \infty) \to \mathbb{R}$ is completely monotone if the restriction of f to $(0, \infty)$ is completely monotone.

For instance, $x^{-\alpha}$ for $\alpha \leq 0$ is completely monotone on $(0, \infty)$.

We start with two results which are easily reformulated in the language of kernels:

(1) In his 1974 paper [297], Ressel characterized the functions that are positive definite in a different sense: given an abelian semigroup $(S, +)$, a function $f : S \to \mathbb{R}$ is said to be *positive semidefinite* if f is bounded and for any finite set of elements $s_1, \dots, s_n \in S$, the matrix $(f(s_j + s_k))_{j,k=1}^n$ is positive semidefinite. Ressel then showed for all $p \geq 1$ that the continuous and positive semidefinite functions on the semigroup $[0, \infty)^p$ are precisely Laplace transforms of finite non-negative Borel measures on $[0, \infty)^p$. In particular, for $p = 1$, this is further equivalent – by a result attributed to Bernstein, Hausdorff, and Widder – to f being completely monotone on $[0, \infty)$.

(2) A related result to this was shown by Schoenberg [316] in *Ann. of Math.* (1938). It says that a continuous function $f : [0, \infty) \to \mathbb{R}$ satisfies the property that for all integers $m, n \geq 1$ and vectors $x_1, \dots, x_m \in \mathbb{R}^n$, the matrix $(f(\|x_j - x_k\|^2))_{j,k=1}^m$ is positive semidefinite, if and only if f is completely monotone – i.e., as mentioned in the previous part, there exists a finite non-negative measure μ on $[0, \infty)$, such that

$$f(x) = \int_0^\infty \exp(-xt)\, d\mu(t), \qquad \forall x \geq 0.$$

Completely monotone functions also feature in the study of metric "endomorphisms" of Euclidean spaces. For instance, Schoenberg proved (in the aforementioned 1938 paper):

Theorem 27.19. *Given a continuous map $f : [0, \infty) \to [0, \infty)$, the following are equivalent:*

(1) For all integers $m, n \geq 1$ and vectors $x_1, \dots, x_m \in \mathbb{R}^n$, the matrix $(f(\|x_j - x_k\|))_{j,k=1}^m$ is Euclidean – i.e., $\{x_j\}$ with the metric $f \circ \|\cdot\|$ isometrically embed into ℓ^2.

(2) We have $f(0) = 0$ and the function $\frac{d}{dx}(f(\sqrt{x})^2)$ is completely monotone on $(0, \infty)$.

This result and paper are part of Schoenberg's program [313, 316, 317, 318, 355] to understand the transforms taking distance matrices from Euclidean space E_n of one dimension n, *isometrically* to those from another, say E_m. Schoenberg denoted this problem by $\{E_n; E_m\}$, with $1 \leq m, n \leq \infty$, where $E_\infty \cong \ell^2$ is Hilbert space. Schoenberg showed:

(1) If $n > m$, then $\{E_n; E_m\}$ is given by only the trivial function $f(t) \equiv 0$. Indeed, first observe by induction on n that the only Euclidean configuration of $n+1$ points that are equidistant from one another is an "equilateral" $(n+1)$-simplex Δ in $\mathbb{R}^n$ (or E_n), hence in any higher-dimensional Euclidean space – and this cannot exist in $\mathbb{R}^{n-1}$, hence not in $\mathbb{R}^m$ for $m < n$. If now $f(x_0) \neq 0$ for some $x_0 > 0$, then applying f to the distance matrix between vertices of the rescaled simplex $x_0\Delta \subset \mathbb{R}^n$, produces $n+1$ equidistant points in $\mathbb{R}^m$, which is not possible if $1 \leq m < n$.

(2) If $2 \leq n \leq m < \infty$, then $\{E_n; E_m\}$ consists only of the homotheties $f(x) = cx$ for some $c \geq 0$. (With von Neumann in 1941 [355], Schoenberg then extended this to answer the question for $n = 1 \leq m \leq \infty$.) Schoenberg also provided answers for $\{E_2; E_\infty\}$.

(3) The solution to the problem $\{E_\infty; E_\infty\}$ is precisely the content of Theorem 27.19.

As a special case of Theorem 27.19, all powers $\delta \in (0,1)$ of the Euclidean metric embed into Euclidean space. We provide an alternate proof using the above results on metric geometry.

Corollary 27.20 (Schoenberg, [315, 316])**.** *Hilbert space ℓ^2, with the metric $\|x-y\|^\delta$, embeds isometrically in "usual" ℓ^2 for any $\delta \in (0,1)$.*

This was shown in 1936 by Blumenthal in *Duke Math. J.* [52] for four-point subsets of ℓ^2 and $\delta \in (0, 1/2)$. Schoenberg extended this to all finite sets.

Proof. As observed by Schoenberg in [315], it suffices to show the result for $(n+1)$-element subsets $\{x_0, \dots, x_n\} \subset \ell^2$, by Menger's theorem 27.1. Now note that for $c > 0$, the function $g(u) := (1-e^{-cu})/u$ is bounded and continuous on $(0,1]$, hence admits a continuous extension to $[0,1]$. Since $u^{-\delta}$ is integrable in $(0,1]$, so is the product

$$\varphi : (0,1] \to \mathbb{R}, \qquad u \mapsto u^{-1-\delta}(1-e^{-cu}).$$

Clearly, $\varphi : [1,\infty) \to \mathbb{R}$ is also integrable, being continuous, non-negative, and bounded above by $u^{-1-\delta}$. By changing variables, we obtain a normalization constant $c_\delta > 0$, such that

$$t^{2\delta} = c_\delta \int_0^\infty (1 - e^{-\lambda^2 t^2}) \lambda^{-1-2\delta} \, d\lambda, \qquad \forall t > 0.$$

Set $t := \|x_j - x_k\|$, and let $\mathbf{u} = (u_0, \dots, u_n)^T \in \mathbb{R}^{n+1}$ with $\sum_j u_j = 0$. Then,

$$\sum_{j,k=0}^n u_j u_k \|x_j - x_k\|^{2\delta} = c_\delta \int_0^\infty \left(\sum_{j,k=0}^n u_j u_k (1 - e^{-\lambda^2 \|x_j - x_k\|^2}) \right) \lambda^{-1-2\delta} \, d\lambda.$$

But the double-sum inside the integrand equals $-\sum_{j,k} u_j u_k e^{-\lambda^2\|x_j - x_k\|^2}$, and this is non-positive by Theorem 16.17. It follows that the matrix $(-(\|x_j - x_k\|^\delta)^2)_{j,k=0}^n$ is conditionally positive semidefinite, and so we are done by Theorem 16.10 and Lemma 16.13. $\square$

Bibliographic notes and references

Lemma 16.1 is taken from the well-known monograph [289] by Pólya and Szegő. Theorems 16.2, 16.3, 16.4, 16.5, 16.6, and 16.7, classifying the dimension-free entrywise positivity preservers on various domains, are due to Schoenberg [318], Rudin [306], Vasudeva [353], Herz [170], and Christensen and Ressel [82, 83], respectively. The theorems of Schoenberg and Rudin and Vasudeva were recently shown by Belton–Guillot–Khare–Putinar in [30] using significantly smaller test sets than all positive semidefinite matrices of all sizes; proving these results is the main focus of this part. These results turn out to be further useful in Part 4, in fully classifying the total positivity preservers on bi-infinite domains.

Loewner's theorem 16.8 on operator/matrix monotone functions is from [240]; see also Donoghue's book [103], as well as the recent monograph by Simon [338] that contains a dozen different proofs.

The results in distance geometry are but a sampling from the numerous works of Schoenberg. Theorem 16.10, relating Euclidean embedding of a metric and the conditional negativity of the corresponding squared-distance matrix, is from [313], following then-recent works by Menger [258, 259, 260] and Fréchet [131]. Schoenberg's theorem 16.17 (respectively, Proposition 16.19), characterizing Hilbert space (respectively, the Hilbert sphere) in terms of positive definiteness of the Gaussian family (respectively, the cosine), is from [317] (respectively, from [313]). We point out that Schoenberg proved these results more generally for separable (not just finite) metric spaces, as discussed in Section 27. See also Schoenberg's paper [316] and another with von Neumann [355] (and its related work [223] by Kolmogorov). For the works of Bochner in this context, we restrict ourselves to mentioning [56, 57]. Theorems 16.20 and 16.24 by Schoenberg on positive definite functions on spheres are from [318]. Also, see the survey [346] of positive definite functions by James Drewry Stewart (who is perhaps somewhat better known for his series of calculus textbooks).

While Schoenberg's motivations in arriving at his theorem lay in metric geometry, as described above, Rudin's motivations were from Fourier analysis. More precisely, Rudin was studying functions operating on spaces of Fourier transforms of L^1 functions on groups G, or of measures on G. Here, G is a locally compact abelian group equipped with its Haar measure; Rudin worked with the torus $G = \mathbb{T}$, while Kahane and Katznelson worked with its dual group $\mathbb{Z}$. These authors together with Helson proved [164] a remarkable result in a converse direction to Wiener–Levy theory, in *Acta Math.* 1959. That same year, Rudin showed Schoenberg's theorem without the continuity hypothesis, i.e., Theorem 16.3. For more details on this part, on the metric geometric motivations of Schoenberg, and other topics, the reader is referred to the detailed recent twin surveys of Belton–Guillot–Khare–Putinar [27, 28].

The Horn–Loewner Theorem 17.1 (in a special case) originally appeared in Horn's paper [182], where he attributes it to Loewner. The theorem has since been extended by the author (jointly) in various ways; see, e.g., [30, 155]; a common, overarching generalization of these and other variants has been achieved in [214]. Horn–Loewner's determinant calculation in Proposition 17.5 was also extended to Proposition 17.8 by Khare [214]. The second, direct proof of Theorem 17.1 is essentially due to Vasudeva [353].

Mollifiers were introduced by Friedrichs [132], following the famous paper of Sobolev [343], and their basic properties can be found in standard textbooks in analysis, as can Cauchy's mean value theorem for divided differences. The remainder of the proof of the stronger Horn–Loewner theorem is from [182], and the Boas–Widder theorem 18.10(2) is from [55].

Bernstein's theorem 19.3 is from his well-known memoir on absolutely monotone functions [44]; see also Widder's textbook [370]. Boas' theorem 19.9 on the analyticity of smooth functions with SSR (strictly sign regular) derivatives is from [54]. Hamburger's theorem 19.14 is a folklore result, found in standard reference books – see, e.g., [8, 310, 335]. The remainder of Section 19 is from Belton–Guillot–Khare–Putinar [30]. Section 20 is taken from the same paper, with the exception of Theorem 20.11 which is new, as are "Proofs 1 and 2" of the existence of a positivity certificate/limiting sum-of-squares representation for $(1 \pm t)(1 - t^2)^n$. The former proof cites a result by Berg–Christensen–Ressel [39], and the latter, direct proof, is new.

Sections 21 and 22 are again from [30], save for the standard Identity Theorem 21.4, and Proposition 21.2 on the closure of real analytic functions under composition; these can be found in, e.g., [227]. The complex analysis basics, including Montel and Morera's theorems, can be found in standard textbooks; we cite [88]. The multivariate Schoenberg–Rudin theorem was proved by FitzGerald, Micchelli, and Pinkus in [124], and subsequently, under significantly weaker hypotheses in [30].

Appendix A on the Boas–Widder theorem is from [55] except for the initial observations. Boas and Widder mention Popoviciu [290] had proved the same result previously, using unequally spaced difference operators. The very last "calculus" Proposition 23.12 can be found in standard textbooks. Appendix B, classifying the dimension-free preservers of positivity when not acting on diagonal blocks, is from the recent work of Vishwakarma [354], with the exception of the textbook Proposition 24.4, and Theorem 24.1 by Guillot–Rajaratnam [157].

Theorems 26.1 and 26.2 in Appendix D, understanding and relating the Loewner positive, monotone, and convex maps, were originally proved without rank constraints on the test sets, by Hiai in [171]. Lemma 26.3 is partly taken from [171] and partly from Rockafellar's book – see [302, Theorems 24.1 and 25.3]. The remainder of Section 26 (i.e., Appendix D), as well as Section 25 (i.e., Appendix C), are taken from Belton–Guillot–Khare–Putinar [32].

Theorem 27.1 is due to Menger [258]. Theorem 27.2 comes from various works of Schoenberg on metric geometry (cited on the preceding page). Theorem 27.3 was already known to experts at the time; we cite here Birkhoff's famous paper [50]. The first part of Proposition 27.10 was shown by Cayley [78], and features Cayley–Menger determinants. The proof of Theorem 27.14 can be found in numerous sources, including online. The results in Section 27.3 are taken from the sources mentioned in it (e.g., the proof of Corollary 27.20).

Part 4:

Pólya frequency functions and sequences

Part 4: Pólya frequency functions and sequences

28. Totally non-negative functions of finite order.

In this part and the next, we approach the preserver problem in the above settings through a more classical viewpoint: that of spaces of *kernels* and their endomorphisms. The material in this part of the text is drawn primarily from the work of Schoenberg and his coauthors (as well as its account in Karlin's monograph), and a few recent preprints, all from 2020.

We begin by describing this part of the text and the next. We are interested in characterizing the preservers of totally non-negative or totally positive kernels, on $X \times Y$ for *arbitrary* totally ordered sets X, Y. The case of X, Y finite was studied in Part 2, and if $|X|, |Y| \geq 2$, the only such functions – up to rescaling – are powers. Next, if exactly one of X, Y is finite, a workaround can be achieved by using a generalization of Whitney's density theorem 6.7; we show this in Section 40. The difficulty lies in the remaining case: classifying the total positivity preservers for kernels on $X \times Y$ when both X, Y are infinite. In this case, the families of kernels we have encountered so far, do not suffice to yield a complete classification.

Thus, we begin by relaxing our goal, to classifying such preservers for *structured* kernels, on *specific* domains. Our first goal is to study the inner transforms of TN/TP Toeplitz kernels on $\mathbb{R} \times \mathbb{R}$. This is not an arbitrary choice: indeed, such kernels have long been studied in the analysis literature, under the name of *Pólya frequency functions*, and so it is natural to study this test set – as well as the related *Pólya frequency sequences* – and to understand the endomorphisms of these classes. This understanding is achieved in the next part of the text.

The class of Pólya frequency functions is fundamental to time-frequency analysis and to interpolation theory, the latter via splines (a subject which begins with many papers by Schoenberg and his coauthors). Pólya frequency functions possess beautiful properties that were established by Schoenberg and others in the twentieth century, and that allow us to exploit tools from harmonic analysis to try and classify their preservers. Looking ahead (and using these tools), we will find that Toeplitz TN kernels turn out to be quite rigid, and the results in the next few sections will help resolve – again in the next part – the original problem, of classifying the TN/TP kernel preservers on arbitrary domains. (See Section 40.)

Thus, a **roadmap** of this part and the next: We begin by discussing some preliminaries on Pólya frequency (PF) functions and sequences, including the variation diminishing property and its history. This is followed by a selection of results from the landmark paper of Schoenberg in *J. d'Analyse Math.* (1951), which establishes a host of properties of PF functions and surveys the development of the subject until that point. (Following Schoenberg's papers, Karlin's book also develops the theory comprehensively.) We also discuss several examples of PF functions and sequences. We next discuss several classical results on root-location, the Laguerre–Pólya class of entire functions, and its connection to both Pólya frequency functions as well as the Pólya–Schur theory of multiplier sequences (and some well-known modern achievements). Finally, we discuss very recent results (2020) on TN_p functions. This part can be read from scratch, requiring only Sections 6 and 12.1 and Lemma 26.3 when invoked.

In the next part: in Section 36 we will see results of Schoenberg and Karlin (and a converse to the latter) which reveal a 'critical exponent' phenomenon in total positivity, akin to Section 9. In Sections 37 and 38, we prove a host of classification results on preservers: of Pólya frequency functions and sequences, one-sided variants, and other structured Toeplitz kernels on various sub-domains of $\mathbb{R}$. In Section 39, we classify the preservers of TN and TP Hankel kernels. Finally, these results all come together in Section 40, along with discretization techniques and set-theoretic embedding arguments, to solve the overarching problem of classifying the preservers of totally positive kernels on all totally ordered domains.

28.1. TN_2 **functions – basic properties.** We begin with notation.

Definition 28.1. Given an integer $p \geq 1$, we term a function $f : \mathbb{R} \to \mathbb{R}$ *totally non-negative of order* p, or TN_p, if f is Lebesgue measurable and the associated Toeplitz kernel

$$T_f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}, \qquad (x, y) \mapsto f(x - y) \tag{28.2}$$

is TN_p (see Definition 5.10). We will say f is *totally non-negative*, or TN, if f is Lebesgue measurable and T_f is TN. This definition extends to $T_f : X \times Y \to \mathbb{R}$ (where $X, Y \subset \mathbb{R}$) if f is only defined on the Minkowski difference $X - Y \subset \mathbb{R}$.

We discuss examples in the next section; first, we explore basic properties of these functions.

Lemma 28.3. *Suppose $p \geq 1$ is an integer, and $f : \mathbb{R} \to \mathbb{R}$ is a TN_p function. Then $cf(ax + b)$ and $ce^{ax} f(x)$ are also TN_p functions, for any $a, b \in \mathbb{R}$ and $c \geq 0$.*

Proof. The first part is left as an exercise to the reader, noting if $a < 0$ that (the sign of) a determinant remains unchanged upon reversing all rows as well as columns. For the second part, if $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{r,\uparrow}$ for $1 \leq r \leq p$ (see Definition 25.1), and $g(x) := ce^{ax} f(x)$, then

$$T_g[\mathbf{x}; \mathbf{y}] = \operatorname{diag}(ce^{ax_j})_{j=1}^r T_f[\mathbf{x}; \mathbf{y}] \operatorname{diag}(e^{-ay_k})_{k=1}^r,$$

and hence $\det T_g[\mathbf{x}; \mathbf{y}] \geq 0$ by the hypotheses. $\square$

The following result reveals the nature of TN_2 functions over symmetric intervals:

Theorem 28.4. *Suppose an interval $J \subset \mathbb{R}$ contains the origin and has positive length, and $f : J - J \to \mathbb{R}$ is Lebesgue measurable. The following are equivalent:*

(1) *There exists an interval $I \subset J - J$ such that f is positive on I and vanishes outside I, and $\log(f)$ is concave on I.*

(2) *f is TN_2, i.e., given $a < b$ and $c < d$ in J, the matrix $\begin{pmatrix} f(a-c) & f(a-d) \\ f(b-c) & f(b-d) \end{pmatrix}$ is TN.*

If so, f is continuous in the interior of I, so discontinuous in $J - J$ at most at two points.

The proof uses a preliminary fact on concave functions:

Lemma 28.5. *Suppose $p < q$ are real numbers and $g : [p, q] \to \mathbb{R}$ is concave. If $p \leq r \leq s \leq q$ and $r + s = p + q$, then $g(r) + g(s) \geq g(p) + g(q)$.*

Proof. Suppose for some $\lambda \in [0, 1]$ that $r = \lambda p + (1 - \lambda)q$; then $s = (1 - \lambda)p + \lambda q$. Now concavity implies:

$$g(r) \geq \lambda g(p) + (1 - \lambda) g(q), \qquad g(s) \geq (1 - \lambda) g(p) + \lambda g(q).$$

Adding these inequalities, we are done. $\square$

Proof of Theorem 28.4. First if f is non-zero at most at one point then the result is clear, so assume throughout this proof that f is non-zero at least at two points. Suppose (1) holds, and $a < b$ and $c < d$ index the rows and columns of a '2×2 submatrix' drawn from the kernel T_f on $J \times J$. By the hypotheses, all entries of the matrix $M := \begin{pmatrix} f(a-c) & f(a-d) \\ f(b-c) & f(b-d) \end{pmatrix}$ are non-negative. Also,

$$a - c,\ b - d \in (a - d,\ b - c).$$

Now there are several cases. If either $a - c$ or $b - d$ lie outside I, then M has a zero row or zero column, whence it is TN_2. Otherwise $a - c, b - d \in I$. If now $a - d$ or $b - c$ lie outside I then M is a triangular matrix, hence TN_2. Else we may suppose all entries of M are positive.

Now the above ordering and the concavity of $\log(f)$ easily imply via Lemma 28.5 (applied to $g = \log(f), p = a - d, q = b - c$) that $\det(M) \geq 0$. This proves (1) $\implies$ (2).

Conversely, suppose (2) holds, so that $f \geq 0$ on $J - J$, and suppose $f(x_0) > 0$. First fix $\delta > 0$ such that either $[0, \delta)$ or $(-\delta, 0]$ is contained in J. We now claim that if f vanishes at $x_1 > x_0$ (respectively, $x_2 < x_0$), then it vanishes on the intersection of $J - J$ with $[x_1, \infty)$ (respectively, with $(-\infty, x_2]$). We only show this for $x_1 > x_0$; note moreover that it suffices to show this for $y \in (x_1, x_1 + \delta) \cap (J - J)$. If $(-\delta, 0] \subset J$, consider the TN_2 submatrix

$$T_f[(x_0, x_1); (x_1 - y, 0)] = \begin{pmatrix} f(x_0 - x_1 + y) & f(x_0) \\ f(y) & f(x_1) \end{pmatrix},$$

where we note that $x_0 - x_1 + y \in (x_0, y) \subset J - J$. If instead $[0, \delta) \subset J$, then

$$T_f[(x_0 - x_1 + y, y); (0, y - x_1)] = \begin{pmatrix} f(x_0 - x_1 + y) & f(x_0) \\ f(y) & f(x_1) \end{pmatrix}$$

yields the same TN_2 submatrix. In both cases, taking determinants gives: $-f(x_0)f(y) \geq 0$, and since $f(x_0) > 0$, the claim follows. This shows the existence of an interval $I \subset J - J$ of positive measure, such that $f > 0$ on I and $f \equiv 0$ outside I.

We next claim that $-\log(f)$ is mid-convex on I. Indeed, given $y - \epsilon < y < y + \epsilon$ in I, it suffices to show $f(y) \geq \sqrt{f(y+\epsilon)f(y-\epsilon)}$. The following argument owes its intuition to the theory of discrete-time Markov chains on a finite state space, but can also be made direct. Begin by defining $n_0 := 2\lceil \epsilon/\delta \rceil$; thus $\epsilon/n_0 \in (0, \delta)$. Set $z_k := f(y + k\epsilon/n_0)$ for $-n_0 \leq k \leq n_0$; since all arguments lie in I, $z_k > 0 \ \forall k$. There are now two cases: if $(-\delta, 0] \subset J$, then

$$0 \leq \det T_f[(y - (k+1)\epsilon/n_0, y - k\epsilon/n_0); (-\epsilon/n_0, 0)] = z_k^2 - z_{k-1}z_{k+1}, \qquad \forall - n_0 < k < n_0.$$

If instead $[0, \delta) \subset J$, use $0 \leq \det T_f[(y - k\epsilon/n_0, y - (k-1)\epsilon/n_0); (0, \epsilon/n_0)]$ for the same values of k. Thus, it follows that $z_k \geq \sqrt{z_{k-1}z_{k+1}}$ for $-n_0 < k < n_0$. Now one shows inductively:

$$z_0 \geq (z_1 z_{-1})^{1/2} \geq (z_2 z_0^2 z_{-2})^{1/4} \geq \cdots \geq \prod_{j=0}^{n_0} z_{2j-n_0}^{\binom{n_0}{j}/2^{n_0}} \geq \cdots$$

Note at each step that the powers of $z_{\pm n_0}$ are not touched, while all remaining terms z_k^γ are lower-bounded by $\sqrt{z_{k-1}z_{k+1}}^{\gamma/2}$. Now think of each step as a (positive) integer time $t > 0$, and consider the exponents at each step. These give a probability distribution π_t on the set $\mathcal{S} := \{-n_0, \dots, 0, 1, \dots, n_0\}$. This is a well-studied model in probability theory: the simple random walk with absorbing barriers $\pm n_0$, on the state space $\mathcal{S}$. In particular, this is the Markov chain called the *symmetric gambler's ruin*. Specifically, the transition probabilities are $\pm 1/2$ to go from a non-absorbing state $-n_0 < k < n_0$ are $\pm 1/2$ to $k \pm 1$. Denoting by π_t the probability distribution on $\mathcal{S}$ at each time point $t \in \mathbb{Z}^{\geq 0}$, we thus obtain:

$$z_0 \geq \prod_{j=-n_0}^{n_0} z_j^{\pi_t(j)}, \qquad t = 0, 1, 2, \dots.$$

Moreover, each π_t has equal mass at $\pm n_0$, so the same holds as $t \to \infty$. Now by Markov chain theory, the limiting probability distribution as $t \to \infty$ exists and has mass only at the absorbing states $\pm n_0$. As these masses must be equal, we obtain by translating back:

$$f(y) = z_0 \geq \sqrt{z_{-n_0} z_{n_0}} = \sqrt{f(y-\epsilon)f(y+\epsilon)}.$$

In this special case, the argument can be made direct as well; here is a sketch. Let $c_t := \sum_{j=-(n_0-1)}^{n_0-1} \pi_t(j)$ for $t = 1, 2, \dots$. If $-n_0 < j < n_0$, then after $2n_0 - 1$ steps, the power

$z_j^{\pi_t(j)}$ will 'contribute' at least $z_{-n_0}^{\pi_t(j)/2^{n_0+j}}$ or $z_{n_0}^{\pi_t(j)/2^{n_0-j}}$. From this it follows that

$$c_{t+2n_0-1} \leq c_t(1-2^{1-n_0}), \qquad \forall t \in \mathbb{Z}^{\geq 0}$$

via the AM–GM inequality. Choosing $t = m(2n_0 - 1)$ for $m = 0, 1, \dots,$ and recalling that $\pi_t(-n_0) = \pi_t(n_0)$ for all t, we have the desired conclusion as $m \to \infty$.

Thus $-\log(f)$ is mid-convex on I. Moreover, $-\log(f)$ is Lebesgue measurable on I by assumption, so Theorem 12.4 implies $-\log(f)$, hence f, is continuous on the interior of I. (This shows the final assertion.) In particular, a continuous mid-convex function is convex, so that $-\log(f)$ is convex in the interior of I. To show it is convex on I, it remains to show:

$$\log f(\lambda a + (1-\lambda)b) \geq \lambda \log f(a) + (1-\lambda)\log f(b), \qquad \forall \lambda \in (0,1),\ a, b \in I. \tag{28.6}$$

To show this, approximate λ by a sequence of dyadic rationals $\lambda_n \in (0,1)$, and note by mid-concavity/mid-convexity that

$$\log f(\lambda_n a + (1-\lambda_n)b) \geq \lambda_n \log f(a) + (1-\lambda_n)\log f(b), \qquad \forall n \geq 1.$$

Since $\lambda_n, \lambda \in (0,1)$, the arguments on the left are always in the interior of I, where f is positive and continuous. Thus, the preceding inequality shows (28.6) as $n \to \infty$. □

An immediate consequence of Theorem 28.4 is:

Corollary 28.7. *If $0 \in J \subset \mathbb{R}$ is as in Theorem 28.4, and functions $f, g : J - J \to \mathbb{R}$ are TN_2, then so are $f \cdot g$ and f^α for $\alpha \geq 0$.*

This follows from the fact that $\log(f) + \log(g), \alpha \log(f)$ are concave if $\log(f), \log(g)$ are.

We now revert to the 'classical' setting of TN_2 functions, i.e., $f : \mathbb{R} \to \mathbb{R}$. These functions decay exponentially at infinity, except for the exponentials themselves:

Proposition 28.8. *Suppose $f : \mathbb{R} \to \mathbb{R}$ is TN_2 (whence measurable), and has unbounded support. Then either $f(x) = e^{ax+b}$ for suitable scalars $a, b \in \mathbb{R}$, or there exists $\gamma \in \mathbb{R}$ such that the TN_2 function $e^{-\gamma x} f(x)$ tends to zero exponentially fast as $|x| \to \infty$, whence is integrable.*

Proof. Via Theorem 28.4, let I denote the largest interval on which f is positive, and set $g(x) := \log f(x)$. We show the result for $I = \mathbb{R}$; the proof is similar (but easier) for other I. First note that $-g(x)$ is convex on I. Now recall Lemma 26.3: g' exists on a co-countable, dense subset of I, and $g'_\pm$ exist and are non-increasing on the interior of I (whence have only jump discontinuities).

Suppose $g(x)$ is not linear (i.e., $f(x)$ is not of the form e^{ax+b}). Then g'_+ is not constant on I, so there exist points $x_1 < x_2$ in I such that $g'(x_1) > g'(x_2)$. It follows that there exist constants $c_1, c_2 \in \mathbb{R}$ satisfying:

$$\log f(x) \leq g'(x_j)x + c_j, \qquad j = 1, 2.$$

Choose $\gamma \in (g'(x_2), g'(x_1))$. Then,

$$\log f(x) - \gamma x \leq \begin{cases} (g'(x_1) - \gamma)x + c_1, & x \in I \cap (-\infty, 0), \\ (g'(x_2) - \gamma)x + c_2, & x \in I \cap (0, \infty). \end{cases}$$

Now exponentiating both sides gives the result. □

28.2. **Classification of TN_p functions for higher p.** The preceding subsection saw a characterization of TN_2 functions, by Schoenberg (1951). It has taken longer to obtain a characterization for TN_p functions for $p \geq 3$. (A 1983 result in *Appl. Anal.* of Weinberger [365] toward characterizing TN_3 functions has a small gap; see Section 35.) We conclude this section by showing such a characterization: to check if a non-negative function is TN_p (with some decay properties), it suffices to check the signs of all $p \times p$ minors, but no smaller ones. Remarkably, such a result was discovered only in 2020, leading to a characterization (by this author) in a subsequent preprint. We begin with the earlier result, for integrable functions:

Lemma 28.9 (Förster–Kieburg–Kösters, [128])**.** *Suppose $f : \mathbb{R} \to [0, \infty)$ is integrable and every $p \times p$ matrix $(f(x_j - y_k))_{j,k=1}^p$ has non-negative determinant, for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{p,\uparrow}$. Then f is TN_p.*

Notice that not every TN_p (or even TN) function is integrable – for instance, e^{ax+b}, which is a TN function by Lemma 28.3 since $f(x) \equiv 1$ is TN. Thus, Lemma 28.9 cannot characterize the TN_p functions. However, we shall presently see a characterization result along these lines, among other variants of Lemma 28.9. As a first variant, the result can be extended to hold for more general domains $X, Y \subset \mathbb{R}$, and for functions that merely decay, and at one of $\pm\infty$:

Proposition 28.10. *Fix scalars $t_*, \rho \in \mathbb{R}$ and a subset $Y \subset \mathbb{R}$ unbounded above. Suppose $X \subset \mathbb{R}$ contains $t_* + y$ for all $\rho < y \in Y$. Let $f : X - Y \to [0, \infty)$ be such that $f(t_*) > 0$ and*

$$\lim_{y \in Y,\ \rho < y \to \infty} f(x_0 - y) f(t_* + y - y_0) = 0, \qquad \forall x_0 \in X,\ y_0 \in Y.$$

If $\det T_f[\mathbf{x}; \mathbf{y}] \geq 0$ $\forall \mathbf{x} \in X^{p,\uparrow}, \mathbf{y} \in Y^{p,\uparrow}$, then the Toeplitz kernel T_f is TN_p on $X \times Y$.

Lemma 28.9 is the special case $X = Y = \mathbb{R}$, where $\rho \in \mathbb{R}$ is arbitrary and $t_* \in \mathbb{R}$ any value at which $f : \mathbb{R} \to \mathbb{R}$ is non-zero. (If no such t_* exists, i.e., $f \equiv 0$, the result is immediate.) Thus it applies to detect "Pólya frequency functions of order p" (i.e., integrable TN_p functions on $\mathbb{R}$). Proposition 28.10 is more general in two ways: first, it also specializes to other domains – for instance, $X = Y = \mathbb{Z}$, i.e., to detect "Pólya frequency sequences of order p" that vanish at $\pm\infty$ (with t_* an integer). More generally, one can specialize to $X = Y = G$, for any additive subgroup $G \subset (\mathbb{R}, +)$. Second, Proposition 28.10 can accommodate non-integrable functions such as $e^{x - |x|}$ (which is seen to be TN in (32.14) below).

Proof. It suffices to show that if $\det T_f[\mathbf{x}; \mathbf{y}] \geq 0$ for all $\mathbf{x} \in X^{p,\uparrow}$ and $\mathbf{y} \in Y^{p,\uparrow}$, then the same condition holds for all $\mathbf{x}' \in X^{p-1,\uparrow}$ and $\mathbf{y}' \in Y^{p-1,\uparrow}$. Thus, fix such $\mathbf{x}', \mathbf{y}'$. We are to show

$$\psi(x_p, y_p) := \det T_f[(\mathbf{x}', x_p); (\mathbf{y}', y_p)] \geq 0 \ \forall x_p > x_{p-1}, y_p > y_{p-1} \quad \implies \quad \det T_f[\mathbf{x}'; \mathbf{y}'] \geq 0.$$

To see why, first define the $(p-1) \times (p-1)$ matrix $A := T_f[\mathbf{x}'; \mathbf{y}']$ and $A^{(j,k)}$ to be the submatrix of A with the jth row and kth column removed. Note that the following maximum does not depend on x_p, y_p:

$$L := \max_{1 \leq j,k \leq p-1} |\det A^{(j,k)}| \geq 0.$$

Next, given $m \geq 1$, define $t_m \in Y$ such that $t_m > t_0 := \max\{x_{p-1} - t_*, y_{p-1}, \rho\}$ and

$$f(x_j - t_m) f(t_* + t_m - y_k) < 1/m, \qquad \forall 0 < j, k < p.$$

We now turn to the proof. Expand the determinant $\psi(x_p, y_p)$ along the last column, and apart from the cofactor for (p, p), expand every other cofactor along the last row. This yields:

$$\psi(x_p, y_p) = f(x_p - y_p) \det(A) + \sum_{j,k=1}^{p-1} (-1)^{j+k-1} f(x_j - y_p) f(x_p - y_k) \det A^{(j,k)}.$$

Given t_* as above, and $m \geq 1$, let

$$y_p^{(m)} := t_m \in Y, \qquad x_p^{(m)} := t_* + t_m \in X.$$

Then $x_p^{(m)} > x_{p-1}$ and $y_p^{(m)} > y_{p-1}$, and $\psi(x_p^{(m)}, y_p^{(m)}) \geq 0$ by the hypotheses, so from above,

$$f(t_*) \det(A) \geq \psi(x_p^{(m)}, y_p^{(m)}) - L \sum_{j,k=1}^{p-1} f(x_j - y_p^{(m)}) f(x_p^{(m)} - y_k) \geq -L \frac{(p-1)^2}{m}.$$

As this holds for all $m \geq 1$, we have $\det(A) = \det T_f[\mathbf{x}'; \mathbf{y}'] \geq 0$, as desired. $\square$

This result specializes to provide a characterization of TN_p kernels for arbitrary $p \geq 2$:

Corollary 28.11. *Given $f : \mathbb{R} \to [0, \infty)$ and an integer $p \geq 2$, the following are equivalent:*

(1) Either $f(x) = e^{ax+b}$ for $a, b \in \mathbb{R}$, or: (a) f is Lebesgue measurable; (b) for all $x_0, y_0 \in \mathbb{R}$, $f(x_0 - y)f(y - y_0) \to 0$ as $y \to \infty$; and (c) $\det T_f[\mathbf{x}; \mathbf{y}] \geq 0$ $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^{p,\uparrow}$.
(2) The function $f : \mathbb{R} \to \mathbb{R}$ is TN_p.

This result improves on Lemma 28.9 in that not every TN_p function is integrable or an exponential e^{ax+b}. For example, we will see in (32.14) below such a function, when $\alpha\beta < 0$.

Proof. The result is obvious for $f \equiv 0$ on $\mathbb{R}$, so we assume henceforth that this is not the case. That (1)(a)–(c) $\implies$ (2) is now immediate from Proposition 28.10, specialized to $X = Y = \mathbb{R}$ and arbitrary $\rho \in \mathbb{R}$. If instead $f(x) = e^{ax+b}$, then by Lemma 28.3 it is TN because the constant function $f \equiv 1$ is obviously TN.

Conversely, suppose (2) holds and $f(x)$ is not of the form e^{ax+b}. Now (1)(a) and (1)(c) are immediate; if the support of f is bounded, then (1)(b) is also immediate. Otherwise f is TN_2 with support $\mathbb{R}$, whence by the proof of Proposition 28.8, there exist $\beta, \gamma \in \mathbb{R}$ and $\delta > 0$ such that $e^{-\gamma x} f(x) < e^{\beta - \delta|x|}$. Now a straightforward computation shows (1)(b). $\square$

We conclude this section with another variant for arbitrary positive-valued kernels on $X \times Y$, for arbitrary $X, Y \subset \mathbb{R}$:

Proposition 28.12. *Fix nonempty subsets $X, Y \subset \mathbb{R}$, and suppose a kernel $K : X \times Y \to (0, \infty)$ satisfies one of the following four decay conditions:*

$$\begin{aligned}
\sup Y \notin Y, &\qquad \lim_{y \in Y,\ y \to (\sup Y)^-} K(x_0, y) = 0, \qquad \forall x_0 \in X,\\
\inf Y \notin Y, &\qquad \lim_{y \in Y,\ y \to (\inf Y)^+} K(x_0, y) = 0, \qquad \forall x_0 \in X,\\
\sup X \notin X, &\qquad \lim_{x \in X,\ x \to (\sup X)^-} K(x, y_0) = 0, \qquad \forall y_0 \in Y,\\
\inf X \notin X, &\qquad \lim_{x \in X,\ x \to (\inf X)^+} K(x, y_0) = 0, \qquad \forall y_0 \in Y.
\end{aligned}$$

Then the following are equivalent for an integer $p \geq 2$:

(1) Every $p \times p$ minor of K is non-negative.
(2) K is TN_p on $X \times Y$.

Proof. Clearly (2) $\implies$ (1); conversely, it suffices to show by induction that $\det K[\mathbf{x}'; \mathbf{y}'] \geq 0$ for $\mathbf{x}' \in X^{p-1,\uparrow}$ and $\mathbf{y}' \in Y^{p-1,\uparrow}$. We show this under the last of the four assumptions; the other cases are similar to the following proof and to the proof of Proposition 28.10. Thus, let

$$\mathbf{x}' = (x_2, \dots, x_p), \qquad \mathbf{y}' = (y_2, \dots, y_p)$$

and $y_1 < y_2, y_1 \in Y$ be fixed. Let $A := K[\mathbf{x}'; \mathbf{y}']$; thus the maximum

$$L := \max_{1 \leq j,k \leq p-1} |\det A^{(j,k)}| \geq 0.$$

Next, construct a sequence $x_1^{(m)} \in X$, $m \geq 1$ such that $x_1^{(m)} < x_2$ and

$$K(x_1^{(m)}, y_k) < 1/m, \ \forall k = 2, \dots, p.$$

Now compute as in the proof of Proposition 28.10:

$$\begin{aligned} K(x_1^{(m)}, y_1) \det(A) \geq & \ \det K[(x_1^{(m)}, \mathbf{x}'); (y_1, \mathbf{y}')] - L \sum_{j,k=2}^{p} K(x_j, y_1) K(x_1^{(m)}, y_k) \\ \geq & \ \det K[(x_1^{(m)}, \mathbf{x}'); (y_1, \mathbf{y}')] - \frac{L(p-1)}{m} \sum_{j=2}^{p} K(x_j, y_1), \end{aligned}$$

where we expand the determinant along the first row and column. Since the first term on the right is non-negative, and $K > 0$ on $X \times Y$, taking $m \to \infty$ shows $\det(A) \geq 0$, as desired. $\square$

29. Pólya frequency functions. The variation diminishing and sign non-reversal properties. Single-vector characterizations of TP and TN matrices.

In this section, we introduce and study the distinguished class of Pólya frequency functions, as well as their variation diminishing property (including a look at its early history, from Descartes to Motzkin). Both of these studies were carried out by Schoenberg in his landmark 1951 paper [324]. The latter property will require first studying the variation diminishing property of TN matrices – carried out even earlier, by Schoenberg in 1930 in *Math. Z.*

We begin by introducing the titular class of functions in this part of the text.

Definition 29.1. A function $\Lambda : \mathbb{R} \to \mathbb{R}$ is said to be a *Pólya frequency (PF) function* if Λ is an integrable TN function that does not vanish at least at two points.

Some historical remarks on terminology follow. The term *frequency function* traditionally meant being integrable, or (up to normalization) a density function. Pólya frequency functions were introduced by Schoenberg in his landmark 1951 paper in *J. d'Analyse Math.* His definition means that the class of 'Dirac' TN functions $\Lambda(x) = \mathbf{1}_{x=c}$ (see Example 29.6 below) are not Pólya frequency functions. In fact, Schoenberg also studied a wider class of TN functions in *loc. cit.*, again excluding the Dirac functions. Specifically, he worked with (what he called) *totally positive functions* – which are (measurable) TN functions that do not vanish at least at two points.

Remark 29.2. We also refrain here from discussing either Schoenberg's motivations or the prior results by Laguerre, Pólya, Schur, and Hamburger that led Schoenberg to developing the theory of PF functions. This discussion will take place in Sections 33 and 34.

We begin by specializing the results in Section 28 to Pólya frequency (PF) functions:

Proposition 29.3.

(1) *The results in Section 28.1 hold for all PF functions.*

(2) *The class of PF functions is closed under the change of variables $x \mapsto ax + b$ for $a \neq 0$, and under convolution.*

(3) *If f is a TN function that is non-vanishing at least at two points, and f is not of the form e^{ax+b} for $a, b \in \mathbb{R}$, then there exists $\gamma \in \mathbb{R}$ such that $e^{-\gamma x} f(x)$ is a PF function. In other words, there is a strict trichotomy for (measurable) TN functions $f : \mathbb{R} \to \mathbb{R}$:*
   - (a) *$f(x)$ is monotone, or equivalently, an exponential e^{ax+b} for some $a, b \in \mathbb{R}$.*
   - (b) *f is supported (and positive) at a single point in the line.*
   - (c) *Up to an exponential factor $e^{\gamma x}$ with $\gamma \in \mathbb{R}$, the function f is integrable, whence a Pólya frequency function – in fact, this latter decays exponentially as $|x| \to \infty$.*

The final trichotomy holds more generally for all TN_2 functions, by Proposition 28.8. (The final part of the second assertion is not involved, and follows from Corollary 32.9 below, using basic properties of convolution.)

We next provide examples. The results in the preceding section studied TN_p functions for $p \geq 2$. We now discuss several important examples of all of these: in fact, of Pólya frequency functions (so, TN_p for all $p \geq 1$). The first is the Gaussian family.

Example 29.4. For all $\sigma > 0$, the Gaussian function $G_\sigma(x) := e^{-\sigma x^2}$ is a PF function, as shown in Lemma 6.8. For future use, we record its Laplace transform (discussed later):

$$\mathcal{B}(G_\sigma)(s) = e^{s^2/4\sigma} \sqrt{\pi/\sigma}, \qquad s \in \mathbb{C}.$$

Example 29.5. The kernel $f(x) = e^{-e^x}$ is totally non-negative. Indeed, the 'submatrix'

$$T_f[\mathbf{x}; \mathbf{y}] = (\exp(\alpha_j \beta_k))_{j,k=1}^n, \qquad n \geq 1, \ \mathbf{x}, \mathbf{y} \in \mathbb{R}^{n,\uparrow}$$

is a generalized Vandermonde matrix, where both $\alpha_j = e^{x_j}, \beta_k = -e^{-y_k}$ form increasing sequences. While $e^{-e^x} \not\to 0$ as $x \to -\infty$, by Lemma 28.3 e^{x-e^x} is integrable, so a PF function.

The next example is an integrable TN function that is pathological in nature, so not a PF function:

Example 29.6. The Dirac function $f(x) = \mathbf{1}_{x=c}$ can be easily verified to be TN, for $c \in \mathbb{R}$.

29.1. **Variation diminishing property for TP and TN matrices.** A widely-used property exhibited by TP and TN matrices is the variation diminishing property (this phrase – or 'variationsvermindernd' in German – was coined by Pólya; see Section 29.3 for more on its history). We now prove this property; it will be useful in Section 30 in explaining the real-rootedness of generating functions of finite Pólya frequency sequences. To proceed, we require some notation.

Definition 29.7. Given a vector $x \in \mathbb{R}^m$, let $S^-(x)$ denote the number of changes in sign, after removing all zero entries in x. Next, assign to each zero coordinate of x a value of ± 1, and let $S^+(x)$ denote the largest possible number of sign changes in the resulting sequence (running over all assignments of ± 1). We also set $S^-(\mathbf{0}) := 0$ and $S^+(\mathbf{0}) := m$, for $\mathbf{0} \in \mathbb{R}^m$.

For instance, $S^-(1,0,0,-1) = 1$ and $S^+(1,0,0,-1) = 3$. In general, it is easy to see that $S^+(x), S^-(x)$ are the largest and smallest number of sign changes possible, when one keeps the non-zero coordinates of x unchanged and modifies the zero coordinates.

We begin by characterizing TP matrices in terms of the variation diminishing property and an additional property.

Theorem 29.8. *Suppose $m, n \geq 1$ are integers, and $A \in \mathbb{R}^{n \times m}$. The following are equivalent:*

(1) A is totally positive.

(2) For all $x \in \mathbb{R}^m \setminus \{0\}$, $S^+(Ax) \leq S^-(x)$. Moreover, if equality holds and $Ax \neq 0$, then the first (last) component of Ax has the same sign as the first (last) non-zero component of x. (If either component in Ax is zero, we replace it by the sign of the changed component in computing S^+.)

This result can be found in Pinkus's book, and follows prior work by Gantmacher–Krein and also the 1981 paper in [*J. Amer. Statist. Assoc.*] by Brown–Johnstone–MacGibbon. Very recently, Choudhury has refined this result to require only a finite set of test vectors – exactly one vector x^B for every contiguous square submatrix B of A. Moreover, $x^B = \operatorname{adj}(B) v^B$ (with $\operatorname{adj}(B)$ the adjugate matrix of B), where v^B can be chosen to be *any* non-zero vector with alternating signs, belonging to a closed orthant:

Theorem 29.9 (Choudhury, [80]). *The assertions in Theorem 29.8 are equivalent to:*

(3) For all integers $1 \leq r \leq \min(m,n)$ and contiguous $r \times r$ submatrices B of A, and given any fixed vector $0 \neq v^B := (\alpha_1, -\alpha_2, \ldots, (-1)^{r-1}\alpha_r)^T$ with all $\alpha_j \geq 0$, we have $S^+(Bx^B) \leq S^-(x^B)$, where $x^B := \operatorname{adj}(B) v^B$. If equality occurs here, then the first (last) component of Bx^B has the same sign as the first (last) non-zero component of x^B. (If either component in Bx^B is zero, we replace it by the sign of the changed component in computing S^+.)

Proof. For this proof, define $\mathbf{d}_p := (1, -1, \dots, (-1)^{p-1})^T \in \mathbb{R}^p$, for any integer $p \geq 1$.

We begin by showing (1) $\implies$ (2). We first take a TP matrix A and draw some conclusions. Let $0 \neq x \in \mathbb{R}^m$ with $S^-(x) = p \leq m-1$. Thus there exists $\mathbf{s} \in [m]^{p,\uparrow}$ such that

$$(x_1, \dots, x_{s_1}),\ (x_{s_1+1}, \dots, x_{s_2}),\ \dots, (x_{s_p+1}, \dots, x_m) \tag{29.10}$$

are components of x^T, with all non-zero coordinates in the tth component having the same sign, which we choose up to 'orientation' to be $(-1)^{t-1}$. Here we also have that not all coordinates in each sub-tuple $(x_{s_{t-1}+1}, \dots, x_{s_t})$ are zero. Moreover, we set $s_0 := 0$ and $s_{p+1} = m$ for convenience.

Denote the columns of A by $\mathbf{c}_1, \dots, \mathbf{c}_m \in \mathbb{R}^n$, let

$$\mathbf{y}_t := \sum_{i=s_{t-1}+1}^{s_t} |x_i| \mathbf{c}_i \in \mathbb{R}^n, \qquad t = 1, \dots, p+1,$$

and let $Y_{n \times (p+1)} := [\mathbf{y}_1 | \cdots | \mathbf{y}_{p+1}]$.

The first claim is that Y is TP. Indeed, given $1 \leq r \leq \min(n, p+1)$ and subsets $J \in [n]^{r,\uparrow}$, $I \in [p+1]^{r,\uparrow}$, standard properties of determinants imply

$$\det Y_{J \times I} = \sum_{k_1 = s_{i_1 - 1}+1}^{s_{i_1}} \cdots \sum_{k_r = s_{i_r - 1}+1}^{s_{i_r}} |x_{k_1}| \cdots |x_{k_r}| \det A_{J \times K},$$

where $K = \{k_1, \dots, k_r\}$ and we have $\det A_{J \times K} > 0$ for all J, K. Since there exist suitable indices $k_t \in [s_{i_t - 1} + 1, s_{i_t}]$ such that $\prod_{t=1}^r |x_{k_t}| > 0$, it follows that $\det Y_{J \times I} > 0$ for all I, J as desired. Hence Y is TP.

With this analysis in hand, we now show (1) $\implies$ (2). The first claim is that $S^+(Ax) \leq S^-(x)$; we consider two cases:

- Suppose $n \leq p+1$. If $Ax \neq 0$, then $S^+(Ax) \leq n-1 \leq p = S^-(x)$. Otherwise $Ax = 0$. If $n = p+1$ then Y would be non-singular. But since $0 = Ax = Y\mathbf{d}_{p+1}$, where $\mathbf{d}_{p+1}$ was defined at the start of this proof, this would imply $\mathbf{d}_{p+1} = 0$, a contradiction. Hence if $Ax = 0$ then $n \leq p$, whence $S^+(Ax) = n \leq p = S^-(x)$.
- Otherwise, $n > p+1$. Define $\mathbf{w} := Ax = Y\mathbf{d}_{p+1}$, and assume for contradiction that $S^+(Ax) \geq p+1 > p = S^-(x)$. Thus there exist indices $1 \leq j_1 < \cdots < j_{p+2} \leq n$ and a sign $\varepsilon = \pm 1$ such that $\varepsilon w_{j_t} (-1)^{t-1} \geq 0$ for $t \in [1, p+2]$. Moreover, not all w_{j_t} are zero, given the rank of Y. Now consider the matrix

  $$M_{(p+2)\times(p+2)} = [\mathbf{w}_J | Y_{J \times [p+1]}], \quad \text{where} \quad J = \{j_1, \dots, j_{p+2}\}.$$

  This is singular because the first column is an alternating sum of the rest. Expanding along the first column,

  $$0 = \det M = \sum_{t=1}^{p+2} (-1)^{t-1} w_{j_t} \det Y_{(J \setminus \{j_t\}) \times [p+1]}.$$

  But all determinants in this sum are positive, all terms $(-1)^{t-1} w_{j_t}$ have the same sign ε, and not all w_{j_t} are zero. This produces the desired contradiction.

Thus if A is TP, then $S^+(Ax) \leq S^-(x)$, and it remains to show the remainder of the assertion (2). Using the notation in the preceding analysis, it remains to show that if $S^+(Ax) = S^-(x) = p$ with $Ax \neq 0$, and if moreover $\varepsilon w_{j_t} (-1)^{t-1} \geq 0$ for $t \in [1, p+1]$ – as opposed to $[1, p+2]$ in the second subcase above – then $\varepsilon = 1$ (given our original choice of 'orientation' above).

To show this, we use that Y is totally positive, so the submatrix $Y_{J\times[p+1]}$ is non-singular, where $J = \{j_1, \dots, j_{p+1}\}$. In particular, $Y_{J\times[p+1]}\mathbf{d}_{p+1} = \mathbf{w}_J$. Using Cramer's rule to solve this system for the first coordinate of $\mathbf{d}_{p+1}$,

$$1 = d_1 = \frac{\det[\mathbf{w}_J | Y_{J\times([p+1]\setminus\{1\})}]}{\det Y_{J\times[p+1]}}.$$

Multiplying both sides by $\varepsilon \det Y_{J\times[p+1]}$, we have

$$\varepsilon \det Y_{J\times[p+1]} = \sum_{t=1}^{p+1} \varepsilon(-1)^{t-1} w_{j_t} \det Y_{(J\setminus\{j_t\})\times([p+1]\setminus\{1\})}.$$

Since each summand on the right is non-negative, and Y is TP, it follows that $\varepsilon = 1$.

We next show (2) $\implies$ (3), where B is not necessarily required to be a contiguous submatrix of A, and $x^B \neq 0$ can be arbitrary. Suppose $B = A_{J\times K}$ with $J \subset [n], K \subset [m]$ both of size $1 \leq r \leq \min(m, n)$. Let $K = (k_1 < \cdots < k_r) \subset [m]$, and define $\widetilde{x} \in \mathbb{R}^m$ to have coordinates x^B_l at position k_l (for $l \in [r]$) and 0 elsewhere. By (2), we have

$$S^-(x^B) = S^-(\widetilde{x}) \geq S^+(A\widetilde{x}),$$

and this last quantity is at least $S^+(Bx^B)$ because Bx^B is a 'sub-string' of the vector $A\widetilde{x}$. Next, if $S^-(x^B) = S^+(Bx^B)$ then we draw the following conclusions:

(1) $S^+(Bx^B) \leq r - 1$, whence $Bx^B \neq 0$, so $x^B \neq 0$.
(2) $\widetilde{x} \neq 0$, and $A\widetilde{x}$ is non-zero as it contains Bx^B as a sub-string.
(3) Let $\varepsilon \in \{\pm 1\}$ be the sign of the first (respectively last) component, in any 'filling' of $A\widetilde{x}$ that attains $S^+(A\widetilde{x})$-many sign changes. Also suppose the first (respectively last) non-zero component in its sub-string Bx^B occurs in position $l \in [1, r]$. Then since $S^+(A\widetilde{x}) = S^+(Bx^B)$, the coordinates of an 'S^+-completion/filling' of $A\widetilde{x}$ in positions $1, \dots, j_1$ (respectively $j_r, \dots, n$) are all non-zero with the same sign $(-1)^{l-1}\varepsilon$ (respectively $(-1)^{r-l}\varepsilon$).

    From this it follows that the first/last coordinates in any S^+-completion/filling of $A\widetilde{x}$ and of Bx^B have the same sign. Clearly, so do the first/last non-zero coordinates of $\widetilde{x}$ and x^B.
(4) Finally, we also have $S^-(\widetilde{x}) = S^+(A\widetilde{x})$ by the above calculation. Hence by the hypotheses in assertion (2) and the preceding paragraph, we deduce assertion (3).

Finally, we show that the (restricted) assertion (3) implies that A is totally positive. By the Fekete–Schoenberg lemma 4.9, it suffices to show for all contiguous submatrices $B_{r\times r}$ of A that $\det B > 0$. This is shown by induction on $r \geq 1$. If $r = 1$, then $\operatorname{adj}(B) = (1)_{1\times 1}$, and so $0 = S^-(x^B) \geq S^+(Bx^B)$, whence $S^+(Bx^B) < 1$. It follows that $Bx^B \neq 0$, and hence that all entries of A are positive.

For the induction step, we suppose all contiguous minors of A of size at most $r - 1$ are positive, whence given a contiguous submatrix $B_{r\times r}$, it is TP_{r-1} by the Fekete–Schoenberg lemma 4.9. In particular, its adjugate matrix $\operatorname{adj}(B)$ has a checkerboard sign pattern: the (j, k) entry has sign $(-1)^{j+k}$. Now it is not hard to verify that x^B has all entries non-zero, with alternating sign pattern $(+, -, +, \dots)^T$. In particular, $S^-(x^B) = r - 1$.

The first claim is that B is invertible. If not, then $Bx^B = (\det B)v^B = 0$, whence $r = S^+(Bx^B) > S^-(x^B)$. But this is false from above, hence shows the claim. Now we show that $\det B > 0$. Indeed, the same computation as just above gives:

$$r - 1 = S^-(x^B) \geq S^+(Bx^B) = S^+((\det B)v^B).$$

Now note that regardless of the zero entries in v^B, the conditions on the α_j imply that v^B can be 'S^+-completed' to a vector with all non-zero entries and alternating signs. In particular, $S^+((\det B)v^B) = r-1$, so the hypotheses in (3) now imply that $\det B > 0$. This completes the induction step, and hence the proof. $\square$

A similar characterization via variation diminution holds for totally non-negative matrices.

Theorem 29.11. *Suppose $m, n \geq 1$ are integers, and $A \in \mathbb{R}^{n \times m}$. The following are equivalent:*

(1) A is TN.

(2) For all $x \in \mathbb{R}^m \setminus \{0\}$, $S^-(Ax) \leq S^-(x)$. Moreover, if equality holds and $Ax \neq 0$, then the first (last) non-zero component of Ax has the same sign as the first (last) non-zero component of x.

This result is again taken from Pinkus's book; and as above, Choudhury recently provided a single-vector strengthening:

Theorem 29.12 (Choudhury, [80])**.** *The assertions in Theorem 29.11 are equivalent to:*

(3) For all integers $1 \leq r \leq \min(m,n)$ and $r \times r$ submatrices B of A, and given any fixed vector $v^B := (\alpha_1, -\alpha_2, \dots, (-1)^{r-1}\alpha_r)^T$ with all $\alpha_j > 0$, we have $S^-(Bx^B) \leq S^-(x^B)$, where $x^B := \operatorname{adj}(B)v^B$. If equality occurs here and $Bx^B \neq 0$, then the first (last) non-zero component of Bx^B has the same sign as the first (last) non-zero component of x^B.

The proofs require a preliminary lemma on sign changes of limits of vectors:

Lemma 29.13. *Given $\mathbf{x} = (x_1, \dots, x_m) \in \mathbb{R}^m$, define $\widetilde{\mathbf{x}} := (x_1, -x_2, x_3, \dots, (-1)^{m-1}x_m)$. Then*

$$S^+(\mathbf{x}) + S^-(\widetilde{\mathbf{x}}) = m - 1.$$

Also, if $\mathbf{x}_k \to \mathbf{x} \neq 0$ in $\mathbb{R}^m$, then

$$\liminf_{k\to\infty} S^-(\mathbf{x}_k) \geq S^-(\mathbf{x}), \qquad \limsup_{k\to\infty} S^+(\mathbf{x}_k) \leq S^+(\mathbf{x}).$$

Proof. We begin with the first identity. Note that both $S^\pm$ are (a) invariant under the automorphism $\mathbf{y} \mapsto -\mathbf{y}$ of $\mathbb{R}^m$, and (b) additive over substrings intersecting at one non-zero number (where this common 'endpoint' is considered in both strings). Thus it suffices to prove the result for vectors $\mathbf{x} \in \mathbb{R}^m \setminus \{0\}$ with $m \geq 2$ and $x_2 = \cdots = x_{m-1} = 0$. This is a straightforward verification.

For the second part, by considering k large enough, we may assume that if the jth coordinate of $\mathbf{x}$ is non-zero, then it has the same sign as the jth coordinate of every $\mathbf{x}_k$ – in fact, we may take these coordinates to all have the same value, since this does not affect $S^\pm(\mathbf{x}_k)$. Now by the observation in the paragraph following Definition 29.7,

$$S^-(\mathbf{x}) \leq S^-(\mathbf{x}_k) \leq S^+(\mathbf{x}_k) \leq S^+(\mathbf{x}),$$

and the result follows. $\square$

We can now prove the above characterizations of TN matrices.

Proof of Theorems 29.11 and 29.12. The proof of (1) $\implies$ (2) uses the two preceding results and Whitney's density Theorem 6.7. Since A is TN, there exists a sequence A_k of TP matrices converging entrywise to A. Now use Theorem 29.8 and Lemma 29.13 to compute:

$$S^-(Ax) \leq \liminf_{k\to\infty} S^-(A_k x) \leq \liminf_{k\to\infty} S^+(A_k x) \leq \liminf_{k\to\infty} S^-(x) = S^-(x).$$

Next, if equality holds and $Ax \neq 0$, then for all k large enough, we have

$$p := S^-(Ax) \leq S^-(A_k x) \leq S^+(A_k x) \leq S^-(x) = p,$$

by Theorem 29.8 and Lemma 29.13. In particular, these are all equalities, whence $S^-(A_k x) = S^+(A_k x)$ for $k \gg 0$. This implies (for large k) that the sign changes/patterns in $A_k x$ have no dependence on zero entries. At the same time, both vectors x and $A_k x$ admit 'partitions' of the form (29.10) with alternating signs, with precisely p sign changes. Now Theorem 29.8 implies that these signs are in perfect agreement, for all large k. Hence the same holds for the sign patterns of x and Ax.

We next show that (2) $\implies$ (3), again for arbitrary vectors $0 \neq x^B \in \mathbb{R}^r$, where $1 \leq r \leq \min(m,n)$; this is similar to the proof of Theorems 29.8 and 29.9. Say $B = A_{J \times K}$ for some $J \subset [n], K \subset [m]$ of equal size r. Embed x into a vector $\widetilde{x} \in \mathbb{R}^m$ at positions K, with zero entries in the other positions. Then Bx^B is a sub-string of $A\widetilde{x}$, so by (2),

$$S^-(Bx^B) \leq S^-(A\widetilde{x}) \leq S^-(\widetilde{x}) = S^-(x).$$

Moreover, if $S^-(Bx^B) = S^-(x)$ and $Bx^B \neq 0$, then:

- All four S^--terms here are equal.
- Bx^B is non-zero, whence so are x^B, so $\widetilde{x}$, and hence $A\widetilde{x}$.

Now suppose the first (last) non-zero entry of $Bx^B \in \mathbb{R}^r$ occurs in position $l \in [1,r]$, and $J = (j_1 < \cdots < j_r) \subset [n]$. Since $S^-(Bx^B) = S^-(A\widetilde{x})$, all entries of $A\widetilde{x}$ before (after) position j_l must have the same sign. This shows (2) $\implies$ (3) for any $0 \neq x^B$.

Finally, we show that (3) $\implies$ (1), with the vectors v^B, x^B as specified. The claim that all $r \times r$ minors of A are non-negative, is shown by induction on $r \geq 1$. For the base case, if $B = (a_{jk})_{1\times 1} = 0$ then there is nothing to prove; otherwise $\operatorname{adj}(B) = (1)$, so $x^B = v^B = (\alpha_1)$, where $\alpha_1 > 0$. Hence $S^-(Bx^B) = 1 = S^-(x^B)$, and so a_{jk} and 1 have the same sign.

For the induction step, suppose $B_{r\times r}$ is a submatrix of A, which we may assume is TN_{r-1}. If $\det B = 0$ then there is nothing to prove, so suppose B is non-singular. Then every row / column of $\operatorname{adj}(B)$ is non-zero (otherwise one can expand B along a suitable column / row and get $\det B = 0$). Moreover, $\operatorname{adj}(B)$ has entries in a checkerboard pattern:

$$\operatorname{sgn}(\operatorname{adj}(B)_{jk}) = (-1)^{j+k}, \qquad \forall 1 \leq j,k \leq r.$$

Since all coordinates of v^B are non-zero, it follows that $x^B = (\beta_1, -\beta_2, \dots, (-1)^{r-1}\beta_r)^T$ with all $\beta_j > 0$. Hence,

$$S^-(x^B) = r - 1 = S^-(v^B), \qquad S^-(Bx^B) = S^-((\det B)v^B) = r - 1.$$

Hence by (3), the first non-zero coordinates of $x^B, (\det B)v^B$ have the same signs, which implies $\det B > 0$. This completes the induction step, and with it the proof. $\square$

29.2. **Variation diminishing property for Pólya frequency functions.** The above characterizations of TN and TP matrices have many applications in mathematics and other sciences; we do not expound on these, referring the reader to Karlin's treatise [200] and numerous follow-up papers in the literature. Here we present continuous analogues of the above results on the variation diminishing property, albeit only in one direction. These are again found in Schoenberg's 1951 paper in *J. d'Analyse Math.* We begin with the definition of the 'variation' that will diminish under the action of a TN kernel.

Definition 29.14. Suppose $I \subset \mathbb{R}$ is an interval with positive measure, and a function $f : I \to \mathbb{R}$. The number $S_I^-(f)$ of variations of sign of $f(x)$ on I is defined as follows:

$$S_I^-(f) := \sup\{S^-((f(x_1), \dots, f(x_p))) : p \geq 1, \mathbf{x} \in I^{p,\uparrow}\}.$$

With this notion at hand, one can show the variation diminishing property for Pólya frequency functions.

Proposition 29.15 (Schoenberg, [324]). *Suppose $\Lambda : \mathbb{R} \to [0,\infty)$ is a Pólya frequency function. Suppose $f : \mathbb{R} \to \mathbb{R}$ is integrable on all finite intervals in $\mathbb{R}$, and such that the integral*

$$g(x) := \int_{\mathbb{R}} \Lambda(x-t) f(t)\ dt$$

converges at all $x \in \mathbb{R}$. Then $S^-_{\mathbb{R}}(g) \leq S^-_{\mathbb{R}}(f)$.

Proof. We will write $S^-(\cdot)$ for $S^-_{\mathbb{R}}(\cdot)$ in what follows. We may assume in the sequel that $S^-(g) > 0$ and $S^-(f) < \infty$. Now if $\mathbf{x} \in \mathbb{R}^{m+1,\uparrow}$ satisfies: $g(x_1), g(x_2), \dots, g(x_{m+1})$ are all non-zero and alternate in sign, then it suffices to show that $S^-(f) \geq m$, for one can now take the supremum over all such tuples $\mathbf{x}$ to deduce $S^-(g) \leq S^-(f)$. Note that

$$g(x) = \lim_{a \to -\infty,\ b \to \infty} \int_a^b \Lambda(x-t) f(t)\ dt,$$

and this convergence is uniform when considered simultaneously at the $m+1$ coordinates of $\mathbf{x}$. Thus, select $-\infty < a < b < \infty$ such that the function

$$g_1(x) := \int_a^b \Lambda(x-t) f(t)\ dt$$

also alternates in sign at $x_1, \dots, x_{m+1}$. Approximating this function by Riemann sums over $n > 0$ sub-intervals of $[a,b]$ of equal length, simultaneously at $x_1, \dots, x_{m+1}$, it is possible to choose $n \gg 0$ (large enough) such that the sequence of Riemann sums

$$z_j^{(n)} := \frac{b-a}{n} \sum_{k=1}^n \Lambda\left(x_j - \tfrac{kb+(n-k)a}{n}\right) f\left(\tfrac{kb+(n-k)a}{n}\right), \qquad j = 1, \dots, m+1 \tag{29.16}$$

also alternates in sign. In other words, $S^-(\mathbf{z}^{(n)}) = m,\ \forall n \gg 0$.

We now invoke the total non-negativity of the kernel T_Λ, applied to $\mathbf{x} \in \mathbb{R}^{m+1,\uparrow}$ as above, and $\mathbf{y} \in \mathbb{R}^{n,\uparrow}$ given by $y_k = (kb + (n-k)a)/n$ for $k = 1, \dots, n$. Thus the matrix

$$A_{(m+1)\times n} := T_\Lambda[\mathbf{x}; \mathbf{y}] = \left(\Lambda\left(x_j - \tfrac{kb+(n-k)a}{n}\right)\right)_{j \in [m+1],\ k \in [n]}$$

is totally non-negative. Now (29.16) and Theorem 29.11 imply that

$$m = S^-(\mathbf{z}^{(n)}) = S^-(Av) \leq S^-(v), \quad \text{where} \quad v = (f(y_k))_{k=1}^n.$$

But then $m \leq S^-(f)$, as desired. $\square$

Schoenberg goes on to prove an analogue of this variation diminishing property, when f and hence g (in the preceding result) are polynomials. The diminution is now in the number of real roots.

Proposition 29.17. *Suppose Λ is a Pólya frequency function, and $f \in \mathbb{R}[x]$ is a polynomial of degree n. Then $g(x) := \int_{\mathbb{R}} \Lambda(x-t) f(t)\ dt$ is also a polynomial of degree n, with $Z(g) \leq Z(f)$. Here $Z(f)$ denotes the number of real roots of f, counted with multiplicity.*

Proof. By Proposition 29.3(3)(c), Λ has finite moments

$$\mu_j := \int_{\mathbb{R}} \Lambda(t) t^j\ dt < \infty, \qquad j = 0, 1, \dots. \tag{29.18}$$

This implies in particular that g is well-defined everywhere, since

$$g(x) = \int_{\mathbb{R}} \Lambda(t) f(x-t)\ dt = \int_{\mathbb{R}} \Lambda(t) \sum_{j=0}^n f^{(j)}(x) \frac{(-t)^j}{j!}\ dt = \sum_{j=0}^n \frac{(-1)^j f^{(j)}(x) \mu_j}{j!}. \qquad (29.19)$$

Since $\mu_0 > 0$ by assumption on Λ, it follows that $g(x)$ is a polynomial of degree n. Now change the 'monomial' basis diagonally, and write the coefficients of f, g as follows:

$$f(x) = a_0 x^n + \binom{n}{1} a_1 x^{n-1} + \cdots + a_n,$$
$$g(x) = b_0 x^n + \binom{n}{1} b_1 x^{n-1} + \cdots + b_n,$$

where $a_0 \neq 0$. This yields a triangular, invertible change of basis from $(a_k)_k$ to $(b_k)_k$:

$$b_k = \sum_{j=0}^k \binom{k}{j} (-1)^j \mu_j a_{k-j}. \qquad (29.20)$$

We now turn to the real roots. If g has a real root c, say a factor of $(x-c)^n$, then replace this by $(x-c)(x-c-\epsilon)\cdots(x-c-(n-1)\epsilon)$ for sufficiently small $\epsilon > 0$. Carrying this out for every real root yields a perturbed polynomial, which we call $g^{(\epsilon)}(x)$. Inverting the triangular transformation (29.20) yields a perturbed polynomial $f^{(\epsilon)}$, with $g^{(\epsilon)}(x) = \int_{\mathbb{R}} \Lambda(x-t) f^{(\epsilon)}(t)\ dt$ for all $\epsilon > 0$. Note moreover that $f^{(\epsilon)}, g^{(\epsilon)}$ still have degree n for all $\epsilon > 0$, and converge coefficientwise to f, g respectively, as $\epsilon \to 0^+$. Hence for $\epsilon > 0$ small enough, the continuity of roots implies $Z(f^{(\epsilon)}) \leq Z(f)$.

The key observation now is that if if $p(x)$ is a real polynomial, then $S^-_{\mathbb{R}}(p) \leq Z(p)$, with equality if and only if all real roots of p are simple. Applying this to the above analysis, we conclude for small $\epsilon > 0$:

$$Z(g) = Z(g^{(\epsilon)}) = S^-_{\mathbb{R}}(g^{(\epsilon)}) \leq S^-_{\mathbb{R}}(f^{(\epsilon)}) \leq Z(f^{(\epsilon)}) \leq Z(f),$$

where the first inequality follows by Proposition 29.15. □

We conclude with analogues of the above results for *one-sided Pólya frequency (PF) functions* – also proved by Schoenberg in [324]. A distinguished class of PF functions consists of those vanishing on a semi-axis, and we will see examples of such functions later in this part of the text. By a linear change of variables, we may assume such a function Λ vanishes on $(-\infty, 0)$. In this case the above transformation becomes

$$g(x) = \int_{-\infty}^x \Lambda(x-t) f(t)\ dt.$$

If we now consider $f : (-\infty, 0] \to \mathbb{R}$, then the upper limit in the preceding integral may be changed to 0, to yield another function g defined on $(-\infty, 0]$. In this case, Schoenberg proved in [324] similar results to above, and we now state these without proof.

Proposition 29.21. *Suppose Λ is a Pólya frequency function which vanishes on $(-\infty, 0)$.*

1. *Let $f : (-\infty, 0] \to \mathbb{R}$ be integrable on every finite interval, and such that $g(x) := \int_{-\infty}^0 \Lambda(x-t) f(t)\ dt$ converges for all $x \leq 0$. Then $S^-_{(-\infty,0]}(g) \leq S^-_{(-\infty,0]}(f)$.*
2. *Suppose f is a real polynomial of degree n, and let $g(x) := \int_0^\infty \Lambda(t) f(x-t)\ dt$. Then g is also a real polynomial of degree n, and $Z_-(g) \leq Z_-(f)$, where $Z_-(f)$ denotes the number of non-positive real roots of f, counted with multiplicity.*

29.3. **Early results on total non-negativity and variation diminution.** We next discuss some of the origins of the variation diminishing property, and its connection to total non-negativity. First observe by Theorem 4.1 and Remark 4.4 for Hankel moment matrices, that total non-negativity is implicit in the solution of the Stieltjes moment problem – as also in the Routh–Hurwitz criterion for stability; see Theorem 31.3 – both from the 1800s.

Coming to the variation diminishing property: it shows up in the correspondence [118] between Fekete and Pólya (published in 1912) that has been mentioned above in the context of proving the 'contiguous minor' test for total positivity of a matrix. (As a bit of trivia: Fekete, Pólya, and Szász were three of the earliest students of L. Fejér; his last student was Vera Sós; and other students include Erdős, von Neumann, Turán, Aczel, Egerváry, Tóth, and Marcel Riesz, among others.)[5]

We now discuss a result of Fekete from his correspondence with Pólya, which not only dealt with variation diminution but also gave rise to the notion of Pólya frequency sequences – discussed in the next section.

A common, historical theme underlying this section and the next, as well as Section 33 below (on the Laguerre–Pólya–Schur program), involves understanding real polynomials and their roots. In fact this theme dates back to Descartes, who in his 1637 work [100] proposed his 'Rule of Signs' – see e.g. Lemma 5.2. The question of understanding the roots remained popular (and does so to this day; see Section 33). As a notable example, we recall Laguerre's 1883 paper [229], which deals with this theme, and opens by recalling Descartes' rule of signs and proving it using Rolle's theorem (the 'standard' proof these days). In [229], Laguerre used the word 'variations' to denote the sign changes in the Maclaurin coefficients of a polynomial or power series. Among the many results found in his memoir, we list two:

Theorem 29.22 (Laguerre, 1883, [229])**.**

1. *Given an interval $[a,b] \subset \mathbb{R}$ and an integrable function $\Phi : [a,b] \to \mathbb{R}$, the number of zeros of the Fourier–Laplace transform $\int_a^b e^{-sx}\Phi(x)\,dx$ is at most the number of zeros of the antiderivative $\int_a^t \Phi(x)\,dx$ for $t \in [a,b]$.*
2. *Suppose $f(x)$ is a polynomial. Then the number of variations (sign changes in the Maclaurin coefficients) of the power series $e^{sx}f(x)$ is a non-increasing function of $s \in [0,\infty)$, and is uniformly bounded below by the number of positive roots of f.*

In particular, in part (2), the variation for any $s > 0$ is 'diminishing', and always finite, since it is bounded above by the variation at $s = 0$. See Section 30.4 for the proof of part (2).

Especially this latter result was pursued by Fekete, who wrote to Pólya to the effect that "Laguerre did not fully justify" Theorem 29.22(2) in his work [229]. To address this, Fekete considered a formal power series with real coefficients $c_0 + c_1 t + \cdots$, acting by multiplication on the space of such power series. With respect to the standard basis of monomials, this

[5] As another bit of trivia, the descendants of K.T.W. Weierstrass and E.E. Kummer feature prominently in the study of (total) positivity. Their joint students include:

1. F.G. Frobenius and L.I. Fuchs, who in turn mentored I. Schur, who was the advisor of I.J. Schoenberg;
2. L. Königsberger, who mentored G. Pick, who advised C. Loewner; and
3. L. Fejér, who mentored G. Pólya and M. Fekete.

transformation is given by the triangular matrix

$$T_{\mathbf{c}} := \begin{pmatrix} c_0 & 0 & 0 & \cdots \\ c_1 & c_0 & 0 & \cdots \\ c_2 & c_1 & c_0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Fekete then stated the following result on variation diminution (a term coined by Pólya) – further asserting that it could be used to prove Theorem 29.22(2):

Proposition 29.23 (Fekete, 1912, [118])**.** *Suppose for an integer $p \geq 2$ that the matrix $T_{\mathbf{c}}$ (supported on $\mathbb{Z}^{\geq 0} \times \mathbb{Z}^{\geq 0}$) is TN_p – i.e., every finite submatrix is TN_p. Given a vector $\mathbf{x} = (x_0, x_1, \dots, x_{p-1}, 0, 0, \dots)^T$ (the coefficients of a polynomial), we have $S^-(T_{\mathbf{c}}\mathbf{x}) \leq S^-(\mathbf{x})$.*

Note here that $T_{\mathbf{c}}\mathbf{x}$ represents the coefficients of a formal power series, and hence can form a sequence with infinitely many non-zero terms. Nevertheless, every such sequence has fewer sign changes than the *finite* sequence $\mathbf{x}$. (To deduce Laguerre's result, Fekete showed that for $s \geq 0$, Proposition 29.23 applies to the special case $c_k = s^k/k!$.)

Thus, we have journeyed from Descartes (1637), to Laguerre (1883), to Fekete (1912), to Schoenberg (1930) and Motzkin (1936) – see Theorem 3.22 from Motzkin's thesis and the preceding paragraph – to Schoenberg (1951), in studying the origins of variation diminution and subsequent developments. (This omits, with due apologies, the substantial contributions of Gantmacher, Krein, and others; the connections to the work of Sturm and to Sturm–Liouville eigenvalue problems; as well as the 1915 paper [285] of Pólya, which proved a different variation-diminishing property of PF functions on polynomials, led Schoenberg to coin the phrase 'Pólya frequency functions', and is briefly discussed in Section 34.) Certainly Fekete's result and the aforementioned developments led Schoenberg and Gantmacher–Krein to develop the theories of total positivity, Pólya frequency functions, and variation diminution.

We conclude this historical section with yet another connection between total non-negativity and Descartes' rule of signs. In his 1934 paper [312] in *Math. Z.*, Schoenberg proved the following result. In it, we use the notation that a finite sequence $f_0, f_1, \dots, f_n$ of functions *obeys Descartes' rule of signs* in an open sub-interval $(a,b) \subset \mathbb{R}$ if the number of zeros in (a,b) of the nontrivial real linear combination

$$c_0 f_0(x) + \cdots + c_n f_n(x), \qquad c_j \in \mathbb{R}, \ \sum_{j=0}^{n} c_j^2 > 0$$

is no more than the number of sign changes in the sequence $(c_0, \dots, c_n)$.

Theorem 29.24 (Schoenberg, [312])**.** *Fix a sequence of real polynomials $p_j(x) := a_{j0} + a_{j1}x + \cdots + a_{jj}x^j$ for $0 \leq j \leq n$, with all $a_{jj} > 0$. The following are equivalent:*

(1) The sequence $(p_0, \dots, p_n)$ obeys Descartes' rule of signs in $(0, \infty)$.

(2) The upper triangular matrix $\begin{pmatrix} a_{00} & a_{10} & \cdots & a_{n0} \\ & a_{11} & \cdots & a_{n1} \\ & & \ddots & \vdots \\ & & & a_{nn} \end{pmatrix}$ *is totally non-negative.*

(E.g., for the 'usual' Descartes' rule of signs, this matrix is precisely the identity matrix.) Schoenberg also extended Descartes and Laguerre's results to more general domains in $\mathbb{C}$; see his 1936 paper [314] in *Duke Math. J.*

29.4. **The sign non-reversal property.** We now discuss another fundamental – and very recent – characterization of total positivity. To motivate this result, begin with Schoenberg's 1930 result asserting that every TN matrix has the variation diminishing property. From here, it is natural to proceed in two directions:

(1) Characterize all matrices with the variation diminishing property. This was carried out by Motzkin in his 1936 thesis; see Theorem 3.22.
(2) Find additional conditions that, together with the variation diminishing property, characterize TN and TP matrices. This was carried out in Theorems 29.8 and 29.11, and involves a 'sign non-reversal' phenomenon.

Our goal here is to show that in fact, it is the variation diminishing property that is not necessary in Theorems 29.8 and 29.11. To proceed, we isolate the key notion into the following definition.

Definition 29.25. Fix integers $n \geq 1$.

(1) A square real matrix $A_{n\times n}$ has the *sign non-reversal property* on a set of vectors $S \subset \mathbb{R}^n$, if for all non-zero vectors $x \in S$ there exists a coordinate $j \in [n]$ such that $x_j(Ax)_j > 0$. If the set S is not specified, it is taken to be $\mathbb{R}^n$.
(2) We also require the *non-strict sign non-reversal property* for a matrix $A \in \mathbb{R}^{n\times n}$, on a set of vectors $S \subset \mathbb{R}^n$. This says that for all $0 \neq x \in S$ there exists a coordinate $j \in [n]$ such that $x_j \neq 0$ and $x_j(Ax)_j \geq 0$.
(3) Define $\mathbb{R}^n_{\text{alt}}$ to be the open bi-orthant consisting of the vectors in $\mathbb{R}^n$ whose coordinates are all non-zero and have alternating signs.
(4) Finally, define $\mathbf{d}_r := (1, -1, \dots, (-1)^{n-1})^T \in \mathbb{R}^n_{\text{alt}}$.

With these notions defined, in their recent work in *Bull. London Math. Soc.*, Choudhury, Kannan, and Khare have characterized TP/TN matrices purely in terms of their sign non-reversal property. More strongly:

Theorem 29.26 ([81])**.** *Fix integers $m, n \geq p \geq 1$ and a real matrix $A \in \mathbb{R}^{m\times n}$. The following are equivalent:*

(1) The matrix A is TP_p.
(2) Every square submatrix B of A of size $r \in [1,p]$ has the sign non-reversal property.
(3) Every contiguous square submatrix B of A of size $r \in [1,p]$ has the sign non-reversal property on $\mathbb{R}^r_{\text{alt}} \subset \mathbb{R}^r$.

A part of this theorem was previously shown in 1966 by Ky Fan [114]. Also note that (2) $\implies$ (3), so we will show (1) $\implies$ (2) and (3) $\implies$ (1). The latter implication can be weakened even further, to require the sign non-reversal property *for a single alternating vector* (chosen from an orthant) – the same orthant as in Theorem 29.9. This was shown very recently:

Theorem 29.27 (Choudhury, [80])**.** *The three conditions above are further equivalent to:*

(4) For each contiguous square submatrix B of A of size $r \in [1,p]$, and any choice of vector $0 \neq v^B := (\alpha_1, -\alpha_2, \dots, (-1)^{r-1}\alpha_r)^T$ with all $\alpha_j \geq 0$, B has the sign non-reversal property for $z^B := \operatorname{adj}(B)v^B$, where $\operatorname{adj}(B)$ is the adjugate matrix of B.

(As the proof will reveal, one can also work with $-v^B$ instead of v^B.) Following the proof of these results, we will show that $A_{m\times n}$ is TN_p if and only if every square submatrix of size $\leq p$ satisfies a condition similar to this one. See Theorem 29.28 and the subsequent result.

Proof. We first show that (1) implies the sign non-reversal property on vectors with non-negative coordinates – under the weaker assumption of working only with the positivity of the *principal* minors of B, of size $\leq r$. (Such matrices are called P-matrices; see e.g. [105, 121, 134] for this argument.) The result is by induction on $r \geq 1$, with the $r = 1$ case obvious. Suppose the result holds for all $(r-1) \times (r-1)$ real matrices with positive principal minors. Now let $B_{r \times r}$ have the same property, with $x, -Bx \in \mathbb{R}^r$ having all non-negative coordinates. We need to prove $x = 0$.

By choice (and Cramer's rule), B^{-1} has positive diagonal entries. Let $\mathbf{b}$ denote the first column of B^{-1}, and define

$$\theta := \min_{j \in [r] : b_j > 0} \frac{x_j}{b_j},$$

noting that the minimum is taken over a non-empty set. Hence it is attained: $0 \leq \theta = x_{j_0}/b_{j_0}$ at some $j_0 \in [r]$. Then $\mathbf{y} := x - \theta \mathbf{b}$ has non-negative coordinates, by choice of θ, and it has j_0th coordinate zero. But we also have

$$-B\mathbf{y} = -Bx + \theta B\mathbf{b} = -Bx + \theta \mathbf{e}_1,$$

and this has non-negative coordinates again.

We now claim $\mathbf{y} = 0$. Indeed, if we obtain $\mathbf{y}'$ by deleting the j_0th coordinate, and B' by deleting the j_0th row and column, then an easy verification yields that $-B'\mathbf{y}', \mathbf{y}'$ have non-negative coordinates. By the induction hypothesis, $\mathbf{y}' = 0$, whence $\mathbf{y} = 0$. But then $Bx = \theta \mathbf{e}_1$ has non-negative coordinates. Since so does $-Bx$, we have $Bx = 0$. Since $\det(B) > 0$, we obtain $x = 0$. This completes the proof of (1) $\implies$ (2) by induction – for x with non-negative coordinates.

Now suppose $x \in \mathbb{R}^r$ and $x_j(Bx)_j \leq 0$ for all j. Let $J \subset [r]$ index the negative coordinates of x, and define the diagonal matrix D_J with (k,k) entry $(-1)^{\mathbf{1}(k \in J)}$. If now x_j and $(Bx)_j$ have opposite signs for all j (meaning their product is non-positive), then so do $(\mathbf{e}_j^T D)x = \mathbf{e}_j^T(Dx)$ and $(\mathbf{e}_j^T D)(Bx) = \mathbf{e}_j^T(DBD)(Dx)$. Thus $Dx, (DBD)(Dx)$ have corresponding coordinates of opposite signs. As DBD also has all principal minors positive, the above analysis implies $Dx = 0$, and so $x = 0$. Hence (1) $\implies$ (2) for all vectors $x \in \mathbb{R}^r$.

(Though we do not require it, we also mention quickly why (2) $\implies$ (1). Let B be any square submatrix of A; as the set of non-real eigenvalues is closed under conjugation, their product is strictly positive. It thus suffices to show that every real eigenvalue is positive. By the sign non-reversal property, if $Bx = \lambda x$ with $\lambda \in \mathbb{R}$ and $x \in \mathbb{R}^r$, and if $x_j \neq 0$, then $x_j(Bx)_j = \lambda x_j^2 > 0$, whence $\lambda > 0$.)

Next, that (2) $\implies$ (3) $\implies$ (4) is immediate. Finally, suppose (4) holds. By the Fekete–Schoenberg lemma 4.9, it suffices to show that all contiguous $r \times r$ minors are positive, for $1 \leq r \leq p$. The proof is by induction on $r \leq p$; the $r = 1$ case directly follows from (4) using $r = 1$ and $B_{1 \times 1} = (a_{jk})$, using that $\operatorname{adj}(B) = (1)$.

For the induction step, suppose $B_{r \times r}$ is a contiguous square submatrix of A, with $r \leq p$, and all contiguous minors of B of size $\leq r - 1$ are positive. Then the same holds for all proper minors of B by Lemma 4.9, and so $\operatorname{adj}(B)$ is a matrix with a 'checkerboard' pattern of signs: $\operatorname{sgn}(\operatorname{adj}(B)_{jk}) = (-1)^{j+k}$ for $1 \leq j, k \leq r$. It follows for any $v^B \neq 0$ as specified that $z^B = \operatorname{adj}(B)v^B \in \mathbb{R}^r_{\text{alt}}$. Now compute for $j_0 \in [1, r]$:

$$0 < (z^B)_{j_0} \cdot (Bz^B)_{j_0} = (-1)^{j_0 - 1}|(z^B)_{j_0}| \cdot (\det B)(v^B)_{j_0} = \det(B)|(z^B)_{j_0}|\alpha_{j_0}.$$

It follows that all three factors are non-zero, and $\det B > 0$, which completes the proof. $\square$

The final results here, again by Choudhury–Kannan–Khare and Choudhury (2021), characterize TN_p matrices through their sign non-reversal:

Theorem 29.28 ([81])**.** *Fix integers $m, n \geq p \geq 1$ and a real matrix $A \in \mathbb{R}^{m \times n}$. The following are equivalent:*

(1) The matrix A is TN_p.

(2) Every square submatrix B of A of size $r \in [1, p]$ has the non-strict sign non-reversal property on $\mathbb{R}^r$.

(3) Every square submatrix B of A of size $r \in [1, p]$ has the non-strict sign non-reversal property on $\mathbb{R}^r_{\mathrm{alt}}$.

As in Theorem 29.27(4) above, this can be further weakened:

Theorem 29.29 (Choudhury, [80])**.** *The preceding three conditions are further equivalent to:*

(4) Every square submatrix B of A of size $r \in [1, p]$, has the non-strict sign non-reversal property for the single vector $z^B := \mathrm{adj}(B) v^B$, where $v^B \in \mathbb{R}^r_{\mathrm{alt}}$ is arbitrarily chosen.

Proof. First suppose (1) holds. By Whitney's density theorem 6.7 there is a sequence $A^{(l)}$ of TP_p matrices converging entrywise to A. Now given B, let $B^{(l)}_{r \times r}$ be the submatrix of $A^{(l)}$ indexed by the same rows and columns as B. Now fix a vector $0 \neq x \in \mathbb{R}^r$, and index by $J \subset [r]$ the non-zero entries in x. Since $B^{(l)}$ is TP, by Theorem 29.26 there exists $j_l \in [r]$ such that $x_{j_l} (B^{(l)} x)_{j_l} > 0$; moreover, $j_l \in J \ \forall l$. As J is finite, there exists $j_0 \in [r]$ and an increasing subsequence of positive integers l_q such that $j_{l_q} = j_0$ for all $q \geq 1$. Now (2) follows:

$$x_{j_0} (Bx)_{j_0} = \lim_{q \to \infty} x_{j_{l_q}} (B^{(l_q)} x)_{j_{l_q}} \geq 0, \qquad x_{j_0} \neq 0.$$

(The proof of (2) $\implies$ (1) is essentially the same as in the preceding proof, with $\lambda x_j^2 \geq 0$ now; once again, we do not require it.) That (2) $\implies$ (3) $\implies$ (4) is immediate. Finally, assume (4) and claim by induction on $r \leq p$, that every $r \times r$ minor of A is non-negative. The base case is immediate; for the induction step, let $B_{r \times r}$ be a submatrix of A, and assume that all $(r-1) \times (r-1)$ minors of B are non-negative. If $\det B = 0$ then we are done; else assume B is invertible. Now no row or column of $\mathrm{adj}(B)$ is zero, and $\mathrm{adj}(B)_{jk}$ is either zero or has sign $(-1)^{j+k}$ for all j, k. Thus, $\mathrm{adj}(B) v^B \in \mathbb{R}^r_{\mathrm{alt}}$. Now a similar computation as in the preceding proof shows that for some $j_0 \in [1, r]$, we have by (4):

$$0 \leq (B \, \mathrm{adj}(B) v^B)_{j_0} (\mathrm{adj}(B) v^B)_{j_0} = (\det B) v^B_{j_0} z^B_{j_0}.$$

Since all factors here are non-zero and $v^B_{j_0}, z^B_{j_0}$ have the same sign, it follows that $\det B > 0$, and so we are done by induction. $\square$

Remark 29.30. In this section, we have seen classical results which help characterize totally positive/non-negative matrices A using the variation diminishing property and the sign non-reversal property – both of which involve certain conditions holding for all vectors in $\mathbb{R}^n$. We also explained recent results of Choudhury [80, 81] that provided *single test vectors*, one for every (contiguous) submatrix of A. In [80], Choudhury also provides a *third* characterization of total positivity – via the Linear Complementarity Problem, which has applications in bimatrix games, linear programming, and other areas. Once again, he is able to improve this characterization to use a single test vector. See [80] for details.

30. Pólya frequency sequences and their generating functions.

In this section, we introduce and study Pólya frequency sequences, and take a look at root-location phenomena for their generating functions. Recall Fekete's proposition 29.23 (1912) from the preceding section, on the variation diminishing property of TN_p triangular Toeplitz matrices that are supported on $\mathbb{Z} \times \mathbb{Z}$. Such matrices are now known as PF sequences:

Definition 30.1.

(1) A real sequence $\mathbf{a} = (a_n)_{n \in \mathbb{Z}}$ is a *Pólya frequency (PF) sequence* if the associated Toeplitz kernel $T_{\mathbf{a}} : \mathbb{Z} \times \mathbb{Z} \to \mathbb{R}$, sending $(x, y) \mapsto a_{x-y}$, is TN.
(2) The PF sequence $\mathbf{a}$ is said to be *one-sided* if there exist $n_\pm \in \mathbb{Z}$ such that $a_n = 0$ either for all $n < n_-$ or for all $n > n_+$ (or both).
(3) More general is the notion of a *p times (i.e., multiply) positive sequence*, or a TN_p sequence $\mathbf{a}$, which corresponds to the matrix $T_{\mathbf{a}} : \mathbb{Z} \times \mathbb{Z} \to \mathbb{R}$ being TN_p. (This too has a one-sided version, as above.)

For instance, one can specialize the results of Section 28.2 to $X = Y = \mathbb{Z}$, to obtain characterizations of TN_p sequences or PF sequences.

30.1. **Examples.** We begin by studying examples of PF sequences, and of the subclass of one-sided PF sequences (TN or TN_p). Clearly, every constant (non-negative) sequence $a_n \equiv c$ for $c \geq 0$ is a PF sequence. Our first source of non-constant PF sequences comes from Pólya frequency functions:

Lemma 30.2. *If $\Lambda : \mathbb{R} \to \mathbb{R}$ is a PF function, or even a TN function, then $(\Lambda(an + b))_{n \in \mathbb{Z}}$ is a PF sequence for $a, b \in \mathbb{R}$. If $a \neq 0$ and Λ is TP, then so is $(\Lambda(an + b))_{n \in \mathbb{Z}}$.*

Proof. If $a = 0$ then the result is immediate. Now suppose $a \neq 0$, and $\mathbf{x}, \mathbf{y} \in \mathbb{Z}^{p,\uparrow}$ for some $p \geq 1$. Then the $p \times p$ matrix

$$(\Lambda(a(x_j - y_k) + b))_{j,k=1}^p = \begin{cases} T_\Lambda[a\mathbf{x} + b\mathbf{1}; a\mathbf{y}], & \text{if } a > 0, \\ T_\Lambda[|a|\mathbf{y} + b\mathbf{1}; |a|\mathbf{x}]^T, & \text{if } a < 0, \end{cases}$$

and both choices have non-negative determinant. The final assertion is showed similarly. □

Another elementary result is the closure of the set of TN_p sequences. The proof is easy.

Lemma 30.3. *Suppose $p \geq 1$ and $\mathbf{a}^{(k)} = (a_n^{(k)})_{n \in \mathbb{Z}}$ is a TN_p sequence for all $k \geq 1$. If $\mathbf{a}^{(k)} \to \mathbf{a}$ pointwise as $k \to \infty$, then $\mathbf{a}$ is also TN_p.*

The above lemmas allow one to 'draw' PF sequences from PF functions, along any infinite arithmetic progression. Here are two examples:

Example 30.4. For a real number $q > 0$, the sequences $(q^n)_{n \in \mathbb{Z}}$ and $(q^n \mathbf{1}_{n \geq 0})_{n \in \mathbb{Z}}$ are both PF sequences. The former is because $(q^{x_j - y_k})_{j,k=1}^n$ is a rank-1 matrix with positive entries for any choice of integers x_j, y_k, so TN. If $q \in (0, 1)$, then the latter kernel is drawn from $\lambda_1(x) = e^{-x} \mathbf{1}_{x \geq 0}$ (which is shown in Example 32.11 below to be a PF function), at the arguments $x = -n \log(q)$. If $q > 1$ then it is drawn from $f_c(x) = e^{cx} \lambda_1(x)$ at the arguments $n \log(q)$ and with $c = 2$. (By Lemma 28.3, f_c is now a TN function as well.) Finally, if $q = 1$ then it is drawn from the Heaviside kernel

$$H_1(x) = \mathbf{1}_{x \geq 0} = \lim_{c \to 1} f_c(x)$$

at any infinite arithmetic progression. (Alternately, for all q one can use the analysis above (30.11), later in this section.)

Example 30.5. For a real number $q \in (0,1)$, the sequence $(q^{n^2})_{n\in\mathbb{Z}}$ is a TP PF sequence. Indeed, this sequence is drawn from the Gaussian PF function G_σ for $\sigma = -\log(q) > 0$ (see Example 29.4).

30.2. **Generating functions and representation theorem for one-sided PF sequences.** Given a real sequence $\mathbf{a} = (a_n)_{n\geq 0}$ of finite or infinite length, it is natural to encode it by the corresponding *generating function*

$$\Psi_{\mathbf{a}}(x) = \sum_{n\geq 0} a_n x^n.$$

When the sequence $(\dots, 0, 0, a_0, a_1, a_2, \dots)$ is a TN_p sequence, and only finitely many terms a_n are non-zero, one can deduce results about the locations of the roots of the corresponding generating polynomial $\Psi_{\mathbf{a}}(x)$ – and in turn, use this information to classify all finite TN sequences. We carry out this analysis in the present subsection and the next; it will be useful in the next part of this text, in classifying the preservers of total positivity for PF sequences. The present subsection ends by revealing the most general form of the generating function of one-sided PF sequences.

Proposition 30.6. *Suppose $(a_n)_{n\in\mathbb{Z}}$ is TN_2, shifted and normalized such that $a_0 = 1$.*

(1) If $a_k = 0$ for some $k > 0$ (respectively $k < 0$), then $a_l = 0$ for all $l > k$ (respectively $l < k$).
(2) If $a_n = 0$ for all $n < 0$, and the sequence $\mathbf{a}$ is TN_p for some $p \geq 2$, then its generating function $\Psi_{\mathbf{a}}$ has a non-zero radius of convergence.
(3) If $a_n = b_n = 0$ for all $n < 0$, and $\mathbf{a}, \mathbf{b}$ are TN_p sequences, then so is the sequence of Maclaurin coefficients of $\Psi_{\mathbf{a}}(x)\Psi_{\mathbf{b}}(x)$.
(4) If $a_n = 0$ for all $n < 0$, and $\mathbf{a}$ is TN, then so is the sequence of Maclaurin coefficients of $1/\Psi_{\mathbf{a}}(-x)$.

Note that the final assertion here does not go through if we assume $\mathbf{a}$ to be merely TN_p and not TN; see e.g. [294]. Also note the similarity of the first assertion to Theorem 28.4.

Proof. The first part is easy to check: given integers k, m with $a_k = 0$, note that

$$\det T_{\mathbf{a}}[(k, k+m); (0, k)] = \det \begin{pmatrix} a_k & a_0 \\ a_{k+m} & a_m \end{pmatrix} = -a_{k+m}.$$

Now if $0 < k, m$ then work with this matrix; if $0 > k, m$ then reverse its rows and columns. Either way, the non-negativity of the determinant implies $a_{k+m} = 0$ for all $m > 0$ (or all $m < 0$), as desired.

For the second part, if only finitely many terms are non-zero then the result is obvious; otherwise by the first part, $a_n > 0$ for all $n \geq 0$. Now given $n \geq 1$, we have

$$0 \leq \det T_{\mathbf{a}}[(n, n+1); (0, 1)] = \det \begin{pmatrix} a_n & a_{n-1} \\ a_{n+1} & a_n \end{pmatrix} = a_n^2 - a_{n+1}a_{n-1}.$$

From this it follows that $a_{n+1}/a_n > 0$ is non-increasing in n. Let $\beta_1 \geq 0$ be the infimum/limit of this sequence of ratios; then the power series $\Psi_{\mathbf{a}}(x) = \sum_{n\geq 0} a_n x^n$ has radius of convergence $1/\beta_1$, by basic calculus.

To show the third part, first observe that the $\mathbb{Z} \times \mathbb{Z}$ Toeplitz matrices $T_{\mathbf{a}}, T_{\mathbf{b}}$ are both TN_p by definition. Moreover, their product is a well-defined, lower-triangular Toeplitz matrix by inspection, and corresponds precisely to 'convolving' the two sequences. But this 'convolution' also corresponds to multiplying the two generating functions. Thus, it suffices to show that $T_{\mathbf{a}}T_{\mathbf{b}}$ is TN_p. This follows by the Cauchy–Binet formula (see Theorem 5.5).

Finally, for the fourth part, the reciprocal of $\Psi_{\mathbf{a}}(x)$ has a positive radius of convergence around 0, by basic calculus. Develop the reciprocal around 0, on the interval of convergence:

$$\frac{1}{\Psi_{\mathbf{a}}(x)} = \sum_{n \geq 0} b_n x^n.$$

Denote the tuple of coefficients by $\mathbf{b}$. By the arguments used to prove the previous part, this yields the lower triangular Toeplitz matrix $T_{\mathbf{b}}$ (indexed by $\mathbb{Z} \times \mathbb{Z}$) such that $T_{\mathbf{a}} T_{\mathbf{b}} = T_{\mathbf{b}} T_{\mathbf{a}} = \mathrm{Id}_{\mathbb{Z} \times \mathbb{Z}}$. Since $T_{\mathbf{a}}$ is TN, we **claim** that so is the matrix with (j,k) entry $(-1)^{j+k} b_{j-k} = (-1)^{j-k} b_{j-k}$ – which is precisely $T_{\mathbf{c}}$ for $c_n = (-1)^n b_n$. But this would correspond to the desired generating function:

$$\sum_{n \geq 0} (-1)^n b_n x^n = \frac{1}{\Psi_{\mathbf{a}}(-x)}.$$

It thus remains to prove the above claim. This requires the well-known *Jacobi complementary minor formula* [190]: given integers $0 < p < n$, an invertible $n \times n$ matrix A (over a commutative ring), and equi-sized subsets $J, K \in [n]^{p,\uparrow}$,

$$\det A \cdot \det(A^{-1})_{K^c \times J^c} = (-1)^{j_1 + k_1 + \cdots + j_p + k_p} \det A_{J \times K}, \tag{30.7}$$

where $J^c := [n] \setminus J$, and similarly for K^c.

We first quickly sketch the proof of this result (via an argument found on the internet); by pre- and post- multiplying A by suitable permutation matrices, one can reduce to the case of $J = K = [p]$ – in which case the sign on the right is $+1$. Now let A_j denote the jth column of A, and A_j^{-1} of A^{-1} respectively. Writing e_j for the standard basis of $\mathbb{R}^n$, we have:

$$A \left[e_1 | \cdots | e_p | A_{p+1}^{-1} | \cdots | A_n^{-1} \right] = \left[A_1 | \cdots | A_p | e_{p+1} | \cdots | e_n \right].$$

This can be rewritten as:

$$\begin{pmatrix} A_{J \times K} & A_{J \times K^c} \\ A_{J^c \times K} & A_{J^c \times K^c} \end{pmatrix} \begin{pmatrix} \mathrm{Id}_{p \times p} & A^{-1}_{K \times J^c} \\ 0 & A^{-1}_{K^c \times J^c} \end{pmatrix} = \begin{pmatrix} A_{J \times K} & 0 \\ A_{J^c \times K} & \mathrm{Id}_{(n-p) \times (n-p)} \end{pmatrix}.$$

Taking the determinant of both sides proves Jacobi's result for $J = K = [p]$.

To conclude the proof of the fourth part, we now use Jacobi's identity to prove the aforementioned claim (in bold). Let M be a square submatrix of $T_{\mathbf{c}}$, with $c_n = (-1)^n b_n$ as above. There exists a suitably large principal submatrix B' of $T_{\mathbf{c}}$, indexed by contiguous rows and columns, of which M is a $p \times p$ submatrix for some p. Let A be the corresponding 'contiguous' principal submatrix of $T_{\mathbf{a}}$. Multiplying every row and column of B' indexed by even numbers by -1, we obtain a matrix B such that $AB = \mathrm{Id}$. Finally, let N be the 'complementary' submatrix of A, indexed by the rows and columns not indexing M in B (or in B'). Applying Jacobi's identity and carefully keeping track of signs shows that $\det(M) \geq 0$, as desired. $\square$

We now proceed toward the form of the generating series $\Psi_{\mathbf{a}}$ for an arbitrary one-sided Pólya frequency sequence. First consider the case when $\mathbf{a}$ contains only finitely many non-zero terms, which by Proposition 30.6 must be 'consecutive'. In other words,

$$\mathbf{a} = (\dots, 0, 0, a_0, \dots, a_m, 0, 0, \dots), \qquad m \geq 0, \ a_0, \dots, a_m > 0.$$

Lemma 30.8. *With* $\mathbf{a}$ *as above, if* $m = 0$ *or* $m = 1$ *(with* a_0, a_m *arbitrary positive scalars), then* $\mathbf{a}$ *is a PF sequence. In particular, if* $\Psi_{\mathbf{a}}(x)$ *is a polynomial with all roots in* $(-\infty, 0)$*, then* $\mathbf{a}$ *is a PF sequence.*

Proof. We first make the useful observation that a block diagonal (but not necessarily square) matrix $\begin{pmatrix} A_{11} & A_{12}=0 \\ A_{21}=0 & A_{22} \end{pmatrix}$ is TN if and only if A_{11}, A_{22} are so. Indeed, the forward implication is obvious, and for the reverse one, consider any square submatrix B indexed by rows $I_1 \sqcup I_2$ and columns $J_1 \sqcup J_2$ (with A_{ij} indexed by rows I_i and columns J_j respectively, for $1 \leq i,j \leq 2$). Suppose $|I_1| > |J_1|$; then the first $|I_1|$ rows of B essentially lie in $\mathbb{R}^{J_1}$, and hence are linearly dependent. Thus $\det B = 0$. Similarly, $\det B = 0$ if $|I_1| < |J_1|$. Thus $|I_1| = |J_1|$, whence $|I_2| = |J_2|$, and hence $B = \begin{pmatrix} B_{11} & 0 \\ 0 & B_{22} \end{pmatrix}$ is block diagonal with each B_{ii} a square submatrix of A_{ii} for $i = 1,2$. As A_{ii} is TN, $\det(B) \geq 0$. Thus, A is TN.

With this observation at hand, we proceed to the proof. First note that the bi-infinite Toeplitz matrix $T_{\mathbf{a}}$ is TN if and only if the semi-infinite matrix $(T_{\mathbf{a}})_{i,j\geq 0}$ is TN, if and only each leading principal submatrix $(T_{\mathbf{a}})_{0\leq i,j\leq n}$ is TN (since every square, finite submatrix of $T_{\mathbf{a}}$ embeds inside one of the latter).

Now let $m = 0$. Then each $(T_{\mathbf{a}})_{0\leq i,j\leq n}$ is a diagonal matrix with each diagonal entry/block $a > 0$. By the preceding observation, $T_{\mathbf{a}}$ is TN for $m = 0$.

For $m = 1$, at first glance it may seem as if there are only two alternatives: either a square submatrix M of $T_{\mathbf{a}}$ has a zero row or zero column; or all entries along the main diagonal of M are a_0 (whence the sub-diagonal has all entries a_1, and all other entries in M are zero), or they are all a_1 (whence the super-diagonal has all entries a_0, and all other entries in M are zero). Both cases yield $\det M > 0$. However, there is at least one more choice: $\begin{pmatrix} a_0 & 0 \\ 0 & a_1 \end{pmatrix}$ (whose determinant is indeed positive).

Thus, here is a cleaner argument for $m = 1$. Notice that since the product of elementary matrices $E_{r,r-1}E_{r+a,r+a-1} = 0$ for all $r, a \geq 1$, one can show by induction on $n \geq 1$ that

$$\begin{aligned}(T_{\mathbf{a}})_{0\leq i,j\leq n} &= a_0 \left(\mathrm{Id}_n + \frac{a_1}{a_0} E_{21} + \cdots + \frac{a_1}{a_0} E_{n,n-1} \right) \\ &= a_0 \left(\mathrm{Id}_n + \frac{a_1}{a_0} E_{21} \right) \left(\mathrm{Id}_n + \frac{a_1}{a_0} E_{32} \right) \cdots \left(\mathrm{Id}_n + \frac{a_1}{a_0} E_{n,n-1} \right).\end{aligned} \tag{30.9}$$

As each factor is a block diagonal matrix with 1×1 blocks (1) and a unique 2×2 block $\begin{pmatrix} 1 & 0 \\ a_1/a_0 & 1 \end{pmatrix}$, each factor is TN by the above observation. Hence so is $(T_{\mathbf{a}})_{0\leq i,j\leq n}$ by (30.9) and Proposition 5.4, and hence $T_{\mathbf{a}}$ is TN as asserted.

Finally, suppose $\Psi_{\mathbf{a}}(x)$ is a polynomial with all roots in $(-\infty, 0)$. Writing this as $a_m(x + \beta_1)(x+\beta_2)\cdots(x+\beta_m)$ (with $-\beta_j$ the roots of $\Psi_{\mathbf{a}}$), we note by the preceding paragraph that $x+\beta_j = \Psi_{(\dots,0,\beta_j,1,0,\dots)}$ is the generating polynomial of a PF sequence, as is $a_m = \Psi_{(\dots,0,a_m,0,\dots)}$. Therefore, so is their product $= \Psi_{\mathbf{a}}(x)$, by Proposition 30.6(3). $\square$

Example 30.10. Let $\delta \geq 0$. Lemma 30.8 and Proposition 30.6(3) show that the finite sequence $\mathbf{a}_m$ with $\Psi_{\mathbf{a}_m}(x) = (1 + \delta x/m)^m$ is a PF sequence. Taking limits via Lemma 30.3, the sequence $a_n = \mathbf{1}_{n\geq 0}\delta^n/n!$ now forms a PF sequence, since its generating power series is $\Psi_{\mathbf{a}}(x) = e^{\delta x} = \lim_{m\to\infty}(1 + \delta x/m)^m$.

We now make some deductions from the above results in this section. Choose scalars $\delta \geq 0$, as well as $\alpha_j, \beta_j \geq 0$ for integers $j \geq 1$ such that $\sum_j (\alpha_j + \beta_j) < \infty$. Then $\prod_{j=1}^n (1 + \alpha_j x)$ generates a PF sequence for all $n \geq 1$, by Lemma 30.8 and Proposition 30.6. Take the limit as $n \to \infty$; since $|1+\alpha_j x| < e^{\alpha_j |x|}$ and $\sum_j \alpha_j < \infty$, the limit, given by

$$\prod_{j=1}^{\infty}(1+\alpha_j x)$$

is an entire function, which generates a PF sequence by Lemma 30.3. Similarly, $1/(1-\beta_j x)$ generates an infinite one-sided PF sequence for each $j \geq 1$, by Lemma 30.8 and Proposition 30.6(4). (For a far easier proof for $(1-\beta_j x)^{-1}$, one can instead use Example 32.10 below to show this assertion for $\beta_j = 1$, then deduce the case of general β_j by pre- and post-multiplying a given submatrix by diagonal matrices having suitable powers of β_j.) It follows as above that $1/\prod_{j=1}^{\infty}(1-\beta_j x)$ also generates a PF sequence. Applying Proposition 30.6(3),

$$e^{\delta x}\prod_{j=1}^{\infty}\frac{1+\alpha_j x}{1-\beta_j x}, \qquad \text{where } \alpha_j, \beta_j, \delta \geq 0, \quad \sum_{j=1}^{\infty}(\alpha_j+\beta_j) < \infty \tag{30.11}$$

is the generating function of a one-sided Pólya frequency sequence $\mathbf{a}$, with $a_0 = 1$.

Remarkably, this form turns out to encompass *all* one-sided Pólya frequency sequences. This is a deep result, shown in a series of papers [4, 5, 106, 107] by Aissen–Schoenberg–Whitney and Edrei – separately and together – and is stated here without proof.

Theorem 30.12 (Aissen–Edrei–Schoenberg–Whitney, 1951–52)**.** *A function $\sum_{n=0}^{\infty} a_n x^n$ with $a_0 = 1$, is the generating function $\Psi_{\mathbf{a}}(x)$ for a one-sided Pólya frequency sequence $\mathbf{a} = (a_0, a_1, \dots)$ if and only if $\Psi_{\mathbf{a}}$ is of the form* (30.11)*.*

(As a historical remark: Whitney was Schoenberg's student, while Aissen and Edrei were students of Pólya.) For proofs, the reader can either look into the aforementioned papers, or follow the treatment in Karlin [200, Chapter 8]; one also finds there a representation theorem for the generating function of a two-sided PF sequence. The proof involves using ideas of Hadamard from his 1892 dissertation [158], as well as Nevanlinna's refinement of Picard's theorem [270].

Remark 30.13. For completeness, we refer the reader to the recent paper [104], in which Dyachenko proves similar representation results involving the total non-negativity of generalized Hurwitz-type matrices.

Remark 30.14. Also for completeness, we mention the analogous, 'two-sided' result:

An arbitrary real sequence $(a_n)_{n\in\mathbb{Z}}$ is a Pólya frequency sequence if and only if it is either of the form $(a\rho^n)_{n\in\mathbb{Z}}$ with $a, \rho > 0$; or else its generating Laurent series converges in some annulus $r_1 < |z| < r_2$ with $0 \leq r_1 < r_2$, and has the factorization

$$\sum_{n\in\mathbb{Z}} a_n z^n = C e^{az+a'z^{-1}} z^m \prod_{j=1}^{\infty}\frac{(1+\alpha_j z)(1+\alpha'_j z^{-1})}{(1-\beta_j z)(1-\beta'_j z^{-1})},$$

where $C \geq 0$, $m \in \mathbb{Z}$, $a, a', \alpha_j, \beta_j, \alpha'_j, \beta'_j \geq 0$ and $\sum_j(\alpha_j+\beta_j+\alpha'_j+\beta'_j) < \infty$.

The 'if' part was proved by Schoenberg [322], and the harder, 'only if' part was shown by Edrei [108].

This part concludes by specializing to the case where $\Psi_{\mathbf{a}}$ has integer coefficients. Here we refer to papers by Davydov [99] and Hô Hai [177], in which they show that the Hilbert series of a quadratic R-matrix algebra (over a field of characteristic zero) generates a PF sequence. The result relevant to this text is:

Theorem 30.15 (Davydov). *A power series $\Psi_{\mathbf{a}}(x) \in 1 + x\mathbb{Z}[[x]]$ generates a PF sequence* **a** *if and only if $\Psi_{\mathbf{a}}$ has the form* (30.11)*, with $\delta = 0$ and all but finitely many α_j, β_j also zero.*

30.3. **Application 1: (Dual) Jacobi–Trudi identities.** We now provide some applications of the above results. The first is to algebra and symmetric function theory. As mentioned above, Theorem 30.15 was applied to the theory of quadratic R-matrix algebras over a field of characteristic zero. We briefly touch upon this area, starting with two examples of this phenomenon, which are Hilbert series of two well-known quadratic (and Koszul dual) algebras. Recall that for a $\mathbb{Z}^{\geq 0}$-graded algebra $A := \oplus_{n \geq 0} A_n$ with finite-dimensional graded pieces, its *Hilbert series* is $H(A, x) := \sum_{n \geq 0} x^n \dim A_n$. Now suppose V is a finite dimensional vector space, say of dimension m. The Hilbert series of its exterior algebra $\wedge^\bullet V$ is

$$H(\wedge^\bullet V, x) = (1 + x)^m,$$

and this is the generating function of a (finite, binomial) PF sequence by the above results. These results also imply the same conclusion for the Hilbert series of the symmetric algebra:

$$H(\mathbb{S}^\bullet V, x) = \frac{1}{(1 - x)^m}.$$

More generally, one fixes an operator, or *R-matrix* $R : V \otimes V \to V \otimes V$, which satisfies two conditions:

- The Yang–Baxter equation $R_{12}R_{23}R_{12} = R_{23}R_{12}R_{23}$, i.e.,
$$(R \otimes \mathrm{Id})(\mathrm{Id} \otimes R)(R \otimes \mathrm{Id}) = (\mathrm{Id} \otimes R)(R \otimes \mathrm{Id})(\mathrm{Id} \otimes R) \quad \text{on} \quad V \otimes V \otimes V.$$
- The Hecke equation $(R + \mathrm{Id})(R - q\,\mathrm{Id}) = 0$, where q is a non-zero element in the underlying ground field.

Associated to this R-matrix, define two $\mathbb{Z}^{\geq 0}$-graded algebras, by quotienting the tensor algebra by two-sided 'quadratic' ideals:

(1) The R-exterior algebra is $\wedge^\bullet_R(V) := T^\bullet(V)/(\mathrm{im}(R + \mathrm{Id}))$.
(2) The R-symmetric algebra is $\mathbb{S}^\bullet_{q,R}(V) := T^\bullet(V)/(\mathrm{im}(R - q\,\mathrm{Id}))$.

(For example, if R is the flip operator $v_1 \otimes v_2 \mapsto v_2 \otimes v_1$ and $q = 1$, then we obtain the usual exterior and symmetric algebras above.) It is well-known that the Hilbert series $H(\wedge^\bullet_R(V), x)$ and $H(\mathbb{S}^\bullet_{q,R}(V), -x)$ are reciprocals of one another. In this general setting, Davydov [99] and Hô Hai [177] showed:

Proposition 30.16. *Let V be finite-dimensional over a field of characteristic zero, q a scalar either equal to* 1 *or not a root of unity, and $R : V \otimes V \to V \otimes V$ an R-matrix as above. Then the Hilbert series $H(\wedge^\bullet_R(V), x)$, whence $H(\mathbb{S}^\bullet_{q,R}(V), x)$ (by Proposition 30.6(4)), generates a PF sequence.*

These results by Davydov and Hô Hai hold in the case when the underlying 'Iwahori–Hecke algebra' (which operates on tensor powers $V^{\otimes n}$ via the R-matrix) is semisimple, which happens when $1 + q + \cdots + q^{n-1} \neq 0$ for all $n > 0$. We add for completeness that when q is a root of unity instead, such a result was proved very recently by Skyrabin [340] under an additional hypothesis: the "1-dimensional source condition".

We now move from algebra to algebraic combinatorics. Restrict to the special case of R being the flip operator and $q = 1$; but now allow for V to have a 'multigraded' basis v_j with degree $\alpha_j x$, where $\alpha_j > 0$ for $j \geq 1$. This leads to distinguished objects in algebraic combinatorics. Indeed, the Hilbert series is now the polynomial $\Psi_{\mathbf{a}}(x) = (1 + \alpha_1 x) \cdots (1 + \alpha_m x)$, with $\alpha_1, \dots, \alpha_m > 0$ (so $a_0 = 1$) – so $a_0, a_1, \dots, a_m$ are precisely the *elementary symmetric polynomials* $a_j = e_j(\alpha_1, \dots, \alpha_m)$ in the roots α_k.

Taking limits as in the above analysis in this section, if $\alpha_j \geq 0$ and $\sum_{j=1}^\infty \alpha_j < \infty$, then $a_j = e_j(\alpha_1, \alpha_2, \dots)$ for all $j \geq 0$. Here, the mth elementary symmetric polynomial in (in)finitely many variables $u_1, u_2, \dots$ is defined to be

$$e_j(\mathbf{u}) := \sum_{1 \leq k_1 < k_2 < \cdots < k_j} u_{k_1} \cdots u_{k_j},$$

with $e_0(\mathbf{u}) := 1$ and $e_j(\mathbf{u}) := 0$ if $\mathbf{u}$ has fewer than j entries.

In fact a stronger phenomenon occurs: if we replace the α_j by variables $\mathbf{u} = (u_1, u_2, \dots)$, then every minor of the infinite triangular Toeplitz matrix

$$\begin{pmatrix} 1 & 0 & 0 & 0 & \cdots \\ e_1(\mathbf{u}) & 1 & 0 & 0 & \cdots \\ e_2(\mathbf{u}) & e_1(\mathbf{u}) & 1 & 0 & \cdots \\ e_3(\mathbf{u}) & e_2(\mathbf{u}) & e_1(\mathbf{u}) & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \tag{30.17}$$

is *monomial positive*, meaning that any minor drawn from it is a non-negative sum of monomials in the u_j. In fact an even stronger result is true: the above matrix is a (skew) Schur polynomial in the u_j. In particular, it is *(skew) Schur positive* (i.e., a non-negative sum of Schur polynomials). This phenomenon is known as the *dual Jacobi–Trudi identity*, and is in a sense, the 'original' case of numerical positivity being monomial positivity – in fact, being Schur positivity. See Appendix F for more on this. (We will see another, more recent such instance in Theorem 44.6, which follows from a more general Schur positivity phenomenon shown by Lam–Postnikov–Pylyavskyy [231].)

An analogous phenomenon holds for the 'usual' Jacobi–Trudi identity. Namely, suppose $\Psi_{\mathbf{a}}(x) = ((1-\alpha_1 x)\cdots(1-\alpha_m x))^{-1}$ for scalars $\alpha_k > 0$ (so once again, $a_0 = 1$). Then $a_0, a_1, \dots$ are precisely the *complete homogeneous symmetric polynomials*

$$a_j = h_j(\alpha_1, \dots, \alpha_m) := \sum_{1 \leq k_1 \leq k_2 \leq \cdots \leq k_j} \alpha_{k_1} \cdots \alpha_{k_j}, \qquad j \geq 1$$

and $h_0(\alpha_1, \dots, \alpha_m) := 1$. Now take limits as above, with $\alpha_j \geq 0$ and $\sum_{j=1}^\infty \alpha_j < \infty$ to obtain $a_j = h_j(\alpha_1, \alpha_2, \dots)$ for $j \geq 0$.

Once again, a stronger phenomenon than 'numerical total non-negativity' holds: if one replaces the α_j by variables $\mathbf{u} = (u_1, u_2, \dots)$ as above, then every minor of the infinite Toeplitz matrix

$$\begin{pmatrix} 1 & 0 & 0 & 0 & \cdots \\ h_1(\mathbf{u}) & 1 & 0 & 0 & \cdots \\ h_2(\mathbf{u}) & h_1(\mathbf{u}) & 1 & 0 & \cdots \\ h_3(\mathbf{u}) & h_2(\mathbf{u}) & h_1(\mathbf{u}) & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \tag{30.18}$$

is *monomial positive*, – and more strongly, (skew) Schur positive. Thus, total positivity connects to the Jacobi–Trudi identity.

30.4. **Application 2: Results of Fekete and Laguerre.** We now complete the proof of a result by Laguerre in the preceding section, as promised there. The first step is to prove a strengthening of a weaker version of Fekete's Proposition 29.23:

Proposition 30.19. *Suppose $T : \mathbb{Z}^{\geq 0} \times \mathbb{Z}^{\geq 0} \to \mathbb{R}$ is TN_p, for some integer $p \geq 2$. Given a real vector $\mathbf{x} = (x_0, \dots, x_{p-1}, 0, 0, \dots)^T$, we have $S^-(T\mathbf{x}) \leq S^-(\mathbf{x})$.*

Proof. Suppose $S^-(T\mathbf{x}) > S^-(\mathbf{x})$ for some finite vector $\mathbf{x} = (x_0, x_1, \dots)^T$ with $x_p = x_{p+1} = \cdots = 0$. Then there exists an initial segment $\mathbf{y}$ of $T\mathbf{x}$ of length $m+1$ such that $S^-(\mathbf{y}) > S^-(\mathbf{x})$. Let $\mathbf{x}' := (x_0, \dots, x_{p-1})^T$, and let the submatrix $T' := (t_{jk})_{j \leq m,\ k<p}$. Then T' is TN by assumption, and $\mathbf{y} = T'\mathbf{x}'$, so by the variation diminishing property (Theorem 29.11), $S^-(\mathbf{y}) = S^-(T'\mathbf{x}') \leq S^-(\mathbf{x}') = S^-(\mathbf{x})$, a contradiction. □

Using this result, we now prove Laguerre's result as mentioned by Fekete in [118] – but using finite matrices instead of power series:

Proof of Theorem 29.22(2). For $s > 0$, let T_s denote the infinite Toeplitz matrix with (j,k) entry $s^{j-k}/(j-k)!$ if $j \geq k$, and 0 otherwise. This corresponds to the sequence with generating function e^{sx}, so by Example 30.10 the matrix T_s is TN; moreover, the map $e^{sx} \mapsto T_s,\ s \geq 0$ is a homomorphism of monoids under multiplication.

We now turn to the proof. Applying Proposition 30.19, $S^-(T_s\mathbf{x}) \leq S^-(\mathbf{x})$ for any finite vector $\mathbf{x}$ (padded by infinitely many zeros). In particular, the integers $\{S^-(T_s\mathbf{x}) : s > 0\}$ are uniformly bounded above by $S^-(\mathbf{x}) < \infty$. Moreover, $S^-(T_s\mathbf{x})$ and $S^-(\mathbf{x})$ are precisely the number of variations in the functions $e^{sx}\Psi_{\mathbf{x}}(x)$ and $\Psi_{\mathbf{x}}(x)$, respectively.

We first show that $S^-(T_{s+t}\mathbf{x}) \leq S^-(T_s\mathbf{x})$ if $s,t > 0$. Given $m \geq 0$ and $s > 0$, let $T_s^{(m)}$ represent the leading principal $(m+1)\times(m+1)$ submatrix of T_s; and let $\mathbf{x}^{(m)} = (x_0, \dots, x_m)^T$ as above. Since $S^-(T_s\mathbf{x}) < \infty$ for all $s > 0$, there exists m such that

$$S^-(T_{s+t}\mathbf{x}) = S^-(T_{s+t}^{(m)}\mathbf{x}^{(m)}), \qquad S^-(T_s\mathbf{x}) = S^-(T_s^{(m)}\mathbf{x}^{(m)}).$$

Now compute, using for $T_t^{(m)}$ the variation diminishing property in Theorem 29.11:

$$S^-(T_{s+t}^{(m)}\mathbf{x}^{(m)}) = S^-(T_t^{(m)}T_s^{(m)}\mathbf{x}^{(m)}) \leq S^-(T_s^{(m)}\mathbf{x}^{(m)}) = S^-(T_s\mathbf{x}).$$

Finally, we claim that $S^-(T_s\mathbf{x}) < \infty$ is at least the (finite) number of positive zeros of $e^{sx}\Psi_{\mathbf{x}}(x)$, which clearly equals the number of zeros of $\Psi_{\mathbf{x}}(x)$ and hence would show the result. This follows from a variant of the (stronger) Descartes' Rule of Signs – see Theorem 10.7. □

30.5. **Location of the roots, for generating functions of finite TN_p sequences.** Recall by Lemma 30.8 that if a polynomial f has only negative (real) roots, then it generates a PF sequence. In the rest of this section, we prove the converse result: namely, if a real sequence $\mathbf{a}$ has only finitely many non-zero entries, then it is a PF sequence only if the polynomial $\Psi_{\mathbf{a}}(x)$ has all negative zeros.

To do so, we will deduce necessary (and sufficient) conditions for a finite sequence to be TN_p. The aforementioned conclusion will then follow by considering all $p \geq 1$. We begin with a result by Schoenberg [325] in *Ann. of Math.* 1955, which says: in order to check whether or not $(\dots, 0, a_0, a_1, \dots, a_m, 0, \dots)$ is a TN_p sequence, we do not need to check infinitely many minors of unbounded size.

Theorem 30.20. *Suppose $(\dots, 0, a_0, a_1, \dots, a_m, 0, \dots)$ is a real sequence with $a_0, a_1, \dots, a_m > 0$ for some $m \geq 0$. This sequence is TN_p (for some integer $p \geq 1$) if and only if the matrix*

$$A_p := \begin{pmatrix} a_0 & a_1 & \cdots & & \cdots & a_m & 0 & 0 & \cdots & 0 \\ 0 & a_0 & \cdots & & \cdots & a_{m-1} & a_m & 0 & \cdots & 0 \\ \vdots & \vdots & & & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & a_0 & \cdots & a_{m-p+1} & a_{m-p+2} & a_{m-p+3} & \cdots & a_m \end{pmatrix}_{p\times(m+p)}$$

is TN. Moreover, in this case the generating polynomial $\Psi_{\mathbf{a}}(z) = a_0 + \cdots + a_m z^m$ has no zeros in the sector $|\arg z| < p\pi/(m+p-1)$.

To show this result, Schoenberg first proved a couple of preliminary lemmas.

Lemma 30.21. *Suppose $(a_n)_{n\in\mathbb{Z}}$ is a summable sequence of non-negative numbers, such that the $p \times \mathbb{Z}$ 'matrix'*

$$A'_p := (a_{k-j})_{0\leq j<p;k\in\mathbb{Z}}$$

has rank p, i.e., an invertible $p \times p$ submatrix. Then $(a_n)_n$ is TN_p if and only if the matrix A'_p is TN.

Proof. One implication is immediate. Conversely, suppose A'_p is TN, and for every $\sigma > 0$, draw a $\mathbb{Z} \times \mathbb{Z}$ matrix from a Gaussian kernel, say $M_\sigma := (e^{-\sigma(j-k)^2})_{j,k\in\mathbb{Z}}$. By Lemma 6.8, this matrix is symmetric and totally positive, and goes entrywise to $\mathrm{Id}_{\mathbb{Z}\times\mathbb{Z}}$ as $\sigma \to \infty$. Now since $\sum_{n\in\mathbb{Z}} a_n$ is finite, one can form the matrix

$$B_\sigma := T_{\mathbf{a}} M_\sigma, \quad (B_\sigma)_{j,k} = \sum_{n\in\mathbb{Z}} a_{j-n} e^{-\sigma(n-k)^2}, \qquad j,k \in \mathbb{Z}.$$

We claim that B_σ is TP_p. To see why, it suffices to check that any contiguous submatrix of B_σ is TP_p; in turn, for this it suffices to check for all $1 \leq r \leq p$ that all contiguous $r \times r$ minors of B_σ are positive, by the Fekete–Schoenberg Lemma 4.9. But each such minor is the determinant of the product of a contiguous $r \times \mathbb{Z}$ submatrix A^* of $T_{\mathbf{a}}$ and a contiguous $\mathbb{Z} \times r$ submatrix $M^*_{\mathbb{Z}\times r}$ of the totally positive matrix M_σ. By the generalized Cauchy–Binet formula, given contiguous r-sized index-sets $J \subset [p]$ and $K \subset \mathbb{Z}$, first note that the submatrix $A^*_{J\times\mathbb{Z}}$ can be moved up and to the left, to assume that $J = [r]$ and $K \rightsquigarrow K' := K + 1 - \min_{j\in J} j$. Now,

$$\det(B_\sigma)_{J\times K} = \det(B_\sigma)_{[r]\times K'} = \sum_{J\in\binom{\mathbb{Z}}{r}} \det(A'_p)_{[r]\times J} \det(M_\sigma)_{J\times K'}.$$

The second factor in the summand is always positive since M_σ is TP, and the first factor is always non-negative, so it remains to find some J such that $(A'_p)_{[r]\times J}$ is invertible. But A'_p has rank p, so there exist p linearly independent columns indexed by $J' := \{j_1 < \cdots < j_p\}$, say. Hence the matrix $(A'_p)_{[r]\times J'}$ has full row rank, hence it contains r-many linearly independent columns. Index them by J, and we are done.

Thus the matrix B_σ is TP for all $\sigma > 0$. It is not hard to see that these matrices converge to A'_p as $\sigma \to \infty$, and the proof is complete. □

The next lemma is interesting in its own right, hence isolated into a standalone result.

Lemma 30.22. *Given matrices A, B, define their 'concatenation' $A \boxplus B$ to be the 'block diagonal' matrix $\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$, but with a horizontal shift such that the final column of A is directly above the initial column of B. Now $A \boxplus B$ is TN if and only if A, B are TN.*

This will use the following observation: *If a square block diagonal matrix $M = M_1 \oplus M_2$ is invertible, then M_1, M_2 are square.* (This is because if M_1 and hence M_2 are not square, their row (or column) ranks do not add up to the size of M, and so $\det(M) = 0$.)

Proof. One implication is immediate. Conversely, suppose A, B are TN, and the column common to them is numbered n in $A \boxplus B$. Choose a square submatrix M of $A \boxplus B$. If M does not have a column of $A \boxplus B$ from before the nth column and one from after the nth, then either M has a zero row or M is a submatrix of A or of B, whence $\det(M) \geq 0$.

Otherwise M has two columns indexed in $A \boxplus B$ by n_1, n_2 with $n_1 < n < n_2$. There are now two cases:

(1) Suppose M does not include the nth column. Then M is a submatrix of $A \boxplus B$ with the nth column removed, which is a block diagonal matrix. Now the diagonal blocks are both square, in which case $\det(M) \geq 0$ by the hypotheses; or neither diagonal block is square, in which case the above observation (following the statement of the lemma) yields $\det(M) = 0$.

(2) Otherwise M also includes the nth column of $A \boxplus B$. Now M is of the form

$$M = \begin{pmatrix} A' & v_A & 0 \\ 0 & v_B & B' \end{pmatrix}.$$

If B' is square, then expanding $\det(M)$ along the 'common' column, we claim that

$$\det(M) = \det(B') \det[A'|v_A],$$

since the cofactors in M corresponding to the coordinates of v_B are of the form $\det \begin{pmatrix} A'' & 0 \\ 0 & B'' \end{pmatrix}$, where B'' is obtained by removing a row from B' and so is not square. Hence each such cofactor vanishes, by the above observation; and we are left with $\det(B') \det[A'|v_A]$. As both determinants are non-negative by assumption, $\det(M) \geq 0$ as desired.

The proof is similar if A' is square. Finally, if neither A' nor B' is square then the cofactor against each coordinate of the 'common' column is of the form $\det \begin{pmatrix} A'' & 0 \\ 0 & B'' \end{pmatrix}$, with at least one of A'', B'' not square. By the above observation (preceding this proof), all such cofactors vanish, so $\det(M) = 0$. $\square$

With these preliminaries at hand, we can proceed.

Proof of Theorem 30.20. The first part follows from Lemma 30.21, since now most of the columns of the matrix A'_p defined in that lemma are zero. Here we use not A_p, but the matrix obtained by reversing the rows and columns of A_p, which is TN if and only if A_p is.

Now suppose A_p is TN. If $m = 0$ then the result is obvious, so we assume $m > 0$ henceforth. By Lemma 30.22, so is the $np \times (nm + np - n + 1)$ matrix

$$M_n = A_p^{\boxplus n} := A_p \boxplus \cdots \boxplus A_p, \qquad n \geq 1,$$

where the 'sum' is n times, and we use that $\boxplus$ is associative.

The first observation is that $M_n = A_p^{\boxplus n}$ has full rank np. Indeed, we may consider from each component A_p the initial p columns, which yields an upper triangular $np \times np$ submatrix with all diagonal entries $a_0 > 0$. For future use, denote the set of these np columns by J.

Let $\alpha = \rho e^{i\theta}$ be a root of $\Psi_{\mathbf{a}}(z)$. If $\theta = \pi$ then we are done, and if $\theta = 0$ then $\Psi_{\mathbf{a}}(z) > 0$. Since $\overline{\alpha}$ is also a root, we may therefore assume that $\theta \in (0, \pi)$. Define $x_j := \Im(\alpha^j) = \rho^j \sin(j\theta)$ for $0 \leq j \leq n(m+p-1)$. We now count the number of sign changes $S^-((x_0, \dots, x_t))$ for $t \geq 0$: this equals the number of times α^j crosses the X-axis in $\mathbb{C}$, so

$$S^-((x_0, \dots, x_t)) = \lfloor t\theta/\pi \rfloor + \varepsilon, \qquad \text{where } \varepsilon \in \{0, -1\}. \tag{30.23}$$

At the same time, $\sum_{j=0}^m a_j \alpha^{\nu+j} = \alpha^\nu \Psi_{\mathbf{a}}(\alpha) = 0$, so taking the imaginary parts yields:

$$\sum_{j=0}^m a_j \alpha^{\nu+j} \sin((\nu + j)\theta)) = 0, \qquad \forall \nu \in \mathbb{Z}^{\geq 0}.$$

In other words, $M_n \mathbf{x} = 0$, where $\mathbf{x} := (x_0, \dots, x_{n(m+p-1)})^T$. Also note that if some $x_j = 0$ for $0 < j < n(m + p - 1)$, then $x_{j-1}x_{j+1} < 0$, since $\theta \in (0, \pi)$ by assumption. Hence small enough perturbations to $\mathbf{x}$ will not change the number of sign-changes except at most at the extremal coordinates.

Now let $\mathbf{y} := (-1, 1, \dots, (-1)^{np})^T$ and let $\mathbf{x}_0 := (M_n)^{-1}_{[np]\times J}\mathbf{y}$, with $J \subset [nm + np - n + 1]$ the set of columns chosen above, containing a_0. Let $\widetilde{\mathbf{x}}_0 \in \mathbb{R}^{nm+np-n+1} \setminus \{0\}$ denote the vector with the coordinates of $\mathbf{x}_0$ in the positions indexed by J, and padded by zeros otherwise. Then $M_n\widetilde{\mathbf{x}}_0 = \mathbf{y}$. Now by the end of the preceding paragraph, choose $\epsilon > 0$ small enough such that $S^-(\mathbf{x} + \epsilon\widetilde{\mathbf{x}}_0) \leq S^-(\mathbf{x}) + 2$. Recalling that $M_n\mathbf{x} = 0$, we have:

$$np - 1 = S^-(\epsilon\mathbf{y}) = S^-(M_n(\mathbf{x} + \epsilon\widetilde{\mathbf{x}}_0)) \leq S^-(\mathbf{x} + \epsilon\widetilde{\mathbf{x}}_0) \leq S^-(\mathbf{x}) + 2,$$

where the first inequality is by the variation diminishing property of the TN matrix $M_n = A_p^{\boxplus n}$ (see Theorem 29.11). Hence by (30.23) with $t = n(m + p - 1)$,

$$np - 1 \leq \lfloor n(m + p - 1)\theta/\pi \rfloor + 2 \leq \frac{n(m + p - 1)\theta}{\pi} + 2.$$

Since n was arbitrary, letting $n \to \infty$ finishes the proof. $\square$

We can now deduce the desired corollary about finite Pólya frequency sequences.

Corollary 30.24. *Suppose* $\mathbf{a} = (\dots, 0, a_0, \dots, a_m, 0, \dots)$ *is a real sequence, with* $a_0, \dots, a_m > 0$. *The following are equivalent:*

(1) $\mathbf{a}$ *is a Pólya frequency sequence.*
(2) $\Psi_{\mathbf{a}}(x) = \sum_{j=0}^m a_j x^j$ *has* m *negative real roots, counted with multiplicity.*
(3) $\Psi_{\mathbf{a}}(x) = \sum_{j=0}^m a_j x^j$ *has* m *real roots, counted with multiplicity.*

Proof. That (1) $\implies$ (2) follows from the final assertion of Theorem 30.20, letting $p \to \infty$. That (2) $\implies$ (1) follows from Lemma 30.8. Clearly (2) $\implies$ (3), and the converse holds since $\Psi_{\mathbf{a}}$ does not vanish at 0, and does not have positive roots by Descartes' Rule of Signs – see e.g. Theorem 10.7 with $I = (0, \infty)$. $\square$

30.6. **Jacobi TN_p matrices and a sufficient condition for TN_p sequences.** We conclude this section with a result by Schoenberg that is 'opposite' to his Theorem 30.20. Namely, if the generating polynomial above has all roots in the sector $|\arg(z) - \pi| \leq \pi/(p + 1)$, then the finite sequence $(\dots, 0, a_0, \dots, a_m, 0, \dots)$ is TN_p. To do so, we first study when 'infinite Toeplitz tri-diagonal (Jacobi) matrices' are TN_p. This was carried out by Karlin in *Trans. Amer. Math. Soc.* (1964), and he showed the following lemma.

Lemma 30.25 (Karlin, [199])**.** *Given* $a, b, c \in (0, \infty)$, *define the corresponding Jacobi matrix* $J(a, b, c)_{\mathbb{Z}\times\mathbb{Z}}$ *via* $J(a, b, c)_{j,k}$ *equals* a, b, c *if* $k = j - 1, j, j + 1$ *respectively, and* 0 *otherwise. Then* $J(a, b, c)$ *is* TN_p *for an integer* $p \geq 1$, *if and only if* $\frac{b}{2\sqrt{ac}} \geq \cos(\pi/(p + 1))$.

Proof. When considering submatrices of $K := J(a, b, c)$ of the form $K[\mathbf{x}; \mathbf{y}]$ for $\mathbf{x}, \mathbf{y} \in \mathbb{Z}^{r,\uparrow}$, it is not hard to verify that if $x_t \neq y_t$ for some $1 \leq t \leq r$, then (a) the matrix $K[\mathbf{x}; \mathbf{y}]$ has a row or a column with at most one non-zero entry; (b) expanding along this row or column breaks up the matrix into the single non-zero entry (if it exists) and a product of two smaller minors of $K = J(a, b, c)$. From this it follows that every minor of K is a product of principal minors and elements of K. Thus, to check if $J(a, b, c)$ is TN_p, we need to examine the principal minors of size at most p. If such a submatrix is not contiguous, then it automatically becomes a block diagonal matrix, and so we only need to consider the contiguous (equivalently, leading) principal minors of $J(a, b, c)$ of size $\leq p$.

For the ensuing discussion, fix $a, c > 0$. Given $\lambda \in \mathbb{C}$, $b \in (0, \infty)$, and an integer $r \geq 1$, define $J_{r,b}(\lambda)$ to be the $r \times r$ leading principal submatrix of $J(a,b,c) - \lambda \operatorname{Id}_{\mathbb{Z}\times\mathbb{Z}}$. Expanding along an initial/terminal row/column, it is easy to observe that the determinants

$$d_{r,b}(\lambda) := \det J_{r,b}(\lambda), \qquad r = 1, 2, \dots$$

are polynomials in λ, which satisfy the recurrence:

$$d_{r+1,b}(x) = (b - x) d_{r,b}(x) - ac\, d_{r-1,b}(x), \quad r \geq 0$$

with the initial conditions $d_{0,b}(x) = 1, d_{-1,b}(x) = 0$. Reformulate this in terms of the 'shifted' polynomials

$$e_{r,b}(x) = \sqrt{ac}^{-r} d_{r,b}(b - 2\sqrt{ac}x), \quad r \geq 1,$$

to obtain the recurrence:

$$x e_{r,b}(x) = \frac{e_{r-1,b}(x) + e_{r+1,b}(x)}{2}, \ r \geq 1; \qquad e_{0,b}(x) = 1, \quad e_{-1,b}(x) = 0.$$

Restricting to $x = \lambda \in (-1, 1)$, so that $x = \cos(\theta)$ for some $\theta \in (0, \pi)$, one can show by induction that $e_r(x) = \sin((r+1)\theta)/\sin(\theta)$ satisfies the initial conditions and the recurrence. In fact, this corresponds to the Chebyshev polynomials of the second kind, given by

$$e_r(\cos\theta) = \frac{\sin(r+1)\theta}{\sin\theta}, \qquad \theta \in (0, \pi), \qquad r \geq -1.$$

Continuing with the fixed scalars $a, c > 0$, we show that $J(a,b,c)$ is TN_p for $b > 0$ if and only if $b \geq 2\sqrt{ac}\cos(\pi/(p+1))$. There are two cases. First if $b \in (0, 2\sqrt{ac})$, then write $\frac{b}{2\sqrt{ac}} = \cos(\theta) \in (0,1)$, where $\theta \in (0, \pi/2)$. Our desired value of interest is the determinant

$$d_r(0) = \sqrt{ac}^r e_{r,b}(b/2\sqrt{ac}) = \sqrt{ac}^r e_{r,b}(\cos\theta) = \sqrt{ac}^r \frac{\sin((r+1)\theta)}{\sin\theta}.$$

This is non-negative if and only if $(r+1)\theta \leq \pi$, if and only if $1 > \frac{b}{2\sqrt{ac}} \geq \cos(\pi/(r+1))$. This completes the classification if $0 < b < 2\sqrt{ac}$, since the sequence $\cos(\pi/(r+1))$ is increasing in $1 \leq r \leq p$.

The second case is if $b \geq 2\sqrt{ac} \geq 2\sqrt{ac}\cos(\pi/(p+1))$. We need to show that $J(a,b,c)$ is TN_p, for which (from above) it suffices to show that $\det J_{r,b}(0) \geq 0$ for $1 \leq r \leq p$. To show this, we use a well-known identity: by the multilinearity of the determinant in the rows/columns, it follows for any square $r \times r$ matrix B over a unital commutative ring that

$$\det(B + \nu \operatorname{Id}) = \nu^r + \sum_{\emptyset \neq I \subset [r]} \nu^{r-|I|} \det B_{I\times I}. \tag{30.26}$$

We now apply this identity to the matrices $J_{r,b}(0)$, if $b \geq 2\sqrt{ac}$. Set

$$b_r := 2\sqrt{ac}\cos(\pi/(r+1)) \in [0, b], \qquad B := J_{r,b_r}(0), \qquad \nu := b - b_r \geq 0.$$

By the analysis above in this proof, $\det B_{I\times I} \geq 0$ for all I. It follows from (30.26) that $\det J_{r,b}(0) \geq 0$. Since this holds for all $1 \leq r \leq p$, the proof is complete. $\square$

With Lemma 30.25 at hand, we present the final proof in this section: that of a sufficient condition for a polynomial to generate a finite Pólya frequency sequence.

Theorem 30.27 (Schoenberg, [326])**.** *Fix an integer $p \geq 1$, and suppose a polynomial $f(z) = a_0 + a_1 z + \cdots + a_m z^m$ with positive coefficients, such that all zeros of f lie in the sector*

$$|\arg(z) - \pi| \leq \frac{\pi}{p+1}.$$

Then the sequence $\mathbf{a} = (\dots, 0, a_0, \dots, a_m, 0, \dots)$ *is* TN_p.

Proof. If $p = 1$ then the result is clear, so we suppose below that $p \geq 2$. Decompose f into linear and complex-conjugate factors:

$$f(z) = a_m \prod_j (z - \alpha_j) \prod_k (z - \rho_k e^{-i\theta_k})(z + \rho_k e^{i\theta_k}),$$

with all $\rho_k > 0$. The hypotheses imply that $\alpha_j < 0$ and $|\theta_k - \pi| \leq \pi/(p+1)$. Now $z - \alpha_j$ generates a PF sequence by Lemma 30.8, and each irreducible quadratic

$$(z - \rho_k e^{-i\theta_k})(z + \rho_k e^{i\theta_k}) = z^2 - 2z\rho_k \cos(\theta_k) + \rho_k^2$$

does the same by Lemma 30.25, since the condition $-\cos\theta_k \geq \cos\pi/(p+1)$ is equivalent to $|\theta_k - \pi| \leq \pi/(p+1)$ (with $p \geq 2$). Hence their product $f(z)$ also generates a PF sequence, by Proposition 30.6(3). $\square$

Finally, we present a (standalone) sufficient condition for a square matrix to be TP:

Theorem 30.28. *Let* $A = (a_{jk})_{j,k\geq 1}$ *with all* $a_{jk} \in (0, \infty)$. *If*

$$a_{jk}a_{j+1,k+1} > a_{j,k+1}a_{k+1,j} \cdot 4\cos^2(\pi/(n+1)), \qquad \forall j, k \geq 1,$$

then A *is* TP_n*; moreover, the constant* $4\cos^2(\pi/(n+1))$ *cannot be reduced.*

This result only uses the positivity of the entries and a growth condition on the 2×2 minors. It was conjectured by Dimitrov–Peña in 2005 [101], and proved (independently) by Katkova–Vishnyakova in 2006 [209]. (Also worth mentioning is their 2008 follow-up paper [210] on (Hurwitz) stability of polynomials.) That the constant is best possible is revealed via Lemma 30.25, as follows: given $0 \leq c < 4\cos^2(\pi/(n+1))$, choose $\theta \in \left(\frac{\pi}{n+1}, \frac{2\pi}{n+1}\right)$ such that $c < 4\cos^2\theta$. Now consider the Jacobi matrix $J(1, 2\cos\theta, 1)_{n\times n}$.

31. Complement: Root-location results and TN Hurwitz matrices

We take a short detour to discuss some well-known matrices associated to finite Pólya frequency sequences and their generating polynomials. First, Corollary 30.24 has received recent attention: instead of working with the matrix A_p as in Theorem 30.20, one can study other variants. This section opens with a few results along these lines, presented without proof.

Definition 31.1. Given a polynomial with real coefficients

$$f(z) = a_0 z^n + a_1 z^{n-1} + \cdots + a_n, \qquad n, a_0, \dots, a_n > 0,$$

and an integer $1 \leq M \leq n$, define $a_j := 0$ for $j < 0$ or $j > n$, and the *Mth generalized Hurwitz matrix* to be the $\mathbb{Z} \times \mathbb{Z}$ matrix $H_M(f)$, given by

$$H_M(f)_{j,k} := a_{Mk-j}, \qquad j, k \in \mathbb{Z}. \tag{31.2}$$

We now present five results (without proof, and perhaps not all of them 'best possible') in the literature, which are similar to each other, and one of which is the Aissen–Edrei–Schoenberg–Whitney corollary 30.24:

Theorem 31.3. *Let $n, a_0, \dots, a_n > 0$ and define $f(z) = a_0 z^n + a_1 z^{n-1} + \cdots + a_n$. Also fix an integer $1 \leq M \leq n$.*

1. *($M = 2$, Routh–Hurwitz, see e.g.* [188, 305]*.) The polynomial f is 'stable', i.e., has no zeros z with $|\arg z| \leq \pi/2$, if and only if all leading principal minors of the Hurwitz matrix $H_2(f)$, of order $\leq n$, are positive.*
2. *($M = 2$, Asner* [14]*, Kemperman* [212]*, Holtz–Tyaglov* [181]*.) The polynomial f has no zeros z with $|\arg z| < \pi/2$ if and only if the Hurwitz matrix $H_2(f)$ is TN.*
3. *($M = 1$, Aissen–Edrei–Schoenberg–Whitney.) The polynomial f has no zeros z with $|\arg z| < \pi$ if and only if the Toeplitz matrix $H_1(f)$ is TN.*
4. *($M = n$, Cowling–Thron* [89]*.) The polynomial f has no zeros z with $|\arg z| \leq \pi/n$.*
5. *($M \in [1, n]$, Holtz–Khrushchev–Kushel* [180]*.) The polynomial f has no roots z with $|\arg z| < \pi/M$ if the generalized Hurwitz matrix $H_M(f)$ is TN.*

We now come to further results on root-location (of real polynomials), in the spirit of Theorems 30.20, 30.27, and 31.3. These three results revealed a connection between Pólya frequency sequences, totally non-negative matrices, and root-location.

As a 'warmup', we show the Gauss–Lucas theorem, found in Lucas's 1874 work [244].

Theorem 31.4 (Gauss–Lucas)**.** *If $p(z)$ is a non-constant polynomial, then the roots of $p'(z)$ in $\mathbb{C}$ are contained in the convex hull of the set of roots of $p(z)$.*

Proof. Let $p(z) = p_n \prod_{j=1}^n (z - \xi_j)$. If ξ_j is a root of p' as well as p, then $\xi_j = 1 \cdot \xi_j + \sum_{k \neq j} 0 \cdot \xi_k$. If ξ is a root of p' but not of p, then we compute:

$$\frac{p'(z)}{p(z)} = \sum_{j=1}^n \frac{1}{z - \xi_j} \quad \implies \quad 0 = \frac{p'(\xi)}{p(\xi)} = \sum_{j=1}^n \frac{\overline{\xi} - \overline{\xi}_j}{|\xi - \xi_j|^2}.$$

Setting $A := \sum_{j=1}^n |\xi - \xi_j|^{-2} > 0$, we obtain via simplifying and conjugating:

$$A\overline{\xi} = \sum_{j=1}^n |\xi - \xi_j|^{-2} \overline{\xi}_j \quad \implies \quad \xi = \sum_{j=1}^n \frac{|\xi - \xi_j|^{-2}}{A} \xi_j,$$

and so ξ is in the convex hull of the ξ_j. $\square$

The next theorem is a 'real' variant of the Hermite–Biehler theorem – the classical version is due to Hermite [165] and Biehler [49] in *J. reine angew. Math.*, in 1856 and 1879 respectively.[6] The real variant presented here requires the following preliminary result.

Lemma 31.5. *Suppose p, q are differentiable functions on a closed interval $[a, b]$, with $p(a) = p(b) = 0$, $p > 0$ on (a, b) and $q < 0$ on $[a, b]$. Then there exist $\lambda, \mu > 0$ such that $\lambda p + \mu q$ has a repeated root in (a, b).*

Proof. Since q is continuous and negative on $[a, b]$, define

$$h(x) := \frac{p(x)}{q(x)}, \qquad x \in [a, b].$$

Clearly h is differentiable on $[a, b]$, negative on (a, b), and vanishes at the endpoints. Hence it has a global minimum, say at $x_0 \in (a, b)$ – whence the function

$$h(x) - h(x_0) : [a, b] \to \mathbb{R}$$

has a repeated root at x_0. Returning to p, q, let $\lambda = 1$ and $\mu = -h(x_0) > 0$; then the function

$$\lambda p(x) + \mu q(x) = q(x)\,(h(x) - h(x_0))$$

can be easily verified to have a repeated root at $x_0 \in (a, b)$. $\square$

We also require the notion of interlacing.

Definition 31.6. Let $f, g \in \mathbb{R}[x]$ be two real-rooted polynomials, with $\deg(f) - 1 \leq \deg(g) \leq \deg(f)$. We say g *interlaces* f if between any two consecutive roots of f (possibly equal), there exists a root of g. We say f, g are *interlacing*, or *interlace (one another)* if either of f, g interlaces the other.

For the next few results, and related variants, and an in-depth treatment, the reader is referred to the monograph [295] of Rahman and Schmeisser. The treatment here is from [179].

Theorem 31.7 (Hermite–Biehler, 'real' version)**.** *Fix polynomials $p, q \in \mathbb{R}[x]$ and set $f(x) := p(x^2) + xq(x^2)$. The following are equivalent:*

1. *The polynomial $f(x)$ has no roots z with $\Re(z) \geq 0$.*
2. *The polynomials $p(-x^2), xq(-x^2)$ have real, simple roots, which are interlacing. Moreover, there exists $z_0 \in \mathbb{C}$, $\Re(z_0) > 0$ such that $\Re\left(\frac{p(z_0^2)}{z_0 q(z_0^2)}\right) > 0$.*

Proof. We begin by assuming (2). Notice that all roots of p, q lie in $(-\infty, 0]$. Thus the ratio $\Re\left(\frac{p(z^2)}{zq(z^2)}\right)$ is non-vanishing, whence always positive by (2), on the half-plane $\Re(z) > 0$ (which is an open 'sector' with aperture π; this is defined and used in a later section). It follows that the equation

$$0 = f(z) = zq(z^2)\left(\frac{p(z^2)}{zq(z^2)} + 1\right)$$

has no solution with $\Re(z) > 0$. Moreover, at a point ix on the imaginary line, we have

$$f(ix) = p(-x^2) + ixq(-x^2),$$

which cannot vanish by (2). This shows that all zeros z of f satisfy: $\Re(z) < 0$.

[6]Two historical asides: Biehler's thesis in the same year 1879 is dedicated to his "master M. Charles Hermite"; and Pierre Fatou was a student of Biehler in the Collége Stanislas.

Conversely, suppose (1) holds. Write $f(z) = a\prod_{j=1}^m (z-\alpha_j)$, with $a \neq 0$ and $\Re(\alpha_j) < 0$. Verify that if $\Im(z) > 0$, then $|iz + \overline{\alpha}_j| > |iz - \alpha_j|$ for all j, so that

$$|f(i\overline{z})| = |a|\prod_{j=1}^m |-i\overline{z} + \alpha_j| = |a|\prod_{j=1}^m |iz + \overline{\alpha}_j| > |a|\prod_{j=1}^m |iz - \alpha_j| = |f(iz)|.$$

Square both sides and expand, to deduce using that $p(x), q(x) \in \mathbb{R}[x]$:

$$0 > \frac{1}{4}(|f(iz)|^2 - |f(i\overline{z})|^2) = \Re(ip(-\overline{z}^2)zq(-z^2)) \quad \Longrightarrow \quad \Im(p(-z^2)\overline{z}q(-\overline{z}^2)) < 0.$$

This holds if $\Im(z) > 0$; taking conjugates, if $\Im(z) < 0$ then $\Im(p(-z^2)\overline{z}q(-\overline{z}^2)) > 0$. That is:

$$z \notin \mathbb{R} \quad \Longrightarrow \quad \Im(z)\Im\left(\frac{p(-z^2)}{zq(-z^2)}\right) < 0. \tag{31.8}$$

In other words, the functions

$$z \mapsto \frac{p(-z^2)}{zq(-z^2)}, \qquad z \mapsto \frac{zq(-z^2)}{p(-z^2)}$$

do not take real values for $z \in \mathbb{C} \setminus \mathbb{R}$. Thus, if the first (or second) function here equals $-\mu/\lambda$ (or $-\lambda/\mu$) for $\lambda, \mu \in \mathbb{R}$, then z must be real – that is, the following functions are real-rooted:

$$\lambda p(-z^2) + \mu zq(-z^2), \qquad \lambda, \mu \in \mathbb{R},\ \lambda^2 + \mu^2 \neq 0 \tag{31.9}$$

We next claim that the polynomials $p(-x^2), xq(-x^2)$ – which are now real-rooted – are moreover coprime. Suppose not, for contradiction. If $p(c) = q(c) = 0$ for $c \in (-\infty, 0)$ then $f(\pm i\sqrt{|c|}) = 0$, which violates (1). If $p(0) = 0$ then $f(0) = 0$, again violating (1). Otherwise $p(-x^2), xq(-x^2)$ must have a pair of common, non-real conjugate roots, say $z_\pm = a \pm ib$ with $b > 0$. Now

$$f(ia \mp b) = f(iz_\pm) = p(-z_\pm^2) + iz_\pm q(-z_\pm^2) = 0,$$

which violates (1) yet again. This contradiction shows that $p(-x^2), xq(-x^2)$ are coprime.

We now explain why no function (31.9) can have a multiple root. (Recall that these functions are all real-rooted.) Indeed, if there existed such a multiple root, then one of the ratio-functions $\frac{p(-z^2)}{zq(-z^2)}, \frac{zq(-z^2)}{p(-z^2)}$ – call it $g(x)$ – would equal a real number, say $r \in \mathbb{R}$, with 'multiplicity'. More precisely, there exists $x_0 \in \mathbb{R}$ such that $(x - x_0)^2$ divides $\lambda p(-x^2) + \mu xq(-x^2)$ with $(\lambda, \mu) \neq (0,0)$. Notice by the coprimality above that $p(-x^2), xq(-x^2)$ do not vanish at x_0. Let $k \geq 2$ denote the order of the multiple root x_0. Thus,

$$g(x) - r = (x - x_0)^k h(x),$$

where we expand near x_0 and so all functions involved are analytic; moreover, $h(x_0) \neq 0$. But then for small $\varepsilon > 0$, the equation $g(x) = r - \varepsilon^k$ has solutions

$$x = x_0 + e^{i\pi(1+2j)/k} h(x_0)^{-1/k}\varepsilon + o(\varepsilon), \qquad j = 1, 2, \dots, k.$$

This implies that the ratio-function $g(x)$ takes real values outside the real axis, which contradicts a conclusion above.

We have thus shown that $p(-x^2), xq(-x^2)$, and indeed, all nontrivial real-linear combinations of them, are real-rooted with simple roots. By (the contrapositive of) Lemma 31.5, it follows that the roots of $p(-x^2), xq(-x^2)$ interlace. Finally, return to (31.8) and let $z := iz_0$ for arbitrary $z_0 \in (0, \infty)$. Then $0 > \Im\left(\frac{p(z_0^2)}{iz_0 q(z_0)^2}\right) = \frac{-p(z_0^2)}{z_0 q(z_0)^2} \in \mathbb{R}$. $\square$

Remark 31.10. Theorem 31.7(2) is equivalent to the polynomials p, q having simple, negative roots, which interlace with the rightmost zero being that of p, and $p(0)q(0) > 0$.

As an application, we show the heart of the Routh–Hurwitz scheme [188, 305]. This will presently lead us back to TN matrices.

Theorem 31.11 (Routh–Hurwitz). *Fix polynomials $p, q \in \mathbb{R}[x]$ and set $f(x) := p(x^2) + xq(x^2)$. The following are equivalent:*

(1) The polynomial $f(x)$ has no roots z with $\Re(z) \geq 0$.
(2) The scalar $c := p(0)/q(0)$ is positive and the polynomial $f_1(x) := p_1(x^2) + xq_1(x^2)$ has no roots z with $\Re(z) \geq 0$ – where $p_1(x) := q(x)$ and $q_1(x) := \frac{1}{x}(p(x) - cq(x))$.

For completeness, we highlight the applicability of the above approach, by deducing another well-known result on interlacing. This is attributed to several authors: Hermite [165] in 1856; Kakeya, whose proof was presented by Fujiwara [133] in 1916; and Obrechkoff [273] in 1963.

Theorem 31.12 (Hermite, Kakeya, Obrechkoff). *Suppose f, g are real polynomials with no common root. The following are equivalent:*

(1) The roots of f, g are real and simple, and f, g interlace (so $|\deg(f) - \deg(g)| \leq 1$).
(2) For all $\lambda, \mu \in \mathbb{R}$ with $\lambda^2 + \mu^2 > 0$, the polynomial $\lambda f(x) + \mu g(x)$ has real, simple roots.

Proof. In both assertions, note that the $\deg(f) + \deg(g)$ roots of f, g are pairwise distinct and all real, so say f (respectively, g) has roots $\alpha_1 < \cdots < \alpha_m$ (respectively, $\beta_1 < \cdots < \beta_n$). These divide the real line into $m + n + 1$ open intervals, on each of which f, g do not change sign. Enumerate these intervals from right to left, so that $I_1 = (\max\{\alpha_m, \beta_n\}, \infty)$. Moreover, on any two adjacent intervals, one of f, g does not change sign, while the other does.

We now turn to the proof. First given (2), we need to show that the polynomials f, g interlace. (Note that both have real, simple roots.) This follows from the claim that no bounded interval I_k has both endpoints as roots of either f or of g. In turn, the claim is a consequence of Lemma 31.5.

Conversely, we assume (1) and show (2). We can assume both $\lambda, \mu \neq 0$, so suppose without loss of generality that (a) $\lambda = 1$, (b) f, g are monic, and (c) $\deg(f) = m \geq n = \deg(g)$, with $n \in \{m-1, m\}$. There are now several cases:

(1) $\mu > 0$. In this case, the function $\lambda f + \mu g$ is positive on I_1, negative on I_3, positive on I_5, and so on, until it has sign $(-1)^{m-1}$ on I_{2m-1}, and $(-1)^m$ as $x \to -\infty$. From this, it follows that $\lambda f + \mu g$ has at least m sign changes on $\mathbb{R}$, and degree m, whence precisely m simple roots.
(2) $\mu < 0$ and $m = n + 1$. In this case, $\lambda f + \mu g$ is positive as $x \to +\infty$, negative on I_2, positive on I_4, and so on, until it has sign $(-1)^m$ on I_{2m}. Now the final sentence of the previous case again applies.
(3) $-1 < \mu < 0$ and $m = n$. Now there are two subcases, corresponding to if $\alpha_m > \beta_n$ or $\alpha_m < \beta_n$. In the former case, $\lambda f + \mu g$ is positive as $x \to +\infty$, negative on I_2, positive on I_4, and so on, until it has sign $(-1)^m$ (as in the preceding case). Hence the final sentence of the first case again applies.

    Otherwise we have $\alpha_m < \beta_n$, whence $\lambda f + \mu g$ is positive on I_2, negative on I_4, and so on, until it has sign $(-1)^{m-1}$ on I_{2m}, and sign $(-1)^m$ as $x \to -\infty$. Hence the final sentence of the first case applies.
(4) If $\mu < -1$ and $m = n$, this reduces to the preceding case, by replacing $f \longleftrightarrow g$ and $(\lambda = 1, \mu \in (-\infty, -1)) \longleftrightarrow (\mu^{-1} \in (-1, 0), 1)$.

(5) This leaves the final subcase, in which $\mu = -1$ and $m = n$ (and recall that $\lambda = 1$ and f, g are monic). Given the interlacing of the roots of f, g, it is not hard to see that $\deg(\lambda f + \mu g) = m - 1$. Notice that shifting the origin simultaneously for both polynomials does not affect either assertion in the theorem, nor does interchanging $\lambda = 1$ with $\mu = -1$. Thus, we assume henceforth that both f and g have negative, simple roots α_j and β_j respectively, and that without loss of generality,

$$0 > \alpha_m > \beta_m > \alpha_{m-1} > \cdots > \alpha_1 > \beta_1.$$

We now 'invert the coefficients' of both polynomials: let $f_{inv}(x) := x^m f(1/x)$ and $g_{inv}(x) := x^m g(1/x)$. These have roots α_j^{-1} and β_j^{-1} respectively, so these roots are once again negative and interlacing (now β_1^{-1} is the closest root to the origin). Moreover, $f_{inv}(0) = g_{inv}(0) = 1$. Applying Remark 31.10 (and Theorem 31.7), the polynomial

$$F(x) := g_{inv}(x^2) + x f_{inv}(x^2)$$

has no roots z with $\Re(z) \geq 0$. Hence by Theorem 31.11 with $c = 1$, the polynomial $F_1(x) := g_1(x^2) + x f_1(x^2)$ has no roots z with $\Re(z) \geq 0$, where

$$g_1(x) = f_{inv}(x), \qquad f_1(x) = \frac{1}{x}(g_{inv}(x) - f_{inv}(x)).$$

Now apply Theorem 31.7 for F_1, and Remark 31.10 for g_1, f_1, to deduce that the roots of g_1, f_1 are simple, negative, and interlace. In particular, the roots of $\frac{1}{x}(g_{inv}(x) - f_{inv}(x)) = (g - f)_{inv}(x)$ (by abuse of notation) are simple and negative. Inverting back the coefficients (via $p(x) \mapsto x^{\deg(p)} p(1/x)$), so are the roots of $g(x) - f(x)$, whence of $\lambda f + \mu g$, as desired. $\square$

32. Examples of Pólya frequency functions: Laplace transform, convolution.

In Section 28, we saw some characterizations of TN_p functions, and also studied the exponential decay of TN_2 (whence all TN_p) functions. We also saw several examples of TN (in fact, Pólya frequency) functions in Section 29. We now return to Pólya frequency functions, and discuss additional examples as well as a recipe to generate new examples of TN_p or PF functions from old ones.

32.1. **The bilateral Laplace transform of a totally non-negative function.** We begin by defining and studying the Laplace transform more generally – for TN_2 functions.

Definition 32.1. The *bilateral Laplace transform* of a (measurable) function $f : \mathbb{R} \to \mathbb{R}$ is denoted by $\mathcal{B}(f)$, and defined at a complex argument $s \in \mathbb{C}$ to be

$$\mathcal{B}(f)(s) := \int_{\mathbb{R}} e^{-sx} f(x)\ dx.$$

This expression is defined if the following integrals both converge as $R \to \infty$:

$$\int_0^R e^{-sx} f(x)\ dx, \qquad \int_{-R}^0 e^{-sx} f(x)\ dx,$$

in which case the sum of their limits is taken to be $\mathcal{B}(f)(s)$.

The following result uses the characterization of TN_2 functions in Theorem 28.4 above, to show the existence of the Laplace transform:

Lemma 32.2. *Suppose $f : \mathbb{R} \to \mathbb{R}$ is TN_2 and not an exponential. Then $\mathcal{B}(f)$ exists in the open vertical strip in $\mathbb{C}$, given by*

$$-\infty \leq \alpha := \inf_{x \in I} (\log f)'(x) < \beta := \sup_{x \in I} (\log f)'(x) \leq \infty,$$

where f' exists on a co-countable set, and we set $\alpha := -\infty$ (respectively, $\beta := \infty$) if $f(x) \equiv 0$ for sufficiently large $x > 0$ (respectively, sufficiently small $x < 0$). If f is integrable then this strip contains the imaginary axis.

Proof. Let I denote the interval of support of f, as in Theorem 28.4. There are three possibilities: (a) I is bounded, in which case the result is easy; (b) I is unbounded only on one side (in which case f can be an exponential function on I); or (c) $I = \mathbb{R}$, in which case $f(x)$ is not an exponential.

We will work in the third case, as the cases for $I \subsetneq \mathbb{R}$ are simpler. It suffices to show that the integral in $\mathcal{B}(f)$ is absolutely convergent on a vertical strip. We now appeal to Proposition 28.8 and its proof, used henceforth without further reference. First by Lemma 26.3, $(\log f)'$ is defined on a co-countable subset of I and is non-increasing there. Thus the following limits make sense, and equal the asserted formulae:

$$\alpha := \lim_{x \to \infty} \frac{f'(x)}{f(x)}, \qquad \beta := \lim_{x \to -\infty} \frac{f'(x)}{f(x)}; \tag{32.3}$$

moreover, $-\infty \leq \alpha < \beta \leq \infty$, since f is not an exponential on $I = \mathbb{R}$.

We claim that the integral in $\mathcal{B}(f)(s)$ is absolutely convergent for $\Re(s) \in (\alpha, \beta)$, whence convergent as desired. Indeed:

$$|\mathcal{B}(f)(s)| \leq \int_{\mathbb{R}} |e^{-sx}| f(x)\ dx = \int_0^\infty e^{-x\Re(s)} f(x)\ dx + \int_{-\infty}^0 e^{-x\Re(s)} f(x)\ dx.$$

Since $\alpha < \Re(s) < \beta$, choose points $x_1 < x_2$ in I such that $\alpha < g'(x_2) < \Re(s) < g'(x_1) < \beta$. Then convergence follows because f shrinks to zero faster than the exponential bounds in the proof of Proposition 28.8.

Finally, suppose f is integrable on I, where we once again assume $I = \mathbb{R}$. Then $f(x) \to 0$ as $|x| \to \infty$, whence $\log f(x) \to -\infty$ as $|x| \to \infty$. It follows that $\alpha < 0 < \beta$. $\square$

We next bring the limits of the vertical strip α, β into the form found in the literature:

Proposition 32.4. *The limits in Lemma 32.2 can also be written as*

$$\alpha := \lim_{x\to\infty} \frac{\log f(x)}{x}, \qquad \beta := \lim_{x\to-\infty} \frac{\log f(x)}{x},$$

where once again, we set $\alpha := -\infty$ (respectively, $\beta := \infty$) if $f(x) \equiv 0$ for sufficiently large $x > 0$ (respectively, sufficiently small $x < 0$).

Proof. The result is nontrivial only for I unbounded on one or both sides of the origin; we show it here only for the case $I = \mathbb{R}$. Since $-\log f$ is convex, the result follows from a more general fact about arbitrary convex functions:

Suppose $g : \mathbb{R} \to \mathbb{R}$ is convex, so that g' is defined on a co-countable set. Then,

$$\sup_{x\in\mathbb{R}} g'(x) = \lim_{x\to\infty} g'(x) = \lim_{x\to\infty} \frac{g(x)}{x},$$
$$\inf_{x\in\mathbb{R}} g'(x) = \lim_{x\to-\infty} g'(x) = \lim_{x\to-\infty} \frac{g(x)}{x}.$$

We only show the first part; the second is similar. By Lemma 26.3(1), the divided difference

$$h(x, y) := \frac{g(y) - g(x)}{y - x}, \qquad x \neq y$$

is coordinatewise non-decreasing. Choose any x_0 at which g is differentiable. Then for $y > x_0$,

$$h(x_0, y) \geq \lim_{y\to x_0^+} h(x_0, y) = g'(x_0);$$

but now taking $y \to \infty$,

$$g'(x_0) \leq \lim_{y\to\infty} \frac{g(y) - g(x_0)}{y - x_0} = \lim_{y\to\infty} \frac{g(y)}{y}.$$

Taking the supremum over $x_0 \in \mathbb{R}$ (or the limit as $x_0 \to \infty$) yields one inequality. For the other, let g be differentiable at x_0 and let $y < x_0$. Then

$$h(y, x_0) \leq \lim_{y\to x_0^-} h(y, x_0) = g'(x_0) \leq \sup_{x\in\mathbb{R}} g'(x).$$

Since this holds for all $x_0 > y$, now taking $x_0 \to \infty$ yields the desired result:

$$\sup_{x\in\mathbb{R}} g'(x) \geq \lim_{x_0\to\infty} \frac{g(x_0) - g(y)}{x_0 - y} = \lim_{x_0\to\infty} \frac{g(x_0)}{x_0}.$$ $\square$

32.2. **Examples of Pólya frequency functions; convolution.** Having discussed the Laplace transform, we next discuss a recipe to generate new examples of TN_p functions (or Pólya frequency functions) from old ones, for all $p \geq 1$.

Definition 32.5. Given Lebesgue measurable functions $f, g : \mathbb{R} \to \mathbb{R}$, define their *convolution*, denoted by $f * g$, to be the function given by the following integral, wherever defined:

$$(f * g)(x) := \int_{\mathbb{R}} f(y)g(x - y)\ dy.$$

The recipe can now be stated: if f, g are integrable TN_p functions for any $1 \leq p \leq \infty$, then so is $f * g$. To show this, we require some basic properties of convolutions, which are now collected together for ease of future reference.

Lemma 32.6 (Convolution properties). *Suppose $f, g : \mathbb{R} \to \mathbb{R}$ are both in $L^1(\mathbb{R})$.*

1. *For almost all $x \in \mathbb{R}$, the function $y \mapsto f(y)g(x-y)$ is Lebesgue measurable and integrable, so that $f * g$ is defined for almost every $x \in \mathbb{R}$.*
2. *$f * g = g * f \in L^1(\mathbb{R})$, and $\|f * g\|_1 \leq \|f\|_1 \|g\|_1$.*
3. *If g is also in $L^\infty(\mathbb{R})$, then $f * g$ is continuous on $\mathbb{R}$.*
4. *If $f, g \geq 0$ on $\mathbb{R}$, then so is $f * g$.*
5. *The Laplace transform is an 'algebra homomorphism' for addition and convolution:*
$$\mathcal{B}(f * g)(s) = \mathcal{B}(f)(s)\mathcal{B}(g)(s),$$
whenever the two terms on the right converge absolutely at a common point $s \in \mathbb{R}$.

Regarding the last part, we leave to the reader the verification that $L^1(\mathbb{R})$ under addition and convolution forms a commutative $\mathbb{R}$-algebra.

Proof.

(1) First note that $A(x,y) := f(x)g(y)$ is Lebesgue measurable, since f, g are. Moreover, $L(x,y) := (y, x-y)$ is an invertible linear transformation of $\mathbb{R}^2$, whence measurable. Thus $(A \circ L)(x,y) := f(y)g(x-y)$ is measurable. But now $A \circ L$ is also integrable:
$$\iint_{\mathbb{R}^2} |(A \circ L)(x,y)| \; dx \; dy = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} |g(x-y)| \; dx \right) |f(y)| \; dy = \|g\|_1 \|f\|_1 < \infty.$$
Hence the assertion follows by Fubini's theorem.

(2) This is now straightforward:
$$\|f * g\|_1 = \int_{\mathbb{R}} |(f * g)(x)| \; dx \leq \iint_{\mathbb{R}^2} |g(x-y)f(y)| \; dx \; dy,$$
and as in the preceding part, the right-hand side equals $\|g\|_1 \|f\|_1$. That $(f * g)(x) = (g * f)(x)$ follows by substituting $y \mapsto x - y$ in the (improper) integral defining the convolution.

(3) Suppose $x_n \to x$ in $\mathbb{R}$ as $n \to \infty$. Then
$$|(f * g)(x_n) - (f * g)(x)| \leq \int_{\mathbb{R}} |f(x_n - y) - f(x - y)||g(y)| \; dy \leq \|\tau_{h_n} f - f\|_1 \|g\|_\infty, \quad (32.7)$$
where h_n is a real sequence going to 0, and $(\tau_h f)(y) := f(y+h)$ for $y, h \in \mathbb{R}$ is the shift operator. Now recall via Urysohn's lemma and properties of the Lebesgue measure that the space of compactly supported functions $C_c(\mathbb{R})$ is dense in $(L^1(\mathbb{R}), \|\cdot\|_1)$. Thus, let $f_k \in C_c(\mathbb{R})$ converge to f as $k \to \infty$; then by the triangle inequality,
$$\|\tau_{h_n} f - f\|_1 \leq \|\tau_{h_n} f - \tau_{h_n} f_k\|_1 + \|\tau_{h_n} f_k - f_k\|_1 + \|f_k - f\|_1. \quad (32.8)$$

The first and third terms on the right agree, since the Lebesgue measure is translation-invariant. Thus, to show the left side of (32.7) goes to zero as $n \to \infty$, it suffices to show that the right side of (32.8) goes to zero. For this, fix $\epsilon > 0$, then fix $k \gg 0$ such that $\|f_k - f\|_1 < \epsilon/3$. Suppose f_k is supported on $[-\rho, \rho]$ for $0 < \rho < \infty$. Choose $n_0 \gg 0$ such that $|h_n| < \rho$ for $n \geq n_0$; then $\tau_{h_n} f_k - f_k$ is continuous and supported on $J := [-2\rho, 2\rho]$. Since f_k is uniformly continuous on J, there exists $\delta > 0$ such that
$$x, y \in J, \; |x - y| < \delta \implies |f_k(x) - f_k(y)| < \frac{\epsilon}{9\rho}.$$

Now choose $n_1 > n_0$ such that $|h_n| < \min(\rho, \delta)$ for $n \geq n_1$. Then $\tau_{h_n} f_k$ and f_k disagree at most on the interval $[-2\rho, \rho]$ if $h_n \geq 0$, and on $[-\rho, 2\rho]$ if $h_n \leq 0$. Hence

$$\|\tau_{h_n} f_k - f_k\|_1 = \int_{\mathbb{R}} |f_k(y + h_n) - f_k(y)| \, dy \leq 3\rho \cdot \frac{\epsilon}{9\rho} = \frac{\epsilon}{3}$$

for all $n \geq n_1$. Using (32.8), it thus follows for each $\epsilon > 0$ that $\|\tau_{h_n} f - f\|_1 < \epsilon$ for all sufficiently large n. This shows continuity on $\mathbb{R}$, by (32.7).

(4) This is immediate from the definition of $f * g$.

(5) We compute:

$$\begin{aligned}\mathcal{B}(f)(s)\mathcal{B}(g)(s) &= \int_{\mathbb{R}} f(y) \, dy \int_{\mathbb{R}} e^{-s(y+u)} g(u) \, du = \int_{\mathbb{R}} f(y) \, dy \int_{\mathbb{R}} e^{-sx} g(x-y) \, dx \\ &= \int_{\mathbb{R}} e^{-sx} \left(\int_{\mathbb{R}} f(y) g(x-y) \, dy \right) \, dx = \int_{\mathbb{R}} e^{-sx} (f * g)(x) \, dx = \mathcal{B}(f * g)(s),\end{aligned}$$

where $f * g$ is defined almost everywhere from above. The interchange of integrals in the first equality on the second line is justified by Fubini's theorem – which applies here because

$$\iint_{\mathbb{R}^2} |f(y) e^{-sx} g(x-y)| \, dy \, dx = \int_{\mathbb{R}} e^{-sy} |f(y)| \, dy \cdot \int_{\mathbb{R}} e^{-su} |g(u)| \, du,$$

and both integrals are finite by assumption. ☐

As an immediate consequence, the above recipe follows:

Corollary 32.9. *Suppose $f, g : \mathbb{R} \to \mathbb{R}$ are integrable TN_p functions for some $p \geq 1$ (or both TN functions). Then so is $f * g$.*

Proof. That $f * g$ is integrable follows from Lemma 32.6. That it is TN_p follows from the Basic Composition Formula (see (5.14)) and Corollary 6.1. ☐

Two applications of this corollary will be provided presently.

Having studied the Laplace transform and convolution, we now come to additional examples of TN and Pólya frequency functions.

Example 32.10. The Heaviside function $H_1(x) := \mathbf{1}_{x \geq 0}$ is TN. This can be shown using direct computations, see e.g. Chapters 1, 3 of Karlin's book [200], where it is shown that the 'transpose' kernel $K(x, y) = \mathbf{1}_{x \leq y}$ satisfies:

$$\det K[\mathbf{x}; \mathbf{y}] = \mathbf{1}(x_1 \leq y_1 < x_2 \leq y_2 < \cdots < x_p \leq y_p),$$

for all integers $p \geq 1$ and tuples $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{p,\uparrow}$.

Example 32.11. We are now interested in convolving the previous example with itself several times. However, the function H_1 is not integrable. Thus, first use Lemma 28.3 to define

$$\lambda_1(x) := e^{-x} H_1(x) = \mathbf{1}_{x \geq 0} e^{-x}.$$

Note, this is an integrable TN function on $\mathbb{R}$, which is discontinuous at the origin. We now claim that *the n-fold convolution f_n of λ_1 with itself is the function $\frac{x^{n-1}}{(n-1)!} \lambda_1(x)$, for all $n \geq 1$.*

The verification is by induction on $n \geq 1$, with the base case immediate. To show the induction step, use that $f_n(x) := x^{n-1} \lambda_1(x)/(n-1)!$, and compute:

$$f_{n+1}(x) = \int_{\mathbb{R}} f_n(y) f_1(x - y) \, dy.$$

By the induction hypothesis, the integrand vanishes unless $y, x-y$ are both non-negative, whence so is x. Thus $f_{n+1}(x) = 0$ for $x < 0$, and for $x \geq 0$, we compute:

$$f_{n+1}(x) = \int_0^x f_n(y) f_1(x-y)\ dy = \int_0^x \frac{y^{n-1}}{(n-1)!} e^{-y} \cdot e^{-(x-y)}\ dy = e^{-x} \frac{x^n}{n!}.$$

In particular, $x^n \lambda_1(x)$ is an integrable TN function, by Corollary 32.9. Similar to the Gaussian, we record the Laplace transform of these functions, for future use. More generally, given a non-negative power $\alpha \geq 0$ and a scalar $\beta > 0$, let $g_{\alpha,\beta}(x) := x^\alpha \lambda_1(\beta x)$. Then,

$$\mathcal{B}(g_{\alpha,\beta}) = \int_0^\infty e^{-xs} x^\alpha e^{-\beta x}\ dx = \int_0^\infty e^{-x(s+\beta)} x^\alpha\ dx = \frac{\Gamma(\alpha+1)}{(s+\beta)^{\alpha+1}}, \qquad s > -\beta. \tag{32.12}$$

These examples will play a role below, in classifying the total-positivity preservers on arbitrary domains. For now we present a final example, again obtained via convolution:

Example 32.13. Let $f(x) := \lambda_1(x)$ and $g(x) := \lambda_1(-x)$. As these are integrable TN functions, so is their convolution, which one verifies is $e^{-|x|}/2$. Hence $e^{(\alpha-\beta)|x|/2}$ is TN for all $\alpha < \beta$. Multiplying by $e^{(\alpha+\beta)x/2}$, it follows by Lemma 28.3 that the function

$$f(x) = \begin{cases} c e^{\beta(x-x_0)}, & \text{if } x \leq x_0, \\ c e^{\alpha(x-x_0)}, & \text{if } x > x_0 \end{cases} \tag{32.14}$$

is TN for $c \geq 0$ and $x_0 \in \mathbb{R}$ – and integrable when $\alpha < 0 < \beta$, as above. Notice also that the limiting cases of $\alpha = -\infty$, $\beta = +\infty$ (both leading to f vanishing on a semi-axis), are integrable TN functions; while if $\alpha = \beta$ then f is an exponential, whence also TN.

32.3. **Pólya frequency functions and the Laguerre–Pólya class.** It is rewarding to place the theory of Pólya frequency functions (and more generally, TN functions) in its historical context before proceeding further. Begin with a scalar $\delta \geq 0$ and a summable positive sequence:

$$\alpha_j > 0,\ j = 1, 2, \dots, \qquad \sum_{j \geq 1} \alpha_j < \infty,$$

so that the terms $1/\alpha_j$ are bounded below by a positive number. Then the convolution

$$f_n(x) := (\varphi_{\alpha_1} * \cdots * \varphi_{\alpha_n})(x), \qquad \varphi_a(x) := \frac{1}{a} \lambda_1(x/a)$$

is a Pólya frequency function with Laplace transform $\prod_{j=1}^n (1 + \alpha_j s)^{-1}$. Given $\delta \geq 0$, the PF function $\Lambda_n(x) := f_n(x - \delta)$ therefore satisfies:

$$\mathcal{B}(\Lambda_n)(s) = \Phi_n^*(s),\ \forall n \geq 1,\ \Re(s) > \max_{j \leq n}(-1/\alpha_j), \qquad \text{where } \Phi_n^*(s) := \frac{e^{-\delta s}}{\prod_{j=1}^n (1 + \alpha_j s)}.$$

Notice for $n \geq 2$ that the Φ_n^* are dominated uniformly on the imaginary axis by an integrable function:

$$|\Phi_n^*(ix)| \leq \frac{1}{|(1 + \alpha_1 ix)(1 + \alpha_2 ix)|}, \qquad \forall x \in \mathbb{R},\ n \geq 2. \tag{32.15}$$

Hence for $n \geq 2$, the Laplace inversion formula recovers Λ_n from Φ_n^* via the Fourier–Mellin integral, which converges absolutely:

$$\Lambda_n(x) = \frac{1}{2\pi i} \lim_{T \to \infty} \int_{-iT}^{iT} e^{sx} \Phi_n^*(s)\ ds, \qquad \forall x \in \mathbb{R}.$$

Now since $\sum_{j\geq 1} \alpha_j < \infty$, we have the convergence of the functions

$$\Phi_n^*(s) \to \Phi^*(s) := \frac{e^{-\delta s}}{\prod_{j=1}^\infty (1+\alpha_j s)},$$

where the infinite product in $\Phi^*(s)$ is defined in the vertical strip $\max_j(-1/\alpha_j) < \Re(s) < \infty$, and converges there since

$$\prod_{j=1}^\infty |1+\alpha_j s| \leq \prod_{j=1}^\infty (1+\alpha_j |s|) \leq \exp \sum_{j=1}^\infty (\alpha_j |s|) < \infty.$$

Hence by Lebesgue's dominated convergence theorem, the integrals Λ_n also converge to a function:

$$\lim_{n\to\infty} \Lambda_n(x) = \lim_{n\to\infty} \frac{1}{2\pi i} \lim_{T\to\infty} \int_{-iT}^{iT} e^{sx} \Phi_n^*(s)\ ds = \frac{1}{2\pi i} \lim_{T\to\infty} \int_{-iT}^{iT} e^{sx} \Phi^*(s)\ ds,$$

and this holds for all real x. Denote the function on the right by $\Lambda(x)$; then one can show that Λ is also a Pólya frequency function, which vanishes on $(-\infty, 0)$ and is such that $\mathcal{B}(\Lambda)(s) = \Phi^*(s)$ for $\Re(s) > \max_j(-1/\alpha_j)$. Moreover, the reciprocal of this bilateral Laplace transform is the restriction to the strip $\Re(s) > \max_j(-1/\alpha_j)$ of an entire function with only (real) negative zeros:

$$\frac{1}{\mathcal{B}(\Lambda)(s)} = \frac{1}{\Phi^*(s)} = e^{\delta s} \prod_{j=1}^\infty (1+\alpha_j s). \tag{32.16}$$

A similar phenomenon occurs when one considers Pólya frequency functions that need not vanish on a semi-axis. In this case one can convolve functions of the form $\varphi_a(\pm x)$ for $0 \neq a \in \mathbb{R}$, as well as the Gaussian kernel (and shifted variants of these). Further taking limits produces Pólya frequency functions whose Laplace transforms are of the form

$$\Phi^*(s) = \frac{e^{\gamma s^2 - \delta s}}{\prod_{j=1}^\infty (1+\alpha_j s) e^{-\alpha_j s}}, \tag{32.17}$$

where

$$\gamma \in [0,\infty),\ \alpha_j, \delta \in \mathbb{R},\ 0 < \gamma + \sum_j \alpha_j^2 < \infty, \quad \max_{\alpha_j > 0}(-1/\alpha_j) < \Re(s) < \min_{\alpha_j < 0}(-1/\alpha_j).$$

Schoenberg showed in *J. d'Analyse Math.* (1951) the following remarkable result: the above toy examples (32.16) and (32.17) (created by convolving variants of λ_1 and the Gaussian) are in fact representative of *all* Pólya frequency (PF) functions Λ satisfying $\int_{\mathbb{R}} \Lambda(x)\ dx = 1$ – with the first toy example a prototype for all PF functions that vanish on $(-\infty, 0)$. E.g. the PF functions Λ as in (32.16) are characterized by the fact that $\frac{1}{\mathcal{B}(\Lambda)(s)}$ is (the restriction to a vertical strip, of) an entire function $ce^{\delta s} \prod_{j=1}^\infty (1+\alpha_j s)$, where $c \in (0,\infty)$, $\delta, \alpha_j \geq 0$, and $\sum_j \alpha_j < \infty$.

In fact, such entire functions were the subject of a beautiful theory built up around the turn of the 20th century, by experts before Schoenberg – including Laguerre, Pólya, and Schur. In the next section, we provide a brief detour into this rich area, before returning to its connections to Pólya frequency functions and TN functions.

33. History – The Laguerre–Pólya–Schur program: multiplier sequences and entire functions. Modern results.

This section undertakes a brief historical journey through one of the most longstanding and mathematically active areas in analysis, with a rich history as well as modern activity: the study of the zeros of (complex) polynomials and entire functions. We already saw some classical results in Section 31; here we see more such results, now from the viewpoint of linear operators on polynomial spaces that preserve real-rootedness and similar properties.

The study of roots of complex polynomials has always attracted tremendous attention. To name two dozen experts with related work before 1930: Descartes (1637); Budan, Gauss, Fourier, Sturm, Cauchy (1800–1840); Chebyshev, Hermite, Poulain, Weierstrass, Routh, Biehler, Lucas (1840–1880); Laguerre, Hadamard, Maló, Markov, Hurwitz, Grace, Van Vleck (1880–1910); Fekete, Kakeya, Pólya, Jensen, Schur, Cohn, Szegő, Walsh, Obrechkoff (1910–1930). For these and many other classical contributions, see e.g. the 1929 survey [351] by Van Vleck in *Bull. Amer. Math. Soc.* In the subsequent nine decades, activity in this area has continued, including papers, surveys, and books; in this section, we briefly allude to the works [91, 92] by Craven–Csordas (and Smith) and the classic text of Levin [235]. (See the monograph [295] by Rahman–Schmeisser for more on this area.) The section concludes by alluding to a few important contributions to this area, all taken from this millennium.

We begin with notation.

Definition 33.1.

(1) Given a region $S \subset \mathbb{C}$, let $\pi(S)$ denote the class of polynomials with all zeros in S, and coefficients in $\mathbb{R}$ (sometimes this is replaced by $\mathbb{C}$). Given an integer $n \geq 1$, let $\pi_n(S) \subset \pi(S)$ denote the subset of polynomials with degree at most n.
(2) Given a complex polynomial p, let $Z_{nr}(p)$ denote the number of non-real roots of p.

A question that has interested analysts for more than a century is to understand operations – even linear ones – under which $\pi_n(S)$ is stable. This is an old question for which not many nontrivial answers were known – especially until 2004; see the discussion preceding Theorem 33.40 below. Certainly, some easy answers have long been known. For example, if $p(x) \in \pi(\mathbb{C}) = \mathbb{R}[x]$ is real-rooted, then so are:

(1) its product $p(x)q(x)$ with a real-rooted polynomial $q(x)$.
(2) the 'shift' $ap(bx + c)$ for scalars $a, b, c \in \mathbb{R}$, $a \neq 0$.
(3) the derivative $p'(x)$ – this is Rolle's theorem. Notice, the derivative operator commutes with all additive shifts/translations.
(4) 'multiplicative differentiation' $xp'(x)$, again by Rolle's theorem. In contrast to the preceding operation, this operator commutes with all multiplicative shifts/dilations.
(5) the 'inversion' $x^{\deg(p)}p(1/x)$, whose roots are 0 if $p(0) = 0$, and $x_0 \neq 0$ if $p(1/x_0) = 0$.

All but the first of these operations also are answers to the more general question, of understanding linear transformations on $\pi_n(\mathbb{C})$ that do not increase the number of non-real roots $Z_{nr}(\cdot)$. In particular, they answer our **first question**, of understanding linear operators preserving real-rootedness. A related, **second question** involves understanding which (linear) operations T preserve real-rootedness, now only on real polynomials with all non-positive roots (or all non-negative roots). Notice that the operations above are also (positive) examples of such linear operations (e.g. assuming $q(x)$ also has one-sided roots).

Often, one assumes a further restriction on the linear map T in order to have more structure to work with. Two such conditions are that T commutes with the usual/additive derivative

∂, or with the multiplicative derivative $x\partial$. Clearly, the latter operators are of the form

$$T(x^k) = \gamma_k x^k, \qquad \gamma_0, \gamma_1, \dots \in \mathbb{R}$$

(or $\gamma_k \in \mathbb{C}$ if one considers the analogous problem). Similarly, one can show that the former operators each have a 'power series expansion'

$$T = \sum_{k=0}^{\infty} \widehat{T}(k) \partial^k, \qquad \widehat{T}(k) \in \mathbb{R}.$$

Notice here that applying such an operator to a polynomial only requires finitely many terms, so the sequence $\widehat{T}(k)$ can be arbitrary. One associates to this operator its *symbol*:

$$G_T(s) := \sum_{k=0}^{\infty} \widehat{T}(k) s^k. \tag{33.2}$$

In this section we present the characterizations of both classes of operators – commuting with additive and multiplicative differentiation $\partial, x\partial$ respectively – and they both bear the name of Pólya (with Benz and Schur, respectively). We will mostly focus on the latter case, in which the scalars γ_k are called *multipliers*. Thus, the Pólya–Schur theorem classifies all multiplier sequences (which preserve real-rooted polynomials).

For now, we return to the opening discussion of preserving real-rootedness, or more generally, diminishing the number of non-real roots. Our journey begins with a classical result due to Poulain [293] in 1867, answering a question of Hermite [166] from the previous year:

Theorem 33.3 (Hermite–Poulain)**.** *Suppose $q(x) = \sum_{k=0}^{m} q_k x^k$ is a polynomial with $q_0, q_m \neq 0$ and all real roots.*

(1) If $p(x) \in \mathbb{R}[x]$, then $Z_{nr}(q(\partial)p) \leq Z_{nr}(p)$; here the differential operator $q(\partial)$ acts via

$$(q(\partial)p)(x) := \sum_{k=0}^{m} q_k p^{(k)}(x).$$

(2) If q has only positive (respectively, negative) zeros, and $A \in \mathbb{R}$, then the number of zeros in $[A, \infty)$ (respectively, $(-\infty, A]$) of $q(\partial)q$ exceeds that of q.

Thus, the Hermite–Poulain theorem extends Rolle's theorem – i.e., that differentiation diminishes the number of non-real roots of a polynomial – which is the special case $q(x) = x$.

Proof. To show (1), write $q(x) = q_m \prod_j (x - \alpha_j)$, where no α_j is zero since $q_0 \neq 0$. Since $q(\partial) = q_m \prod_j (\partial - \alpha_j)$, it suffices to show that $(\partial - \alpha)p(x)$ has at least as many real roots as p. We now present Poulain's proof of this, in a sense 'differential-equation theoretic': since

$$(\partial - \alpha)p(x) = e^{\alpha x} \partial (e^{-\alpha x} p(x)),$$

it suffices to show $p(x)$ has at most as many real roots as $(e^{-\alpha x} p(x))'$, where $\alpha \neq 0$. This follows by Rolle's theorem, since $e^{-\alpha x} p(x)$ vanishes at the roots of p as well as $\alpha x \to \infty$. (This is the trick that was used in proving the weak and strong versions of Descartes' rule of signs, in Lemma 5.2 and Theorem 10.3, respectively.)

This shows the first part, but also the second: the preceding sentence suggests how to proceed if all roots of q are non-zero, with a common sign. Indeed, a small refinement of the preceding proof now works on $[A, \infty)$ (respectively, $(-\infty, A]$). $\square$

Notice that the operator $q(\partial)$ commutes with 'additive' differentiation ∂, a notion discussed above – and it preserves real-rootedness. A complete characterization of such linear preservers was carried out by Benz, in *Comment. Math. Helv.* in 1934:

Theorem 33.4 (Pólya–Benz theorem, [34]). *Suppose T is a linear operator on the space of complex polynomials, which commutes with differentiation ∂. Then T preserves real-rootedness if and only if its symbol $G_T(s) := \sum_{k=0}^\infty \widehat{T}(k) s^k$, defined in (33.2), is an entire function that is either zero or in the Laguerre–Pólya class $\mathcal{LP}_2$ (see Definition 33.20):*

$$G_T(s) = C s^m e^{-\gamma s^2 + \delta s} \prod_{j=1}^\infty (1+\alpha_j s) e^{-\alpha_j s}, \text{ with } C \in \mathbb{C}^\times; m \in \mathbb{Z}^{\geq 0}; \gamma \geq 0; \delta, \alpha_j \in \mathbb{R}; \sum_{j=1}^\infty \alpha_j^2 < \infty.$$

33.1. **Multiplier sequences and early results.** In the remainder of this section, we focus on the linear transformations which preserve real-rootedness, commute with 'multiplicative differentiation' $x\partial$, and turn out to be intimately linked to our main objects of focus: Pólya frequency functions. These are the so-called 'diagonal transforms' of $\pi(\mathbb{C})$ – i.e., multiplying each monomial x^k in a polynomial $p(x)$ by a scalar $\gamma_k \in \mathbb{R}$. They are called multipliers; corresponding to the two related questions after Definition 33.1, they come in two varieties:

Definition 33.5 (Pólya–Schur, [288]). Given a sequence $\Gamma = (\gamma_k)_{k=0}^\infty$ of real numbers, define the linear map $\Gamma[-] : \mathbb{R}[[x]] \to \mathbb{R}[[x]]$ (in particular, acting on polynomials) via

$$\Gamma \left[\sum_{k \geq 0} a_k x^k \right] := \sum_{k \geq 0} \gamma_k a_k x^k.$$

We now say that a sequence Γ is a *multiplier sequence of the first kind* if $\Gamma[p]$ is real-rooted whenever the polynomial $p(x)$ is; and *of the second kind* if $\Gamma[p]$ is real-rooted whenever the polynomial $p(x)$ has all roots real, non-zero, and of the same sign.

Example 33.6 (Laguerre, 1884). Given a real number $a > 0$ and an integer $k > 0$, the sequence

$$a(a+1)\cdots(a+k-1), \quad (a+1)(a+2)\cdots(a+k), \quad \dots$$

is a multiplier sequence which preserves the real-rootedness/one-sidedness of roots – and more generally, diminishes the number of non-real roots. This follows from a more general result by Laguerre – see Theorem 33.8(2), specialized here to $q(x) = (x+a)(x+a+1)\cdots(x+a+k-1)$.

In their famous 1914 work [288] in *J. reine angew. Math.*, Pólya and Schur provided 'algebraic' and 'transcendental' characterizations of the above two classes of multipliers; these are explained below. That said, the study of such multipliers had begun well before. We present here a quintet of well-known results in this direction – these are shown presently:

- by Laguerre, in 1882 in *C. R. Acad.* and in 1884 in *Acta Math.*;
- by Maló in 1895 in *J. Math. Spéc.* (this is briefly used in the next part of the text);
- by Jensen in 1913 in *Acta Math.*; and
- by Schur in 1914, in *J. reine angew. Math.*

To state and prove these, define the *Schur composition* $\odot$ of two polynomials/power series:

$$p(x) = \sum_{j \geq 0} p_j x^j, \ q(x) = \sum_{j \geq 0} q_j x^j, \quad \Longrightarrow \quad (p \odot q)(x) := \sum_{j \geq 0} j! p_j q_j x^j. \tag{33.7}$$

Theorem 33.8. *Suppose $p(x) = p_0 + p_1 x + \cdots + p_n x^n$ and $q(x) = q_0 + q_1 x + \cdots + q_m x^m$ are polynomials in $\mathbb{R}[x]$, with $m, n \geq 0$ and q real-rooted.*

(1) (Laguerre [228], 1882.) *The polynomial $\sum_{k \geq 0} (q_k/k!) x^k$ is real-rooted. In other words, $1, 1, 1/2!, 1/3!, \dots$ is a multiplier sequence of the first kind.*

(2) (Laguerre [230], 1884.) *Suppose $q \neq 0$ on $[0, n]$. Then the polynomial*

$$\tilde{q}(x) := q(0)p_0 + q(1)p_1 x + \cdots + q(n)p_n x^n$$

has at least as many real roots as p, that is, $Z_{nr}(\tilde{q}) \leq Z_{nr}(p)$.

(3) (Maló [249], 1895.) *Suppose the roots of q are all non-zero and of the same sign. If p is real-rooted, so is the Hadamard composition $p_0q_0 + p_1q_1x + \cdots + p_lq_lx^l$ (provided it is not identically zero), where $l = \min\{m, n\}$.*

(4) (Jensen [194], 1913.) *The following is a multiplier sequence of the first kind for $n \geq 1$:*

$$1,\ 1,\ 1 - \frac{1}{n},\ (1 - \frac{1}{n})(1 - \frac{2}{n}),\ \ldots,\ \prod_{j=1}^{n-1}(1 - \frac{j}{n}),\ 0,\ 0,\ \ldots$$

(5) (Schur composition theorem [330], 1914.) *Suppose the roots of q are all non-zero and of the same sign. If p is real-rooted, so is the Schur composition $p \odot q$ (or else $p \odot q \equiv 0$).*

Thus, parts (3) and (5) yield 'finite' multiplier sequences of the first kind.

To show these results, we first present a more general theorem, shown in 1949 by de Bruijn:

Theorem 33.9 (de Bruijn, [73])**.** *For an aperture $0 \leq \alpha \leq \pi$, let S_α denote an open sector in the complex plane, given by*

$$S_\alpha := \{z \in \mathbb{C}^\times : \arg(z) \in (\theta_\alpha, \theta_\alpha + \alpha)\} \tag{33.10}$$

for a fixed 'initial angle' θ_α. Similarly, let S_β denote an open sector for $\beta \in [0, \pi]$ and fixed initial angle θ_β. Now suppose $p(z) = \sum_{k=0}^n p_k z^k$ and $q(z) = \sum_{k=0}^m q_k z^k$ are complex polynomials, with $p_n, q_m \neq 0$. If $p(z), q(z)$ have all roots in the sectors S_α, S_β respectively, then their Schur composition $p \odot q$ has all roots in the open sector

$$-S_\alpha S_\beta := \{-z_1 z_2 \in \mathbb{C} : z_1 \in S_\alpha, z_2 \in S_\beta\}.$$

Proof. First suppose $(p \odot q)(z) = 0$ for some $z \notin -S_\alpha S_\beta$. For any such z, (by abuse of notation) $-zS_\alpha^{-1}$ is then disjoint from S_β, so we can embed both of these in open half-planes (i.e. open sectors of aperture π) $-zS_\alpha^{-1} \subset S_1$ and $S_\beta \subset S_2$ such that $S_1 \cap S_2 = \emptyset$. In particular, it suffices to show the result for $\alpha = \beta = \pi$. Now for a second reduction: the polynomials

$$p_1(z) := p(-ize^{i\theta_\alpha}) = \sum_{j=0}^n p_{1,j} z^j, \qquad q_1(z) := q(-ize^{i\theta_\beta}) = \sum_{j=0}^m q_{1,j} z^j,$$

can be verified to have all of their roots in the left half-plane in $\mathbb{C}$, i.e., in

$$L := \{z \in \mathbb{C} : \Re(z) < 0\}. \tag{33.11}$$

Hence in this 'reduction' case, if we show the following claim – that the polynomial

$$(p_1 \odot q_1)(z) := (p \odot q)(-ze^{i(\theta_\alpha + \theta_\beta)})$$

has no roots in $(0, \infty)$ – then $(p \odot q)(z)$ has no roots in $\mathbb{C} \setminus -S_\alpha S_\beta$, as desired.

While this claim can be shown using Grace's Apolarity Theorem (see e.g. [366]), we mention de Bruijn's direct argument. First claim that if $\lambda \in (0, \infty)$, $\eta \in L$, and $P(z)$ is any non-zero polynomial with degree $n \geq 0$, leading coefficient $P_n \neq 0$, and all roots in L, then $(\lambda\partial - \eta)P$ is not identically zero and its roots still lie in L. Indeed, write $P(z) = P_n \prod_{j=1}^n (z - \xi_j)$; then

$$(\lambda\partial - \eta)P(z) = P(z)\left(\sum_{j=1}^n \frac{\lambda}{z - \xi_j} - \eta\right), \qquad \forall z \in \mathbb{C} \setminus L.$$

The first factor on the right is non-zero by assumption, while each summand and $-\eta$ both have positive real part, so that the second factor on the right is in $-L$. It follows that the non-zero polynomial $(\lambda\partial - \eta)P(z)$ again has all roots in L. Now start with

$$P(z) = p_1(z) = p_{1,n} \prod_{j=1}^{n} (z - \xi_j), \qquad q_1(z) = q_{1,m} \prod_{j=1}^{m} (z - \eta_j),$$

and apply the above reasoning inductively for each $\eta = \eta_j$, to conclude that

$$(q_1(\lambda\partial)p_1)(z) = q_{1,0}p_1(z) + q_{1,1}\lambda p_1'(z) + \cdots + q_n \lambda^n p_1^{(n)}(z)$$

has all its roots in L. In particular, it does not vanish at 0, so

$$0 \neq q_{1,0}p_{1,0} + 1!q_{1,1}p_{1,1}\lambda + \cdots + l!q_{1,l}p_{1,l}\lambda^l, \qquad l = \min\{m, n\}.$$

This precisely says $0 \neq (p_1 \odot q_1)(\lambda)$; as $\lambda \in (0, \infty)$ was arbitrary, the proof is complete. □

This result led de Bruijn to derive a host of corollaries, which we discuss before returning to the proof of Theorem 33.8. The first is the 'closed' sector version of Theorem 33.9:

Corollary 33.12 (de Bruijn, [73])**.** *Suppose $p, q \neq 0$ are complex polynomials, whose roots all lie in closed sectors $\overline{S}_\alpha, \overline{S}_\beta$ respectively. If both $\alpha, \beta \in [0, \pi]$, then either $p \odot q \equiv 0$ or it has all roots in the sector $-\overline{S}_\alpha\, \overline{S}_\beta$. (If an aperture is 0, that closed sector is a half-line.)*

Proof. This is obtained from Theorem 33.9 via limiting arguments (and the continuity of roots, e.g. by Hurwitz's theorem), when the apertures of both sectors are positive. If either aperture is zero, write the corresponding half-line as an intersection of a sequence of nested closed sectors with positive apertures, and apply Corollary 33.12 for each of these. (Note that if the apertures $\alpha, \beta \in [0, \pi)$, then one can write each closed sector as an intersection/limit of a decreasing family of open sectors of apertures $< \pi$, and then the result follows from Theorem 33.9 simply by taking intersectons.) □

The next corollary shows two results of Weisner, from his 1942 paper in *Amer. J. Math.*:

Corollary 33.13 (Weisner, [366, Theorem 1 and its Corollary])**.** *Suppose p, q are polynomials in $\mathbb{R}[x]$ with $q(x)$ having all real roots.*

- *(1) If the roots of $p(x)$ lie in a closed sector $\overline{S}_\alpha$ with aperture $\leq \pi$, then either $p \odot q \equiv 0$ or it has all roots in the 'double sector' $\pm\overline{S}_\alpha$.*
- *(2) If the roots of q are moreover negative, and a closed sector with aperture in $[0, \pi]$ contains the roots of p, it also contains the roots of $p \odot q$.*

Proof. For the first part: if $p \odot q \equiv 0$ or S_α has aperture π then the result is immediate; now suppose neither condition holds. Apply Corollary 33.12 twice: with S_β the lower and upper half-planes $\pm iL$, where L is the left half-plane (33.11). It follows that the roots of $p \odot q$ lie in

$$(-iL \cdot \overline{S}_\alpha) \cap (iL \cdot \overline{S}_\alpha) = \overline{S}_\alpha \cup -\overline{S}_\alpha.$$

This shows the first part; for the second, apply Corollary 33.12 with $\overline{S}_\beta = (-\infty, 0]$. □

The next corollary is a mild strengthening of another result by de Bruijn [73]; his proof works for the following as well.

Corollary 33.14 (de Bruijn, [73]). *Fix scalars $-\infty < \delta < 0 \leq \Delta \leq +\infty$, and suppose the polynomials $p, q \in \mathbb{C}[z] \setminus \{0\}$ both have all roots in the strip $\Im(z) \geq -\Delta$. Then the polynomial*

$$\varphi_\delta(z) := \sum_{k=0}^{\infty} \frac{\delta^k}{k!} p^{(k)}(z) q^{(k)}(z)$$

has all roots also in the same strip $\Im(z) \geq -\Delta$.

Notice that $\varphi_\delta(z)$ has the same form as the Schur composition, but in the 'other' parameter. Namely, $\varphi_z(0) = (p \odot q)(z)$. This is used in the proof.

Proof. Fix a non-real number w with $\Im(w) < -\Delta$. We need to show that $\varphi_\delta(w) \neq 0$. To do so, consider the polynomials

$$P(z) := \sum_{k=0}^{\infty} z^k \frac{p^{(k)}(w)}{k!} = p(z+w), \qquad Q(z) := \sum_{k=0}^{\infty} z^k \frac{q^{(k)}(w)}{k!} = q(z+w).$$

By assumption, the zeros z of P, Q lie in the upper half-plane, which is an open sector with aperture π. Hence by Theorem 33.9, the roots z (not w) of their Schur composition $\varphi_z(w) = (P \odot Q)(z)$ all lie in $\mathbb{C} \setminus (-\infty, 0]$. In particular, $\varphi_\delta(w) = (P \odot Q)(\delta) \neq 0$, by choice of δ. $\square$

Remark 33.15. If instead p, q have all roots in the strip $\Im(z) \leq \Delta$, a similar argument shows that so does φ_δ. Intersecting these two results yields the version in de Bruijn's paper [73], i.e., for the strip $|\Im(z)| \leq \Delta$. In particular, if $\Delta = 0$, this also shows that if p, q are real-rooted, then so is φ_δ. However, the above approach has the advantage that de Bruijn's results also yield root-location results in asymmetric strips $\Im(Z) \in [-\Delta, \Delta']$ for $0 \leq \Delta \neq \Delta'$.

Finally, we prove the classical results stated above.

Proof of Theorem 33.8. We show the five parts in a non-sequential fashion, beginning with part (2) by Laguerre. It turns out this is precisely the counterpart of the Hermite–Poulain theorem 33.3, now for the 'multiplicative differential' operator $x\partial$ instead of the usual derivative ∂. Indeed, write $q(x) = q_m \prod_{j=1}^m (x - \alpha_j)$ with $\alpha_j \in \mathbb{R}$, and compute at each monomial:

$$q(x\partial)(x^k) = q_m \prod_{j=1}^m (x\partial - \alpha_j)(x^k) = x^k \cdot q_m \prod_{j=1}^m (k - \alpha_j) = q(k) x^k.$$

Since the factors $(x\partial - \alpha_j)$ pairwise commute, it again suffices to show $(x\partial - \alpha)p(x)$ has at least as many real roots as p, if $\alpha \in \mathbb{R} \setminus [0, \deg(p)]$ (so we may assume $\deg(p) = n$). We now study the order of the root at 0, and the positive/negative roots of both polynomials p and $(x\partial - \alpha)p$. The orders of the root at zero agree. Coming to positive roots, write using Poulain's idea:

$$(x\partial - \alpha)p(x) = x^{\alpha+1} \partial(x^{-\alpha} p(x)).$$

Now if $\alpha < 0$ then $x^{-\alpha}p(x)$ has an additional zero at $x = 0$; while if $\alpha > n$, $x^{-\alpha}p(x)$ has an additional 'zero' at $x = +\infty$. In both cases, one argues as in the proof of the Hermite–Poulain theorem 33.3 above, using Rolle's theorem to obtain that the positive roots of $(x\partial - \alpha)p(x)$ are at least as many as that of $p(x)$ – if $x > 0$. A similar argument holds for $x < 0$.

Next, part (5), i.e., the Schur composition theorem, is a special case of Corollary 33.13(1), in which the roles of p and q are reversed, and one takes $\overline{S}_\alpha$ to be a closed semi-axis in $\mathbb{R}$. To now obtain the result of Laguerre (part (1)), first apply 'inversion' to observe that if

$$q(x) = q_0 + q_1 x + \cdots + q_m x^m$$

is real-rooted, then so is its 'inversion' (possibly up to a power of x),

$$Q(x) := q_m + q_{m-1}x + \cdots + q_0 x^m.$$

Now let $P(x) := (1+x)^m$; then by part (5) the polynomial

$$(P \odot Q)(x) = q_m + q_{m-1}\frac{m!}{(m-1)!}x + q_{m-2}\frac{m!}{(m-2)!}x^2 + \cdots + q_0 m! x^m$$

is also real-rooted. Again invert this (and multiply by powers of x if necessary), and then divide throughout by $m!$ to obtain Laguerre's result, i.e., part (1).

The proof of Jensen's result (part (4)) is similar to that of part (1). Start with a real-rooted polynomial $p(x)$ and take the Schur composition with $(1+x)^n$. Hence the polynomial

$$\sum_{k \geq 0} p_k \cdot n(n-1)\cdots(n-k+1) \cdot x^k$$

is real-rooted. Replacing x by x/n, so is the polynomial

$$p_0 + p_1 x + (1 - \frac{1}{n})p_2 x^2 + \cdots$$

Since this holds for all polynomials, Jensen's result follows. Finally, Maló's result (part (3)) immediately follows by combining parts (1) and (5). $\square$

Remark 33.16. For additional variants of Theorem 33.8(2) due to Laguerre – involving the multiplicative differentiation operator $x\partial$, see Pinkus's paper [280].

Remark 33.17. By using inversion twice, and Theorem 33.8(1) in between, we obtain yet another result of Laguerre: if $q(x) = q_0 + \cdots + q_m x^m$ is real-rooted, then so is the polynomial

$$\frac{q_0}{m!} + \frac{q_1 x}{(m-1)!} + \cdots + \frac{q_m x^m}{0!}.$$

Some concluding remarks: the study of root-location of real/complex polynomials remains evergreen; see e.g. the very recent work [76] on "zero-sector reducing" linear operators on $\mathbb{R}[x]$ (in addition to related highlights of modern mathematics, described presently). Finally, we present without proof, a result of Schur [330] and Szegő [350] – as well as one by Pólya – which involve a different kind of 'composition':

Theorem 33.18. *Let $n \geq 1$ be an integer, and $p(z), q(z)$ be polynomials given by*

$$p(z) := \sum_{k=0}^{n} \binom{n}{k} p_k z^k, \qquad q(z) := \sum_{k=0}^{n} \binom{n}{k} q_k z^k.$$

(1) (Schur–Szegő composition theorem.) *If q is real-rooted with all roots in $(-1,0)$, and the roots of p lie in a convex region K containing the origin, then all roots of the following 'composition' of theirs lie in K:*

$$h(x) := \sum_{k=0}^{n} \binom{n}{k} p_k q_k z^k.$$

(2) (Pólya – see [366, Theorem 3].) *If q is real-rooted, and the roots of p lie in a sector S, then the roots of $h(x)$ lie in the double sector $\pm S$.*

33.2. **Laguerre–Pólya entire functions.** We now proceed toward Pólya and Schur's characterizations of multiplier sequences. Definition 33.5 implies that the multipliers of the first kind form a subclass of the second kind. It turns out that these characterizations are related to the polynomials being acted upon, and we begin by understanding these polynomials.

Recall that the Hermite–Poulain theorem 33.3(1) says that if p, q are real-rooted polynomials, then $q(\partial)p$ is also thus. (As a special case, Rolle's theorem says that the class of real-rooted polynomials is closed under differentiation.) If we now take *limits* of such polynomials – in a suitable sense – then we may expect that such properties hold as well. This does turn out to be true in several cases:

(1) As just mentioned, Theorem 33.3(1) says that if q is a real-rooted polynomial, then the differential operator $q(\partial)$ preserves real-rootedness on polynomials.
(2) More generally, the same turns out to hold if one considers functions that are the limits, uniform on every compact subset of $\mathbb{C}$, of real-rooted polynomials.
(3) Notice for the exponential function $e^{\alpha x}$ that $q(\partial)e^{\alpha x} = q(\alpha)e^{\alpha x}$. 'Dually', it turns out that the differential operator $e^{\alpha\partial}$ (for $\alpha \in \mathbb{R}$) preserves real-rootedness on polynomials. Indeed, this clearly holds when $e^{\alpha\partial}$ acts on linear polynomials, so we need to show that the same property is preserved under products. But this is a straightforward calculation:

$$\begin{aligned} e^{\alpha\partial}(p(x)q(x)) &= \sum_{n=0}^{\infty} \frac{\alpha^n}{n!} \sum_{k=0}^{\infty} \binom{n}{k} p^{(k)}(x) q^{(n-k)}(x) = \sum_{k=0}^{\infty} \frac{\alpha^k}{k!} p^{(k)}(x) \sum_{j=0}^{\infty} \frac{\alpha^j}{j!} q^{(j)}(x) \\ &= (e^{\alpha\partial}p(x))(e^{\alpha\partial}q(x)), \end{aligned} \tag{33.19}$$

where all sums are finite, since p, q are polynomials. In fact, since the product rule says that the *locally nilpotent* operator ∂ is a derivation of the algebra $\mathbb{R}[x]$ of polynomials, (33.19) is simply an instance of the fact that a nilpotent derivation of an algebra R exponentiates to an algebra automorphism of R.

In the spirit of the two questions following Definition 33.1, note that the final case (among others) works even if $\alpha > 0$ and p, q have non-positive roots, since it works for each linear factor. Thus, we consider limits – again uniform on each compact subset of $\mathbb{C}$ – of polynomials with all roots lying on a real semi-axis (i.e., in $(-\infty, 0]$ or in $[0, \infty)$). Such limiting functions always turn out to be entire; they were famously characterized by Laguerre [228] (1882) and Pólya [284] (1913). These characterizations, which are now presented, preceded – and motivated – Pólya and Schur's work on multipliers and their classification.

Definition 33.20. An entire function $\Psi(z) \not\equiv 0$ is in the *first Laguerre–Pólya class*, denoted $\Psi \in \mathcal{LP}_1$, if it admits a Hadamard–Weierstrass factorization

$$\Psi(s) = C s^m e^{\delta s} \prod_{j=1}^{\infty} (1 + \alpha_j s), \qquad \text{with } C \in \mathbb{R}^\times,\ m \in \mathbb{Z}^{\geq 0},\ \delta, \alpha_j \geq 0,\ \sum_j \alpha_j < \infty. \tag{33.21}$$

Similarly, $\Psi \not\equiv 0$ is in the *second Laguerre–Pólya class*, denoted $\Psi \in \mathcal{LP}_2$, if it admits a Hadamard–Weierstrass factorization

$$\Psi(s) = C s^m e^{-\gamma s^2 + \delta s} \prod_{j=1}^{\infty} (1 + \alpha_j s) e^{-\alpha_j s}, \qquad \text{with } C \in \mathbb{R}^\times,\ m \in \mathbb{Z}^{\geq 0},\ \gamma \geq 0,\ \delta, \alpha_j \in \mathbb{R}, \text{ and } \sum_j \alpha_j^2 < \infty. \tag{33.22}$$

A few observations are in order: the first is the inclusion between these classes: $\mathcal{LP}_1 \subset \mathcal{LP}_2$. Second, all functions in $\mathcal{LP}_2$, whence in $\mathcal{LP}_1$, have real roots. Third, a function $\Psi(s) \in \mathcal{LP}_1$ has all non-positive roots, so $\Psi(-s)$ is also entire, with all non-negative roots.

We now have the following relationship between the Laguerre–Pólya classes $\mathcal{LP}_1, \mathcal{LP}_2$, and the discussion on uniform limits of real-rooted polynomials previous to Definition 33.20.

Theorem 33.23.

(1) (Laguerre, [228].) *Suppose an entire function $\Psi(s)$ lies in the class $\mathcal{LP}_1$ (respectively, in $\mathcal{LP}_2$). Then there exists a sequence $\psi_n(s)$ of polynomials with all roots in $(-\infty, 0]$ (respectively, in $\mathbb{R}$), that converges locally uniformly on $\mathbb{C}$ (i.e., on every compact subset).*

(2) (Pólya, [284].) *Conversely, fix a neighborhood $U \subset \mathbb{C}$ of the origin. Suppose a sequence $\psi_n(s)$ of polynomials, each of which has roots in $(-\infty, 0]$ (respectively, in $\mathbb{R}$), converges uniformly on U to a function not identically zero. Then the ψ_n converge locally uniformly on $\mathbb{C}$ to an entire function Ψ in the class $\mathcal{LP}_1$ (respectively, $\mathcal{LP}_2$).*

In fact, one can write down concrete sequences of polynomials converging to the Laguerre–Pólya entire functions $\Psi(x)$ in Definition 33.20:

$$\begin{aligned}
\psi_{1,n}(s) &:= Cs^m(1 + \frac{\delta s}{n})^n \prod_{j=1}^{n}(1 + \alpha_j s), \\
\psi_{2,n}(s) &:= Cs^m(1 - \frac{\gamma s^2}{n})^n(1 + \frac{\delta s}{n})^n \prod_{j=1}^{n}(1 + \alpha_j s)(1 - \frac{\alpha_j s}{n})^n.
\end{aligned} \tag{33.24}$$

At first a weaker variant of Theorem 33.23(2) was shown by Laguerre, who assumed $U = \mathbb{C}$ to conclude that $\Psi \in \mathcal{LP}_1$ or $\mathcal{LP}_2$. The stronger version above is by Pólya in [284], and is presently used to classify the multiplier sequences. Similarly: considering locally uniform limits of polynomials with non-negative zeros yields functions $\Psi(-s)$ such that $\Psi(s) \in \mathcal{LP}_1$.

Remark 33.25. Lindwart–Pólya showed [239] that in the above cases and more general ones, the uniform convergence of polynomials $\psi_n(s)$ on some disk $D(0,r) \subset \mathbb{C}$ implies uniform convergence on any compact subset of $\mathbb{C}$. (The reader may recall here the 'convergence extension theorems' of Stieltjes and Vitali.) This has since been extended to smaller sets than $D(0,r)$, e.g. by Korevaar–Loewner, Levin, and others.

Theorem 33.23 has seen several generalizations in the literature; see the works of Korevaar and Obrechkoff among others, e.g. [224, 225, 271, 272, 273, 274]. We present here a sample result, taken from Levin [235], on uniform limits of polynomials with zeros in a sector in $\mathbb{C}$:

Theorem 33.26. *Fix a neighborhood $U \subset \mathbb{C}$ of the origin, and a closed sector $\overline{S}_\theta$ with aperture $\theta < \pi$. Suppose a sequence $\psi_n(s)$ of polynomials, each having roots in $\overline{S}_\theta$, converges uniformly on U to a function $\Psi_\theta \not\equiv 0$. Then the ψ_n converge locally uniformly on $\mathbb{C}$ to*

$$\Psi_\theta(s) = Cs^m e^{\delta s} \prod_{j=1}^{\infty}(1 + \alpha_j s),$$

where $C \in \mathbb{C}^\times$, $m \in \mathbb{Z}^{\geq 0}$, α_j, δ are either 0 or lie in $-\overline{S}_\theta^{-1} := \{-1/s : 0 \neq s \in \overline{S}_\theta\}$, and $\sum_j |\alpha_j| < \infty$. Moreover, an entire function Ψ can be locally uniformly approximated by a sequence ψ_n with the above properties (over $\overline{S}_\theta$) if and only if Ψ is of the above form Ψ_θ.

Note, the same polynomials as in (33.24) (with suitable parameter values) converge to Ψ_θ.

33.3. **From Laguerre–Pólya to Pólya–Schur.** Finally, we come to Pólya and Schur's classification of multiplier sequences (see Definition 33.5). We begin with basic properties.

Lemma 33.27. *Suppose $\Gamma = (\gamma_0, \gamma_1, \dots)$ is a multiplier sequence of the first kind.*

1. *Then so is $(\gamma_k, \gamma_{k+1}, \dots)$ for all $k > 0$; this also holds if Γ is of the second kind.*
2. *If $\gamma_0 \neq 0$ but $\gamma_k = 0$ for $k > 0$, then $\gamma_{k+n} = 0$ for all $n \geq 0$.*
3. *If $\gamma_0 \neq 0$, then all non-zero γ_k have either the same sign or alternating signs.*

Proof.

1. Suppose a polynomial $p(x) = \sum_{j \geq 0} p_j x^j$ has all roots real (or real and of the same sign, in which case $p_0 \neq 0$). Then so does $q(x) := x^{-k}\Gamma[x^k p(x)] = \sum_{j \geq 0} \gamma_{k+j} p_j x^j$, and note that if $p_0 \neq 0$ then $q(0) \neq 0$.
2. The first step is the following assertion by Schur [330]: *If $p(x) = \sum_{j=0}^n p_j x^j \in \mathbb{R}[x]$ is real-rooted, with $p_0, p_n \neq 0$, then (i) no two consecutive coefficients p_j vanish; and (ii) if $p_j = 0$ for some $j \in (0, n)$, then $p_{j-1}p_{j+1} < 0$.*

   Indeed, since $p_0 \neq 0$, no root α_j of $p(x)$ is zero, whence an easy computation shows:
$$p_1^2 - 2p_0p_2 = p_0^2 \sum_{j=1}^n \alpha_j^{-2} > 0, \qquad \text{where} \qquad p(x) := p_0 \prod_{j=1}^n (1 - x/\alpha_j),$$
   Fix the least $j > 0$ with $p_j = 0$. Now $p^{(j-1)}(x)$ is real-rooted by Rolle's theorem, and
$$p^{(j-1)}(x) = (j-1)!p_{j-1} + \frac{j!}{1!}p_j x + \frac{(j+1)!}{2!}p_{j+1}x^2 + \cdots + \frac{n!}{(n-j+1)!}p_n x^{n-j+1}.$$
   In particular, applying the preceding analysis to $p^{(j-1)}(x)/(j-1)!$ yields:
$$0 < jp_j^2 - (j+1)p_{j-1}p_{j+1} \quad \implies \quad p_{j-1}p_{j+1} < 0.$$
   This implies Schur's assertion. We now prove the second part of the Lemma. Suppose $\gamma_n \neq 0 = \gamma_k$, for some $n > k > 0$. Then $\gamma_{k-1}\gamma_{k+1} < 0$ by Schur's assertion. On the other hand, the following polynomial is also real-rooted, which is impossible:
$$\Gamma[x^{k+1} - x^{k-1}] = x^{k-1}(\gamma_{k+1}x^2 - \gamma_{k-1}). \tag{33.28}$$
3. Suppose $\gamma_n \neq 0$ for some $n > 0$. By (33.28), $\gamma_{k-1}, \gamma_{k+1}$ have the same sign for all $0 < k < n$. Now consider the signs of γ_0 and γ_1, and let $n \to \infty$. ☐

Now Pólya and Schur provide the following two characterizations of multiplier sequences. Given Lemma 33.27, we work with non-negative sequences, else use $\Psi_\Gamma(-x)$ in place of $\Psi_\Gamma(x)$.

Theorem 33.29 (Pólya–Schur, [288])**.** *Given real $\Gamma = (\gamma_k)_{k=0}^\infty$, the following are equivalent:*

1. *Γ is a multiplier sequence of the first kind.*
2. (Algebraic characterization.) *For all $n \geq 0$, the polynomial $\Gamma[(1+x)^n]$ is real-rooted, with all zeros of the same sign, i.e., in the Laguerre–Pólya class $\mathcal{LP}_1$.*
3. (Transcendental characterization.) *The generating series*
$$\Psi_\Gamma(x) := \sum_{k=0}^\infty \frac{\gamma_k}{k!} x^k = \Gamma[e^x]$$
   belongs to the Laguerre–Pólya class $\mathcal{LP}_1$, or else $\Psi_\Gamma(-x)$ does so.

We now outline why this result holds, modulo Theorem 33.23:

Proof-sketch of Theorem 33.29. Suppose $\Gamma[-]$ is a multiplier of the first kind, and all $\gamma_k \geq 0$. (The case of γ_k alternating, via Lemma 33.27, is handled by considering $\Psi_\Gamma(-x)$.) In particular, defining $p_n(x) := (1 + x/n)^n$ for $n \geq 1$, all roots of

$$\Gamma[p_n(x)] = \sum_{k=0}^{n} \gamma_k \binom{n}{k} (x/n)^k = \sum_{k=0}^{n} \frac{\gamma_k}{k!} \prod_{j=1}^{k-1} \left(1 - \frac{j}{n}\right) x^k$$

are real, and necessarily non-positive since all Maclaurin coefficients of $\Gamma[p_n(x)]$ are non-negative. This last assertion holds by Descartes' Rule of Signs – see e.g. Theorem 10.7 with $I = (0, \infty)$. One then shows that this sequence $\Gamma[p_n]$ of polynomials forms a normal family, hence converges locally uniformly as $n \to \infty$ to an entire function, which is clearly given by $\Psi_\Gamma(x)$. Moreover, $\Psi_\Gamma(x) = \Gamma[e^x]$ must be in $\mathcal{LP}_1$.

Conversely, if $\Psi_\Gamma(x) \in \mathcal{LP}_1$, then it can be approximated locally uniformly by a sequence of polynomials $\psi_n(x) = \sum_{k \geq 0} \psi_{n,k} x^k$, all of whose roots are in $(-\infty, 0]$. Now suppose $p(x) = \sum_{k \geq 0} p_k x^k$ is a polynomial with all real roots. By Corollary 33.13(1), the polynomial $(\psi_n \odot p)(x)$ is real-rooted. Taking $n \to \infty$, the same holds for $\Gamma[p(x)]$, as desired. $\square$

There is a similar characterization of multipliers of the second kind (see also Theorem 34.5):

Theorem 33.30 (Pólya–Schur, [288])**.** *Given real $\Gamma = (\gamma_k)_{k=0}^\infty$, the following are equivalent:*

(1) Γ is a multiplier sequence of the second kind.
(2) (Algebraic characterization.) *For all $n \geq 0$, the polynomial $\Gamma[(1 + x)^n]$ is real-rooted, i.e., in the Laguerre–Pólya class $\mathcal{LP}_2$.*
(3) (Transcendental characterization.) *The generating series $\Psi_\Gamma(x) = \Gamma[e^x]$ is an entire function, and belongs to the Laguerre–Pólya class $\mathcal{LP}_2$.*

Clearly (1) $\implies$ (2). The proof of (2) $\implies$ (3) $\implies$ (1) resembles the corresponding proof for multiplier sequences of the first kind.

Remark 33.31. Notice the 'reversal', in a rough sense: Ψ_Γ in $\mathcal{LP}_1$ acts on – and preserves real-rootedness on – functions in $\mathcal{LP}_2$, and vice versa. This is because acting on larger test sets imposes more constraints and reduces the available functions / generating series.

We close this part with a few connections to Pólya frequency sequences – specifically, to the representations of one-sided PF sequences in Theorem 30.12, following the discussion prior to (30.11). First, the latter theorem implies the following 1951 observation:

Theorem 33.32 (Aissen–Edrei–Schoenberg–Whitney, [4])**.** *Suppose $\mathbf{a} = (a_n)_{n \geq 0}$ is a real sequence, with $a_0 = 1$, such that its generating function $\Psi_{\mathbf{a}}(s) := \sum_{n=0}^\infty a_n s^n$ is entire. Then the following are equivalent:*

(1) $\mathbf{a}$ is a Pólya frequency sequence, i.e., the bi-infinite matrix $(a_{j-k})_{j,k \in \mathbb{Z}}$ is totally non-negative (where we set $a_n := 0$ for $n < 0$).
(2) The function $\Psi_{\mathbf{a}}(s) = e^{\delta s} \prod_{j=1}^\infty (1 + \alpha_j s)$, for some $\delta, \alpha_j \in [0, \infty)$ such that $\sum_j \alpha_j < \infty$.
(3) The function $\Psi_{\mathbf{a}}(s)$ belongs to the Laguerre–Pólya class $\mathcal{LP}_1$.
(4) The sequence $(n! a_n)_{n \geq 0}$ is a multiplier sequence of the first kind.

Proof. That (2) $\iff$ (3) and (3) $\iff$ (4) follow from Theorems 33.23 and 33.29, respectively. Finally, that (1) $\iff$ (2) follows from Theorem 30.12, since $\Psi_{\mathbf{a}}$ is entire. $\square$

Remark 33.33. See also [68] for additional connections. Yet another connection is that Aissen et al's representation theorem 30.12 implies the Laguerre–Pólya theorem 33.23(2) for $\mathcal{LP}_1$. Indeed, as the authors observe in [4], let a sequence $p_n(s)$ of polynomials satisfy:

$p_n(0) = 1$ and all roots negative, and suppose $p_n \to \Psi$. Then the Maclaurin coefficients of each p_n generate a 'finite' Pólya frequency sequence by Corollary 30.24. Hence so do the coefficients of $\Psi(s)$, so $\Psi \in \mathcal{LP}_1$ by Corollary 33.32.

A final result, by Katkova in 1990 [207], connects PF sequences, Laguerre's theorem 29.22(2), and the algebraic and transcendental characterizations of Pólya–Schur multipliers:

Theorem 33.34 (Katkova, [207]). *Fix an integer $p \geq 1$ and a polynomial $f \in \mathbb{R}[x]$ with $f(x) > 0$ on $[0, \infty)$. There exists $n_0(p) > 0$ such that the following sets of Maclaurin coefficients form a (one-sided) TN_p sequence:*

(1) The Maclaurin coefficients of $(1+x)^n f(x)$, for all integers $n \geq n_0(p)$.

(2) The Maclaurin coefficients of $e^{sx} f(x)$, for all real $s \geq n_0(p)$.

33.4. **The Laguerre–Pólya class, the Riemann hypothesis, and modern applications of real-rootedness.** While at first glance, the definitions do not reveal a connection between the Laguerre–Pólya class and Pólya frequency functions, we saw at the end of the preceding section that there is at least a 'one-way' connection. In fact the (remarkable) connection goes both ways – parallel to the theory of Pólya–Schur multipliers – and will be precisely described in the next section, via the bilateral Laplace transform.

To conclude this section, we start with this transform and conduct a very quick tour of some of the gems of modern mathematics – starting with the (not modern) Riemann hypothesis. In 1927 in *J. reine angew. Math.*, Pólya initiated the study of functions $\Lambda(t)$ such that $\mathcal{B}(\Lambda)(s)$ has only pure imaginary zeros.[7,8] Pólya's work [286] was motivated by the Riemann hypothesis, conjectured in 1859 by Riemann [298]. It says that the analytic continuation of the Riemann zeta function

$$\zeta(s) := \sum_{n \geq 1} n^{-s}, \qquad \Re(s) > 1$$

has (trivial zeros at $s = -2, -4, \dots$ and) nontrivial zeros all on the critical line $\Re(s) = 1/2$. An equivalent formulation is via the *Riemann xi-functions*

$$\xi(s) = \frac{1}{2} s(s-1)\pi^{-s/2}\Gamma(s/2)\zeta(s), \qquad \Xi(s) = \xi(1/2 + is),$$

where Γ is Euler's gamma function. Note that $\xi(s) = \xi(1-s)$ and $\Xi(s) = \Xi(-s)$. Now the Riemann hypothesis is equivalent to the fact that Ξ has only real zeros. (In fact this was how Riemann stated his conjecture.) Since the function Ξ is entire of order one, this leads to a folklore result, which can be found in e.g. Pólya's 1927 work [287]:

Theorem 33.35. *The Riemann hypothesis is equivalent to the statement: $\Xi \in \mathcal{LP}_2$.*

Thus, the Laguerre–Pólya class occupies a special place in analytic number theory.

The Riemann hypothesis is one of the most studied problems in modern mathematics. It was originally formulated in the context of the distribution of prime numbers, and has far-reaching consequences. Most of the work toward settling this conjecture has employed methods from complex analysis and analytic number theory.

We now present three mutually inter-related reformulations of the Riemann hypothesis from analysis, very recently presented by Gröchenig, and 'orthogonal' to the aforementioned

[7] Curiously, several authors cite Pólya's paper in *J. London Math. Soc.* (1926), pp. 98–99 for this; but a glance at the journal website reveals that such a paper seems not to exist!

[8] Pólya's question is studied even today; see e.g. the 2019 work [102] of Dimitrov and Xu, which provides a characterization for a class of functions that contains the Riemann xi-function Ξ (defined presently).

methods. The main ingredients are Theorem 33.35 and Schoenberg's theorem 34.9 below, which asserts that a function is a Pólya frequency function if and only if its bilateral Laplace transform is the reciprocal of a function Ψ in the Laguerre–Pólya class with $\Psi(0) > 0$:

Theorem 33.36 (Gröchenig, 2020, [147])**.** *Let $1/2 + it_0$ be the first zero of the zeta function on the critical line $\Re(s) = 1/2$. The Riemann hypothesis holds, if and only if there exists a Pólya frequency function Λ satisfying:*

$$\frac{1}{\Xi(s)} = \mathcal{B}(\Lambda)(s), \qquad |\Re(s)| < t_0.$$

Taking the Fourier transform instead of the Laplace transform, this yields the Riemann hypothesis if and only if the function

$$\Lambda(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \frac{1}{\xi(u+1/2)} e^{-ixu}\, du$$

is a Pólya frequency function. Equivalently, write $\Xi(s) = \Xi_1(-s^2)$ (since Ξ is even); thus Ξ_1 is entire of order $1/2$. Now the Riemann hypothesis holds, if and only if there exists a one-sided Pólya frequency function Λ with support in $[0,\infty)$, and a scalar $\alpha < 0$, satisfying:

$$\frac{1}{\Xi_1(s)} = \mathcal{B}(\Lambda)(s), \qquad \Re(s) > \alpha.$$

See also the 2007 paper of Katkova [208] for more connections between the Riemann hypothesis and total positivity, this time through Pólya frequency *sequences.* We provide here a few details. Let $\xi_1(s) := \xi(1/2 + \sqrt{s})$; then *the Riemann hypothesis is equivalent to $\xi_1 \in \mathcal{LP}_1$, or by Theorem 33.32, $\xi_1 = \Psi_{\mathbf{a}}$ for a PF sequence* $\mathbf{a}$. Now Katkova shows:

Theorem 33.37 (Katkova, [208])**.** *We have $\xi_1 = \Psi_{\mathbf{a}}$ for a sequence* $\mathbf{a}$ *that is Pólya frequency (or totally non-negative) of order at least* 43. *Moreover, the sequence* $\mathbf{a}$ *is* asymptotically PF, *i.e., for all $p \geq 0$ there exists $N_p > 0$ such that the matrix $(a_{N_p+j-k})_{0\leq j,\, 0\leq k\leq p-1}$ is TN.*

We follow the above discussion with a disparate development, in mathematical physics: the *Lee–Yang program.* In material science, it has been observed that certain magnetic materials lose magnetism at a critical temperature. This phase transition is called the Curie point/temperature. Such phenomena in statistical physics led to work on the Ising and other models, by Ising, Onsager, and several others. In the 1950s, Lee and Yang related this study to locating zeros of the 'partition function' associated with the model (and the underlying Lee–Yang measure). As a result, they were able to compute the phase transition for the Ising model. (This was part of their body of work that earned them the 1957 Nobel Prize in Physics.) Lee–Yang showed in [233, 375] that under desirable conditions, all zeros of the partition function are purely imaginary, or under a specialization, all on the circle:

Theorem 33.38 (Lee–Yang, 1952)**.** *Given an integer $n \geq 1$, a matrix $J \in [0,\infty)^{n\times n}$ (the 'ferromagnetic coupling constants'), and magnetic fields $h_1, \dots, h_n \in \mathbb{C}$, define*

$$Z_J(\mathbf{h}) := \sum_{\sigma \in \{-1,1\}^n} \exp(\sigma^T J \sigma + \sigma^T \mathbf{h})$$

to be the corresponding 'partition function'. (Here, the σ_j are the 'spins'.) Then $Z_J(\mathbf{h})$ is non-zero if $\Re(h_j) > 0\ \forall j$, and all zeros $h \in \mathbb{C}$ of $Z_J(h, h, \dots, h)$ are purely imaginary.

This result leads to the so-called *Lee–Yang circle theorem* (see also [307] for its connections to the original work of Lee–Yang):

Theorem 33.39. *Let $A = A^* \in \mathbb{C}^{n\times n}$ be Hermitian, with all a_{jk} in the closed unit disk. Then the multiaffine* Lee–Yang polynomial

$$f(z_1,\dots,z_n) := \sum_{S\subset[n]} \prod_{j\in S} z_j \prod_{k\notin S} a_{jk}$$

has no zeros in $\mathbb{D}^n$, where $\mathbb{D}$ is the open unit disk. In particular, the map $f(z,\dots,z)$ has all zeros on the unit circle S^1.

An essential part of the ensuing analysis in the program initiated by these results, involves understanding operators preserving spaces of polynomials with roots lying in / avoiding a prescribed domain in $\mathbb{C}$. More precisely, we are back to understanding linear operators on spaces of (multivariate) polynomials, preserving (higher dimensional) versions of stability, real-rootedness, and hyperbolicity. This includes higher-dimensional versions of Pólya–Schur multipliers. Such tools were developed, and a host of classification results obtained, around the turn of the millennium by Borcea and Brändén, in a series of remarkable papers.

As late as 2004, Craven–Csordas mention in their survey [91] that a classification of linear maps preserving $\pi_n(S)$ (see the paragraph following Definition 33.1) was not known even for important classes of domains $S \subset \mathbb{C}$, including $S = \mathbb{R}$, or a half-plane, or more generally a strip over an imaginary interval (a, b); or a (double) sector centered at 0. Answers started to come in only a few years after that; we present one result. In their 2009 paper in *Ann. of Math.*, Borcea–Brändén showed:

Theorem 33.40 ([61])**.** *Let $T : \pi_n(\mathbb{C}) \to \pi(\mathbb{C})$ be a linear operator on polynomials. The following are equivalent:*

(1) T preserves real-rootedness, i.e., $T : \pi_n(\mathbb{R}) \to \pi(\mathbb{R})$.
(2) The linear map T has rank at most 2, and is of the form $T(p) = \alpha(p)f + \beta(p)g$, where $\alpha, \beta : \mathbb{R}[x] \to \mathbb{R}$ are linear functionals, and $f, g \in \pi(\mathbb{R})$ have interlacing roots.
(3) The symbol *of $T(x, y)$, given by $G_T(x, y) := T((x + y)^n) = \sum_{k=0}^n \binom{n}{k} T(x^k)y^{n-k}$, is* stable. *In other words, it has no root (x, y) with $\Im(x), \Im(y) > 0$.*
(4) The symbol of $T(x,-y)$, given by $G_T(x,-y) := T((x - y)^n)$, is stable.

Here Borcea–Brändén define stability in keeping with Levin's notion of H-stability; note that a univariate *real* polynomial $q(x)$ is stable (i.e., has no roots in the upper half-plane) if and only if it is real-rooted: $Z_{nr}(q) = 0$.

Returning to the historical account, in [60, 61, 62, 63, 64] Borcea–Brändén also

- characterized linear operators preserving S-stability for other prescribed subsets $S \subset \mathbb{C}$ (including – in [61] – S a line, a circle, a closed half-plane, a closed disk, and the complement of an open disk);
- developed a multivariable Szász principle and multidimensional Jensen multipliers;
- proved three conjectures of C.R. Johnson;
- presented a framework that incorporated a vast number of (proofs of) Lee–Yang and Heilmann–Lieb type theorems;

among other achievements. See also a detailed listing of the modern literature in the field, in [61]. Together with Liggett in [65], they also developed the theory of negative dependence for "strongly Rayleigh (probability) measures", enabling them to prove various conjectures of Liggett, Pemantle, and Wagner, and to construct counterexamples to other conjectures on log-concave sequences. See also the survey of Wagner [358] for more details and connections.

The above works, originating from the Laguerre–Pólya–Schur program on the location of roots of polynomials, were taken forward very recently, by Marcus, Spielman, and Srivastava. In another series of striking papers that used interlacing families of polynomials, real stability, and hyperbolicity (among many other ingredients), the authors proved the longstanding Kadison–Singer conjecture [198], and also showed the existence of bipartite Ramanujan (expander) graphs of every degree and every order (settling conjectures of Lubotzky and Bilu–Linial). See [250, 251, 252]. These contributions are only a small part of a larger and very active current area of research, referring to the geometry of roots of polynomials. Stability of dynamical systems, global optimization, and in particular control theory were and are immediate beneficiaries of the theoretical advances. Under the covers of the newly founded *SIAM Journal of Applied Algebra and Geometry* many exciting discoveries touching the subject have appeared, with key concepts such as linear matrix inequalities, hyperbolic polynomials, spectrahedra, and semi-definite programming.

34. Schoenberg's results on Pólya frequency functions.

Having discussed the Laguerre–Pólya entire functions and the Pólya–Schur multiplier sequences, we return to our primary objects of interest: Pólya frequency functions. The goal in this section is to discuss some of the foundational results on these functions.

We first recall the definition of these functions and two closely related classes of functions:

Definition 34.1. Suppose $\Lambda : \mathbb{R} \to [0, \infty)$ is Lebesgue measurable.

1. Λ is said to be *totally non-negative (TN)* if given an integer $p \geq 1$ and tuples $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{p,\uparrow}$, $\det T_\Lambda[\mathbf{x}; \mathbf{y}] := \det(\Lambda(x_j - y_k))_{j,k=1}^p \geq 0$.
2. We will say Λ is *non-Dirac* if Λ does not vanish at least at two points.
3. Λ is a *Pólya frequency (PF) function* if Λ is non-monotone – equivalently, integrable with mass a positive real number.

The equivalence in the third assertion was proved in Proposition 29.3(3).

Remark 34.2 (Non-Dirac TN functions)**.** The functions in Definition 34.1(2) were termed *totally positive functions* by Schoenberg – recall that in his papers and Karlin's book, and even later, TN and TP matrices/functions were termed TP and STP, i.e., (strictly) totally positive. Finally, by Theorem 28.4, a non-Dirac TN function is strictly positive on an interval of positive length, and continuous on its interior.

34.1. **Precursors by Pólya and Hamburger.** We now discuss the origins of Pólya frequency functions. Recall that in his 1951 paper [324], Schoenberg proved the variation diminishing property for Pólya frequency functions Λ, in terms of the values of functions as well as for real zeros of polynomials. (See Propositions 29.15 and 29.17, respectively.) To the collection of prior results proved about the variation diminishing property – including for power series by Fekete in 1912 [118] and for matrices by Schoenberg in 1930 [311], Motzkin in 1936 in his thesis [263], and others including prominently by Gantmacher–Krein – we now add a result shown in 1915 by Pólya, in connection with the Laguerre–Pólya class $\mathcal{LP}_2$.

Specifically, Pólya studied the reciprocal of a function $\Psi_k(s) \in \mathcal{LP}_k$ for $k = 1, 2$, where $\Psi_k(0) > 0$. He expanded the meromorphic function $1/\Psi_k$ as:

$$\frac{1}{\Psi_k(s)} = \sum_{j=0}^{\infty} \frac{(-1)^j}{j!} \mu_j s^j \tag{34.3}$$

Pólya then applied the differential operator $1/\Psi_2(\partial)$ to a real polynomial $p(x)$, via

$$\frac{1}{\Psi_2(\partial)} p(x) = \mu_0 p(x) - \mu_1 p'(x) + \mu_2 \frac{p''(x)}{2!} - \cdots \pm \mu_n \frac{p^{(n)}(x)}{n!}, \tag{34.4}$$

in the spirit of the preceding section, and where $n = \deg(p) \geq 0$. Notice that since ∂ is locally nilpotent on $\mathbb{C}[x]$, every formal power series in ∂ yields a well-defined operator on $\mathbb{C}[x]$, yielding an algebra homomorphism $\mathbf{T}$ from $\mathbb{C}[[s]]$ to linear operators on $\mathbb{C}[x]$.

With these preliminaries, as a first result Pólya showed another condition equivalent to being in the second Laguerre–Pólya class, i.e., a multiplier sequence of the second kind:

Theorem 34.5 ([285])**.** *Given a formal power series $\psi(s) := \sum_{j \geq 0} \mu_j (-s)^j / j!$ with $\mu_0 > 0$, the following are equivalent:*

(1) Given a polynomial $p \in \mathbb{R}[x]$, $\psi(\partial)(p(x))$ has no more real roots than $p(x)$:

$$Z_{nr}(\psi(\partial) p(x)) \geq Z_{nr}(p(x)).$$

(2) $\Psi_2(s) := 1/\psi(s)$ is in the Laguerre–Pólya class $\mathcal{LP}_2$.

Proof-sketch. First, given the algebra homomorphism $\mathbf{T}$ sending a power series $\psi(s)$ to $\psi(\partial)$, and since $\psi(0) > 0$, the assertion (1) can be rephrased as:

(1′) *Let $\Psi_2(s) := 1/\psi(s)$. If $p \in \mathbb{R}[x]$, then $\Psi_2(\partial)p$ has at least as many real roots as p.*

We now show that (2) $\implies$ (1′). If Ψ_2 is constant then (1′) is immediate; else approximate $\Psi_2(s)$ locally uniformly by a sequence of real-rooted polynomials $\psi_n(s)$; since $\mu_0 = \Psi_2(0) > 0$ and Ψ_2 is non-constant, the same holds for ψ_n for $n \gg 0$. In particular, for large n, the Hermite–Poulain theorem yields $Z_{nr}(\psi_n(\partial)p) \leq Z_{nr}(p)$; and we also have $\deg(\psi_n(\partial)p) = \deg(p) = \deg(\Psi_2(\partial)p)$. Hence $Z_{nr}(\Psi_2(\partial)p) \leq Z_{nr}(p)$ by the continuity of roots.

Conversely, we assume (1′) and show (2). Let $g(x) := x^n$ for some $n \geq 0$. If $\Psi_2(s) = \sum_{j \geq 0} \nu_j s^j/j!$, then by (1′), the polynomial

$$\Psi_2(\partial)(x^n) = \sum_{j=0}^{n} \binom{n}{j} \nu_j x^{n-j}$$

has at least n roots, whence is real-rooted for $n \geq 0$. Define the real sequence $\mathbf{N} := (\nu_0, \nu_1, \dots)$; thus $\Psi_2(\partial)(x^n) = \mathbf{N}[(1+x)^n]$ is real-rooted for all $n \geq 0$. It follows by the Pólya–Schur characterizations of multiplier sequences of the second kind (see Theorem 33.30) that $\Psi_{\mathbf{N}}(x) = \mathbf{N}[e^x] = \Psi_2(x)$ is in the class $\mathcal{LP}_2$, as desired. $\square$

Thus, Pólya showed that $\Psi_2(\partial)$ weakly increases (i.e., does not decrease) the number of real roots. We return to this result and proof presently; first we continue with the account of Pólya's work [285]. Another result dealt with the reciprocal functions $1/\Psi_k(s)$, $k = 1, 2$:

Theorem 34.6 (Pólya, [285]). *Suppose $\Psi_k(s) \in \mathcal{LP}_k$ for $k \in \{1, 2\}$ is such that $\Psi_k(0) > 0$ and $\Psi_k(x) \not\equiv Ce^{\delta x}$. Let $1/\Psi_k(s) = \sum_{j \geq 0} \mu_j (-s)^j/j!$, and define the Hankel matrices*

$$H_{\mu,j,n} := \begin{pmatrix} \mu_j & \mu_{j+1} & \cdots & \mu_{j+n} \\ \mu_{j+1} & \mu_{j+2} & \cdots & \mu_{j+n+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{j+n} & \mu_{j+n+1} & \cdots & \mu_{j+2n} \end{pmatrix}, \qquad j \geq 0.$$

Then $\det H_{\mu,0,n} > 0 \ \forall n \geq 0$. *If $k = 1$ then moreover* $\det H_{\mu,1,n} > 0 \ \forall n \geq 0$.

Proof-sketch. Pólya first showed in [285, Satz II and Satz III] the following claim:

"*Suppose $k = 1$ or 2, and a function Ψ_k is as in the assumptions. If a non-zero real polynomial is non-negative on $\mathbb{R}$ and $k = 2$ (respectively, non-negative on $[a, \infty)$ and $k = 1$ for some $a \in \mathbb{R}$), then the polynomial $\frac{1}{\Psi_k((-1)^k \partial)} p$ is always positive on $\mathbb{R}$ (respectively, on $[a, \infty)$).*"

Now to prove $\det H_{\mu,0,n} > 0$, choose a non-zero vector $u = (u_0, \dots, u_n)^T \in \mathbb{R}^{n+1}$, and set $p(x) := (\sum_{j=0}^n u_j x^j)^2$. By the previous paragraph and a straightforward computation,

$$0 < \frac{1}{\Psi_2(\partial)} p \bigg|_{x=0} = \mu_0 u_0^2 - \mu_1(u_0 u_1 + u_1 u_0) + \mu_2(u_0 u_2 + u_1^2 + u_2 u_0) - \cdots = \sum_{j,k=0}^{n} (-1)^{j+k} \mu_{j+k} u_j u_k.$$

Since this inequality holds for all non-zero vectors u, the matrix $((-1)^{j+k} \mu_{j+k})_{j,k=0}^n$ has positive determinant by Sylvester's criterion (Theorem 2.8). But $H_{\mu,0,n}$ is obtained from this matrix by pre- and post-multiplication by the matrix $\operatorname{diag}(1, -1, 1, \dots, (-1)^{n-1})$.

This shows the result for $k = 2$; similarly for $k = 1$, choose a non-zero vector u as above, and set $p(x) := x(\sum_{j=0}^{n} u_j x^j)^2 \geq 0$ on $[0, \infty)$. By the opening paragraph, we compute:

$$0 < \frac{1}{\Psi_1(-\partial)} \, p \Big|_{x=0} = \sum_{j,k=0}^{n} \mu_{j+k+1} u_j u_k,$$

similar to the $k = 2$ case. Since this holds for all $u \neq 0$, it follows that $\det H_{\mu,1,n} > 0$. □

These positive determinants were taken up at the turn of that decade by Hamburger, who in 1920–21 published his solution to the Hamburger moment problem [162]. Around the same time, he applied this solution to Pólya's theorem 34.6, and showed that the positivity of the Hankel determinants in it is not sufficient to recover the Laguerre–Pólya class. Hamburger also showed, however, that the functions $1/\Psi_k$ are Laplace transforms:

Theorem 34.7 (Hamburger, 1920, [161]). *Fix an entire function $\Psi(s)$ with $\Psi(0) > 0$ and $1/\Psi(s) = \sum_{j \geq 0} \mu_j (-s)^j / j!$.*

1. *If $\det H_{\mu,0,n} > 0$ for all $n \geq 0$, then the reciprocal $1/\Psi(s)$ is the bilateral Laplace transform of a certain function $\Lambda(x) \geq 0$, in the maximal strip $\Re(s) \in (\alpha, \beta)$ containing the origin where $1/\Psi(s)$ is regular.*
2. *If $\det H_{\mu,0,n}, \det H_{\mu,1,n}$ are positive for all $n \geq 0$, then the reciprocal $1/\Psi(s)$ is the (bilateral) Laplace transform of a certain function $\Lambda(x) \geq 0$, with $\Lambda \equiv 0$ on $(-\infty, 0)$, in the maximal half-plane $\Re(s) \in (\alpha, \infty)$ of regularity of $1/\Psi(s)$.*

From these results, one sees that the μ_j are precisely the moments of Λ (see (29.18)), or of the non-negative measure $\Lambda(t)dt$:

$$\mu_j := \int_{\mathbb{R}} \Lambda(t) t^j \, dt < \infty, \qquad j = 0, 1, \ldots.$$

Combined with the Hamburger and Stieltjes moment problems – see Remarks 2.23 and 4.4 respectively – and since $\Lambda(t)dt$ has infinite support, this explains why the moment-matrices $H_{\mu,0,n}, H_{\mu,1,n}$ have positive determinants, i.e., are positive definite by Sylvester's criterion.

34.2. **Schoenberg's characterizations of PF functions.** Schoenberg built upon the above results, by (a) understanding the nature of the functions Λ in Hamburger's theorem 34.7; and (b) characterizing the functions $\Lambda(x) \in L^1(\mathbb{R})$ that satisfy the variation diminishing property:

$$S^-(g_f) \leq S^-(f) \ \forall f : \mathbb{R} \to \mathbb{R}, \quad \text{where} \quad g_f(x) := \int_{\mathbb{R}} \Lambda(x - t) f(t) \, dt. \tag{34.8}$$

Remarkably, the Λ in both cases are essentially one and the same, as explained below.

Theorem 34.9 (Schoenberg, [320, 324]). *Suppose $\Lambda : \mathbb{R} \to \mathbb{R}$ is Lebesgue measurable. If Λ is a non-Dirac TN function (see Definition 34.1), not of the form e^{ax+b} for $a, b \in \mathbb{R}$, then the bilateral Laplace transform of Λ is $1/\Psi(s)$, with $\Psi(0) > 0$ and $\Psi(s)$ in the Laguerre–Pólya class $\mathcal{LP}_2$, i.e., of the form*

$$\Psi(s) = C e^{-\gamma s^2 + \delta s} \prod_{j=1}^{\infty} (1 + \alpha_j s) e^{-\alpha_j s},$$

where $C > 0, \gamma \geq 0$, $\delta, \alpha_j \in \mathbb{R}$, $0 < \gamma + \sum_j \alpha_j^2 < \infty$, and the equality $\mathcal{B}(\Lambda)(s) = 1/\Psi(s)$ holds on a maximal strip $\Re(s) \in (\alpha, \beta)$. Here $-\infty \leq \alpha < \beta \leq \infty$, and if α and/or β is finite then it is a zero of $\Psi(\cdot)$.

Conversely, if $\Psi \in \mathcal{LP}_2$ is as above, then $1/\Psi(s)$ is the (bilateral) Laplace transform of a non-Dirac TN function Λ, not of the form e^{ax+b} for $a, b \in \mathbb{R}$.

In particular, the condition $0 < \gamma + \sum_j \alpha_j^2 < \infty$ implies that $\Psi(s)$ is not of the form $Ce^{\delta s}$. Moreover, in light of Proposition 29.3, one can translate the maximal strip by multiplying by an exponential factor:

Corollary 34.10. *Given a non-Dirac TN function $\Lambda : \mathbb{R} \to \mathbb{R}$, the following are equivalent:*
- *(1) Λ is a Pólya frequency function.*
- *(2) The maximal strip in the preceding theorem contains the imaginary axis.*
- *(3) Λ is non-monotone.*
- *(4) Λ is integrable.*

The above results were for general Pólya frequency functions (or non-Dirac TN functions) vis-à-vis the Laguerre–Pólya class $\mathcal{LP}_2$. The corresponding equivalence for one-sided TN/PF functions was also shown by Schoenberg, in the same works [320, 324]:

Theorem 34.11. *Suppose $\Lambda : \mathbb{R} \to \mathbb{R}$ is Lebesgue measurable. If Λ is a non-Dirac TN function, vanishing on $(-\infty, 0)$, and such that $\mathcal{B}(\Lambda)(s)$ converges for $\Re(s) > 0$ (in particular not an exponential e^{ax+b} for $a > 0$), then the bilateral Laplace transform of Λ is $1/\Psi(s)$, with $\Psi(s) > 0$ for $s > 0$ and $\Psi(s)$ in the Laguerre–Pólya class $\mathcal{LP}_1$, i.e., of the form*

$$\Psi(s) = Ce^{\delta s} \prod_{j=1}^{\infty} (1 + \alpha_j s),$$

where $C > 0$, $\delta, \alpha_j \geq 0$, $0 < \sum_j \alpha_j < \infty$, and the equality $\mathcal{B}(\Lambda)(s) = 1/\Psi(s)$ holds on a maximal strip $\Re(s) \in (\alpha, \infty)$. Here α denotes the first zero of $\Psi(\cdot)$.

Conversely, if $\Psi \in \mathcal{LP}_2$ is as above, then $1/\Psi(s)$ is the (bilateral) Laplace transform of a non-Dirac TN function Λ with the aforementioned properties.

Moreover, such a function Λ is a Pólya frequency function if and only if $\Psi(0) > 0$.

Note that $\Psi(0) > 0$ if and only if Λ is integrable; thus it cannot be constant on $(0, \infty)$.

Schoenberg's theorems 34.9 and 34.11 characterize non-Dirac TN functions and Pólya frequency functions, both one- and two-sided, in terms of the Laguerre–Pólya class.

Proof-sketch of Theorem 34.9. In the concluding portion of Section 32 (see the discussion around (32.16) and (32.17)) we saw an outline of why for $k = 1, 2$, the function $1/\Psi_k(s)$ for any $\Psi_k \in \mathcal{LP}_k$ is the Laplace transform of a Pólya frequency function (one-sided or general, for $k = 1, 2$ respectively). We outline here a proof of why the reverse implication holds for $k = 2$. The outline opens with Schoenberg's words [320]:

"*A proof of Theorem 34.9 is essentially based on the results and methods developed by Pólya and Schur. The only additional element required is a set of sufficient conditions insuring that a linear transformation be variation diminishing.*"

This last sentence of Schoenberg refers to his 1930 paper, in which he showed that TN matrices are variation diminishing. Using this property, he showed the same for Pólya frequency functions, whence for polynomials (and then did the same for one-sided PF functions); see Section 29.2. Now suppose Λ is a PF function; we proceed to outline the proof of why $\mathcal{B}(\Lambda)$, which converges on a maximal strip $\Re(s) \in (\alpha, \beta)$ with $\alpha < 0 < \beta$, is of the form $1/\Psi_2(s)$ for $\Psi_2(s)$ in the Laguerre–Pólya class $\mathcal{LP}_2$. Following Pólya, write

$$\mathcal{B}(\Lambda)(s) = \sum_{j=0}^{\infty} \frac{(-1)^j}{j!} \mu_j s^j, \qquad \Re(s) \in (\alpha, \beta),$$

where $\mu_j = \int_{\mathbb{R}} \Lambda(t)t^j\ dt$ are the moments, as discussed after Theorem 34.7. Also note that $\mathcal{B}(\Lambda)(0) = \mu_0 > 0$, so we write the reciprocal power series:

$$\Psi_2(s) := \frac{1}{\mathcal{B}(\Lambda)(s)} = \sum_{j=0}^{\infty} \frac{\nu_j}{j!} s^j.$$

Now return to the integral transformation $f \mapsto g_f$ as in (34.8). By Proposition 29.17, $Z(g_f) \leq Z(f)$, where $Z(\cdot)$ denotes the number of real roots. Next, recall Schoenberg's computation (29.19), perhaps inspired by Pólya's trick (34.4):

$$g_f(x) = \int_{\mathbb{R}} \Lambda(t) f(x-t)\ dt = \sum_{j=0}^{\infty} \frac{(-1)^j}{j!} \mu_j (\partial^j f)(x) = (\mathcal{B}(\Lambda)(\partial) f)(x).$$

We use here that both sides are polynomials, so that only finitely many terms $\mu_j \partial^j$ act nontrivially. Now take the reciprocal power series, e.g. via the map $\mathbf{T}$ following (34.4), to obtain: $f(x) = \Psi_2(\partial)(g_f)(x)$. In fact since $f \mapsto g_f$ is invertible (see the proof of Proposition 29.17), the linear operator $g \mapsto f_g := \Psi_2(\partial) g$ is also invertible, and it weakly increases the number of real roots. Finally, use $(1') \implies (2)$ in the proof of Pólya's theorem 34.5. $\square$

Remark 34.12. Thus, Schoenberg showed in [321, 324] the connection between his characterization of PF functions via the Laguerre–Pólya class, Pólya–Schur multipliers, and Pólya's theorem 34.5 proving the variation diminishing property over polynomials. This explains why Schoenberg proposed in [321] the name *Pólya frequency functions* for this family of functions.

We conclude with another characterization of Pólya frequency functions, which Schoenberg announced in [321] and showed in [323]. This occurs via the variation diminishing property:

Theorem 34.13 (Schoenberg, [321, 323])**.** *Given $\Lambda : \mathbb{R} \to \mathbb{R}$ Lebesgue integrable, let the kernel $f \mapsto g_f$ as in* (34.8)*, for continuous, bounded $f : \mathbb{R} \to \mathbb{R}$. The following are equivalent:*

(1) Λ is variation diminishing: $S^-_{\mathbb{R}}(g_f) \leq S^-_{\mathbb{R}}(f)$ for all continuous, bounded $f : \mathbb{R} \to \mathbb{R}$.
(2) One of $\pm\Lambda$ is a PF function, or Λ is a Dirac function $C\mathbf{1}_{x=a}$ for some $C, a \in \mathbb{R}$.

Thus, the converse to Proposition 29.15 holds as well. For yet another characterization of Pólya frequency functions – in terms of splines – see [95, 96] by Curry and Schoenberg.

34.3. **Support of PF functions; (strict) total positivity.** We now discuss two consequences of the above results, which were also proved by Schoenberg. The first is that a Pólya frequency function necessarily cannot be compactly supported:

Proposition 34.14. *The support of a Pólya frequency function $\Lambda : \mathbb{R} \to \mathbb{R}$ is unbounded.*

Recall by Theorem 28.4 that the support of a TN_2 function is an interval, so in fact the support of a PF function now must be of the form (a, b) where at least one of $|a|, |b|$ is infinite.

Proof. We already know the support of a Pólya frequency function is not a singleton. Let I be a bounded interval with endpoints $-\infty < a < b < \infty$. The claim is that for any function $\Lambda : \mathbb{R} \to \mathbb{R}$ supported on I and continuous in its interior, the Laplace transform

$$\mathcal{B}(\Lambda)(s) := \int_a^b e^{-sx} \Lambda(x)\ dx, \qquad s \in \mathbb{C}$$

is entire, with kth derivative $(-1)^k \int_a^b e^{-sx} x^k \Lambda(x)\ dx$. Indeed, one shows that the series

$$\sum_{n=0}^{\infty} \frac{(-s)^n}{n!} \int_a^b x^n \Lambda(x)\ dx$$

is uniformly convergent for s in any bounded domain in $\mathbb{C}$. Hence its sum is entire, and equals $\mathcal{B}(\Lambda)(s)$; moreover, one can differentiate term by term to compute its kth derivative:

$$\sum_{n=k}^{\infty} \frac{s^{n-k}}{(n-k)!} \int_a^b (-x)^n \Lambda(x)\ dx = \int_a^b e^{-sx} (-x)^k \Lambda(x)\ dx.$$

Having shown the claim, we proceed to the proof. By Theorem 34.9,

$$\mathcal{B}(\Lambda)(s) = C^{-1} e^{\gamma s^2 - \delta s} \prod_{j=1}^{\infty} \frac{e^{\alpha_j s}}{1 + \alpha_j s}, \tag{34.15}$$

for appropriate values of the parameters. Now if Λ has bounded support, then from above the right-hand side is entire. It follows that $\alpha_j = 0\ \forall j$. But then $\mathcal{B}(\Lambda)(s) = C^{-1} e^{\gamma s^2 - \delta s}$. Now by Laplace inversion and Example 29.4, Λ itself is a (shifted) Gaussian density, hence has unbounded support. This provides the necessary contradiction. $\square$

The final result here, due to Schoenberg and Whitney in *Trans. Amer. Math. Soc.* (1953), provides a sufficient condition when a Pólya frequency function is (strictly) totally positive. We mention only a part of their results, and show only the sub-part relevant for our purposes.

Theorem 34.16 (Schoenberg–Whitney)**.**

(1) If $\Lambda_+, \Lambda_0 : \mathbb{R} \to \mathbb{R}$ are PF functions, and Λ_+ yields a TP kernel, then so does $\Lambda_+ * \Lambda_0$.

(2) Suppose $\Lambda(x)$ has bilateral Laplace transform (34.15)*. If $\gamma > 0$, or $\gamma = 0$ and $\sum_j |a_j|$ diverges, then the kernel T_Λ is TP (of all orders).*

Proof. For the first part: that $\Lambda_+ * \Lambda_0$ is a PF function follows from Corollary 32.9. We now claim that given $p \geq 1$ and $\mathbf{y} \in \mathbb{R}^{p,\uparrow}$, there exists $\mathbf{t} = (t_1, \dots, t_p) \in \mathbb{R}^{p,\uparrow}$ such that $T_{\Lambda_0}[\mathbf{t}; \mathbf{y}]$ is non-singular. The proof is by induction on $p \geq 1$; for the base case, since $\int_{\mathbb{R}} \Lambda_0(x)\ dx \in (0, \infty)$, it is positive on an interval, so we can choose t_1 as desired.

For the induction step, let $p \geq 2$ and $\mathbf{y} \in \mathbb{R}^{p,\uparrow}$. Choose $\mathbf{t} := (t_1, \dots, t_{p-1}) \in \mathbb{R}^{p-1,\uparrow}$ such that the matrix $(\Lambda_0(t_j - y_k))_{j,k=1}^{p-1}$ is non-singular, whence has positive determinant since Λ_0 is TN. Now expand the determinant of $T_{\Lambda_0}[(\mathbf{t}; t); \mathbf{y}]$ along the last row; if this vanishes for all $t \in \mathbb{R}$, we obtain an equation $\sum_{k=1}^{p} a_k \Lambda_0(t - y_k) \equiv 0$, where all $a_k \in \mathbb{R}$ and $a_p > 0$. Taking the bilateral Laplace transform of both sides,

$$\psi(s) \sum_{k=1}^{p} a_k e^{-s y_k}$$

must vanish for s in some interval, where $\psi \not\equiv 0$ by the Schoenberg representation theorems. Hence the sum vanishes identically on an interval – which is false by Descartes' rule of signs (Theorem 10.3). Thus, $T_{\Lambda_0}[(\mathbf{t}; t); \mathbf{y}]$ is non-singular for some $t \in \mathbb{R}$. Clearly $t \neq t_j\ \forall j$, so by suitably permuting the rows (and relabeling the t, t_j if needed), the induction step is proved.

Having shown the claim, and using the continuity on an interval of both Λ_+, Λ_0, one checks using the Basic Composition Formula (5.14) that $\Lambda_+ * \Lambda_0$ is TP_p for each $p \geq 1$:

$$\det((\Lambda_+ * \Lambda_0)(x_i - y_k))_{i,k=1}^p = \int \cdots \int_{t_1 < t_2 < \cdots < t_m \text{ in } \mathbb{R}} \det(\Lambda_+(x_i - t_j))_{i,j=1}^p \det(\Lambda_0(t_j - y_k))_{j,k=1}^p \prod_{j=1}^p d\mu(y_k).$$

We only (require, hence) show the second part for $\gamma > 0$. By Theorem 34.9, $e^{-\gamma s^2/2} \mathcal{B}(\Lambda)(s)$ is the bilateral Laplace transform of a PF function, say Λ_0; and $e^{\gamma s^2/2}$ of a TP (Gaussian) PF function Λ_+, say by Example 29.4. By the preceding part, $\Lambda = \Lambda_+ * \Lambda_0$ is TP. $\square$

35. Further one-sided examples: The Hirschman–Widder densities. Discontinuous PF functions.

Bibliographic notes and references

Section 28 on Pólya frequency functions, and totally non-negative/positive functions, is essentially from the monumental work by Schoenberg [324]. See also the accounts in Karlin's book [200] and in the survey [105]. The modification of Theorem 28.4 (characterizing the TN_2 functions) to arbitrary domains $J-J$ is taken from Khare [215]. Section 28.2, containing the characterization of TN_p functions for all $p \geq 3$ and related results, is by Khare [215], following Lemma 28.9 which is taken from Förster–Kieburg–Kösters [128].

Proposition 29.3, collecting basic properties of Pólya frequency functions, is from Schoenberg [324] – as are the results in Section 29.2 on the variation diminishing properties of these functions. The precursor to these facts is the variation diminishing property for TN and TP matrices. This is the focus of Section 29.1; the treatment in this part can essentially be found in Pinkus' book [282, Chapter 3] – following Gantmacher–Krein [138] and Brown–Johnstone–MacGibbon [72] – as well as the recent work of Choudhury [80]. Section 29.4 on the characterization of TN and TP matrices through sign non-reversal, is taken from Choudhury–Kannan–Khare [81] and Choudhury [80]; see also [114].

Pólya frequency sequences were introduced by Fekete and Pólya [118], in an attempt to prove a result by Laguerre [229]. The treatment in Section 30.2, of generating functions of one-sided Pólya frequency sequences, is taken from the works [4, 5, 106, 107, 325] by Aissen–Schoenberg–Whitney and Edrei, separately and together. Similarly, Corollary 30.24 characterizing finite PF sequences was first shown by Edrei [107], and also follows from the papers cited just above. Theorem 30.20 on the root-location properties of such generating functions, as well as the two subsequent lemmas used in its proof, are from Schoenberg's 1955 paper [325]. The connection to elementary symmetric polynomials is classical; see e.g. Macdonald's monograph [247]. Lemma 30.25, on the total non-negativity of Jacobi matrices, is from Karlin's 1964 paper [199]; and Theorem 30.27 is from Schoenberg's 1955 paper [200].

The proofs in Section 31 of the theorems of Hermite–Biehler and Routh–Hurwitz, as well as their consequences involving Hurwitz matrices, are drawn primarily from the short note [179] by Holtz. The proof of the Hermite–Kakeya–Obrechkoff theorem 31.12 additionally uses arguments in the online book by Fisk [122, Chapter 1]. (The original, classical papers are referred to in the exposition itself.) Theorem **??** ♠ is taken from the work of Garloff and Wagner [140].

The contents of Section 32 (apart from the basics of convolution and the Laplace transform) are from Schoenberg's 1951 paper [324] – including the examples of TN functions: the discontinuous functions $H_1(x)$ and $\lambda_1(x)$, their shifted variants, convolutions of these, and the concluding connection to the Laguerre–Pólya class.

The exposition of the Laguerre–Pólya class and results preceding its development are taken from numerous sources – we mention the surveys [91] by Craven–Csordas and [105] by Dym–Katznelson, the book by Karlin [200], and the paper by Aleman–Beliaev–Hedenmalm [10] for the 'general flavor'. Coming to specific proofs: the treatment of the Hermite–Poulain Theorem 33.3 and Laguerre's theorem 33.8(2) (on the multiplier sequence $q(k)$) are taken from Pinkus's article [280]. The remaining parts of Theorem 33.8 – involving Laguerre's multipliers $1/k!$ and the composition theorems by Maló and Schur – are proved using the approach in Levin's monograph [235]. These proofs also draw from de Bruijn's short note [73], where he showed Theorem 33.9 (on 'roots lying in sectors') and its numerous corollaries listed out above, including the ones by Weisner [366]. (For historical completeness: the Hadamard–Weierstrass factorization emerges primarily out of [159, 364]; see also the monograph [235].)

Theorems 34.5 and 34.6, involving early findings of Pólya about functions $1/\Psi(s)$ for Ψ in the Laguerre–Pólya class, are taken from [285]. Theorem 34.7, showing that $1/\Psi(s)$ has a Laplace transform representation, is taken from Hamburger's paper [161]. Section 34.2, containing characterizations of Pólya frequency functions (both general or one-sided) in terms of the Laguerre–Pólya class or variation diminution, is taken from the two announcements [320, 321] and the subsequent 'full papers' [323, 324] by Schoenberg. Proposition 34.14 on the unbounded support of a Pólya frequency function is from [324], and Theorem 34.16 on the total positivity of certain PF functions is from Schoenberg and Whitney's paper [327].

Part 5:

Composition operators preserving totally positive kernels

Part 5: Composition operators preserving totally positive kernels

36. Critical exponent for powers preserving TN_p. The Jain–Karlin–Schoenberg kernel.

Having classified – in a previous part of this text – the preservers of Loewner positivity, monotonicity, and convexity on infinite domains, we now turn to preservers of total non-negativity and total positivity. This section is concerned with preservers of TN_p for p finite, with emphasis on power functions. In particular, we will see the occurrence of a critical exponent phenomenon in total positivity – this time for powers of one-sided TN_p functions.

36.1. **Connections to representation theory and probability.** We end this second look at critical exponents (the first was in Part 2 of this text) by providing connections to other areas of mathematics – specifically, via the Wallach set (or Gindikin ensemble, or Berezin–Wallach set). The following is a very brief account of these topics, and the references here should provide the reader with starting points for further exploration into this rich area, at the intersection of representation theory, complex analysis, and probability.

Suppose $n \geq 1$ and $D \subset \mathbb{C}^n$ is a tube domain, i.e., of the form $D = \mathbb{R}^n + i\Omega$, where Ω is a homogeneous irreducible self-dual convex cone in $\mathbb{R}^n$. Denote by H_1 the associated Bergman space, consisting of holomorphic functions F on D satisfying:

$$\int_D |F(x+iy)|^2 \, dxdy < \infty,$$

and let H_2 denote the Hardy space, consisting of holomorphic functions F on D satisfying:

$$\sup_{y \in \Omega} \int_{\mathbb{R}^n} |F(x+iy)|^2 \, dx < \infty.$$

Let $P(z - \overline{w})$ denote the Bergman kernel on D; thus $F(w) = \langle F, P_w \rangle_{H_1}$ for all $F \in H_1$, where $P_w(z) := P(z - \overline{w})$. Then the Hardy space H_2 has a reproducing kernel of the form P^p, for some power $p < 1$ of the Bergman kernel. In *Acta Math.* (1976), Rossi and Vergne classified the powers of P which are reproducing kernels for some Hilbert space of holomorphic functions on D. They called the set of such powers p the *Wallach set*, and showed in [303] that it consists of an arithmetic progression and a half-line: $\{0, c/r, 2c/r, \dots, c\} \sqcup (c, \infty)$ for some $c > 0$ and $r \geq 0$. The exact meaning of p, c, r can be found in [303].

Rossi–Vergne named the aforementioned set after Wallach, who was studying it at the time (note, Wallach's papers [359] appeared in print later, in 1979 in *Trans. Amer. Math. Soc.*). Wallach, following up on work of Harish-Chandra, was studying the holomorphic discrete series of connected, simply connected Lie groups G. Specifically, he classified the set of twist-parameters p of the center of K (a maximal compact reductive subgroup of G) for which the corresponding K-finite highest weight module over $\mathfrak{g} = \mathrm{Lie}(G)$ (complexified) is irreducible and unitarizable, or it is reducible and its radical is unitarizable. In [303], Rossi and Vergne obtained the same (Wallach) set of parameters p, with the sufficiently large p leading to the holomorphic discrete series of weighted Bergman spaces. See also [116], where Faraut and Korányi worked over symmetric domains D, and studied Hilbert spaces of holomorphic functions on D.

The Wallach set also appears in at least two other settings, both again in the 1970s:

- Berezin [38] had encountered such a set while studying Kähler potentials of Siegel domains, in the context of quantization. (See also [269] for a recent avatar of this set of exponents, defined in the context of positive semidefinite kernels and recalling various results discussed above).

- Gindikin [143] had shown that given a symmetric cone, and a Riesz distribution R_μ associated to it, R_μ is a positive measure if and only if $\mu \in \mathbb{C}$ lies in an associated Wallach set. A simple proof of this result (and more) was given in 2011 by Sokal [344].

  On a conceptual note: the work of Gindikin on the Wallach set arises through ratios of gamma functions on symmetric cones, which are of the form G/K for a group G and a maximal compact subgroup K. Now, the 'usual' gamma function can be defined via the Laplace transform on the cone $\mathbb{R}^+$. In fact, this can be done over arbitrary symmetric cones in Euclidean Jordan algebras – for instance, in the cone $\mathbb{P}_n(\mathbb{R}) \cong GL_n(\mathbb{R})/O_n(\mathbb{R})$, where $G = GL_n, K = O_n$ stand for the general linear and orthogonal groups of $n \times n$ real matrices, respectively.

  Via the Iwasawa decomposition $G = K \cdot A \cdot N$, carrying out the Laplace transform on such a cone G/K is the same as doing so on $A \cdot N$. This turns out to be a solvable group, and the associated Haar measure is closely related to Lebesgue measure. This provides a 'natural' setting for Gindikin's work and for proving his results – see the 1991 paper of Kostant–Sahi [226] in *Adv. Math.* for details. As the authors remark, the main ingredient in the above working is the Laplace transform on a self-dual cone; the origins of this can be found in the 1935 paper [336] of Siegel in *Ann. of Math.* (see also the book of Hua [185]).

Gindikin's work leads us to another recent manifestation of the critical exponent in matrix analysis – specifically, in random matrix theory. We first mention the 1987 paper [232] of Lassalle in *Invent. Math.*, which approached the same problem through the formalism of positive cones in formally real Jordan algebras. Lassalle recovered the results of Wallach and Rossi–Vergne concerning the Wallach set.

The results of Gindikin, Berezin, Rossi–Vergne, Wallach, and Lassalle parallel a phenomenon for shape parameters of (non-central) Wishart distributions. We begin with the more standard "central" Wishart distribution, defined by Wishart in 1920 [371]. One way to define this is through the Laplace transform of its density function. More precisely, fix an integer $n \geq 1$; now given the shape parameter $p \in [0, \infty)$ and the scale parameter Σ, which is a positive definite $n \times n$ real matrix, $\Gamma(p, \Sigma)$ denotes the Wishart distribution, say with density f, satisfying:

$$\mathcal{L}\{\Gamma(p,\Sigma)\}(s) := \int_{\mathbb{P}_n(\mathbb{R})} e^{-\mathrm{tr}}(sA)\ f(dA) \quad \text{equals} \quad \det(\mathrm{Id}_{n\times n} + 2s\Sigma)^{-p}, \qquad s \in \mathbb{P}_n(\mathbb{R}).$$

It is a well-known fact (see e.g. [117]) that such a distribution exists if and only if p is in the Wallach set $\{0, 1/2, \dots, (n-2)/2\} \sqcup (((n-1)/2), \infty)$.

We now come to recent work along these lines. The *non-central* Wishart distribution is similarly defined – now also using a non-centrality parameter $\Omega \in \mathbb{P}_n(\mathbb{R})$ – via its Laplace transform

$$\mathcal{L}\{\Gamma(p,\Sigma,\Omega)\}(s) = \det(\mathrm{Id}_{n\times n} + 2s\Sigma)^{-p} e^{-2\mathrm{tr}(\Omega s\Sigma(\mathrm{Id}_{n\times n} + 2s\Sigma)^{-1})}, \qquad s \in \mathbb{P}_n(\mathbb{R}).$$

In 2018, Graczyk–Małecki–Mayerhofer [145] and Letac–Massam [234] showed (akin to above) that such a distribution exists if and only if (a) p belongs to the same Wallach set as above:

$$p \in \{0, 1/2, \dots, (n-2)/2\} \sqcup (((n-1)/2), \infty),$$

and (b) if $p < (n-1)/2$ then $\operatorname{rk} \Omega \leq 2p$. The same result was shown by Mayerhofer in *Trans. Amer. Math. Soc.* in 2019, building on ideas of Faraut [115] (1988) and Peddada–St. Richards [277] (1991). To do so, Mayerhofer [255] extended prior analysis by the aforementioned authors, on the positivity of generalized binomial coefficients that occur in Euclidean

Jordan algebras. See e.g. [117] for an introduction to this; also see the 2011 work [308, Theorems 1 and 5] by Sahi, for stronger positivity results.

We end with a (superficial) relation between the Wallach set here and the set of powers preserving positivity on $\mathbb{P}_n((0,\infty))$ (and total non-negativity of order n for the powers of the kernel T_Ω) studied by FitzGerald–Horn, Karlin, and Jain. These powers were studied in Sections 9 and 15 above, as well as in the present section. It would be interesting to find a deeper, conceptual connection between the two problems.

We begin with the positivity preservers side. Recall from the proof of Theorem 9.3 that if $\alpha \in (0, n-2) \setminus \mathbb{Z}$, and $x_1, \dots, x_n \in (0,\infty)$ are pairwise distinct, then the Taylor expansion of the entrywise power $((1 + x_j x_k)^\alpha)_{j,k=1}^n$ yields

$$((1 + x_j x_k)^\alpha)_{j,k=1}^n = \sum_{m \geq 0} \binom{\alpha}{m} x^{\circ m} (x^{\circ m})^T,$$

where $\binom{\alpha}{m} = \alpha(\alpha-1)\cdots(\alpha-m+1)/m!$ and $x^{\circ m} := (x_1^m, \dots, x_n^m)^T$. Now the key is that for $m = \lfloor \alpha \rfloor + 2$, the binomial coefficient is negative. Thus one can pre- and post-multiply the above matrix by u^T, u respectively, for some $u \in \mathbb{R}^n$ orthogonal to the smaller entrywise powers of x. Using this, one can deduce that $((1 + x_j x_k)^\alpha)_{j,k=1}^n$ is not positive semidefinite.

The connection to the Wallach set $\mathcal{W}$ is via the fact that the analysis for powers that do not lie in $\mathcal{W}$ (to show that a (non)central Wishart distribution does not exist) also goes through the negativity of certain generalized binomial coefficients. As a simple example, we look into the argument in the aforementioned work of Peddada–Richards in *Ann. Probab.* 1991. (A similar computation concludes the proof of [255, Theorem 4.10].) Given integers $k_1 \geq \cdots \geq k_n \geq 0$, define the shifted factorial by:

$$(p)_{\mathbf{k}} := \prod_{j=1}^n (p - \tfrac{1}{2}(j-1))_{k_j}, \qquad \text{where} \qquad (p)_k := p(p+1)\cdots(p+k-1)$$

if $k > 0$, and $(p)_0 := 1$. Now it is shown in [277] – via the use of zonal polyomials – that if the (non)central Wishart density with shape parameter p exists, then $(p)_{\mathbf{k}} \geq 0$ for all n-tuples $\mathbf{k}$ as above. But if $q = 2p \in (0, n-1) \setminus \mathbb{Z}$, then set

$$m := \lfloor q \rfloor + 2, \qquad k_1 = \cdots = k_m = 1, \; k_{m+1} = \cdots = k_n = 0.$$

Then the associated generalized binomial coefficient is

$$\frac{q}{2} \frac{q-1}{2} \cdots \frac{q - \lfloor q \rfloor - 1}{2} \cdot (1 \cdot 1 \cdots 1),$$

and this is negative by choice of q.

37. Preservers of Pólya frequency functions.

37.1. **Preservers of TN functions.**

37.2. **Preservers of Pólya frequency functions.**

37.3. **Preservers of TP Pólya frequency functions.**

38. Preservers of Pólya frequency sequences.

Having classified the preservers of TN functions and their subclass of (TP) Pólya frequency functions, we turn to such Toeplitz kernels on distinguished subsets. This section deals with Pólya frequency sequence, i.e., Toeplitz TN kernels on $\mathbb{Z} \times \mathbb{Z}$. However, several of the results will be shown to hold over more general domains $X, Y \subset \mathbb{R}$ with arbitrarily long arithmetic progressions. The full power of these more general domains SX, Y will be revealed in the next section, which concludes this part of the text.

As we saw in the previous three sections, working with kernels on intervals allows one to use powerful tools and results from analysis. These tools are also used in the present section, where we will use PF functions (on $\mathbb{R} \times \mathbb{R}$) to classify the preservers of (TP) PF sequences (on $\mathbb{Z} \times \mathbb{Z}$).

38.1. Preservers of PF sequences.

38.2. Preservers of TP PF sequences.

Discretization: if F is continuous and preserves PF sequences, then F preserves measurable TN Toeplitz kernels.

38.3. Preservers of one-sided PF sequences.

38.4. **Further questions.** To conclude this section, here are a few open questions involving Pólya frequency functions and sequences, and their preservers.

Question 38.1. In light of Schoenberg's theorem 28.4 characterizing the TN_2 functions, classify the preservers of these functions.

For example, the aforementioned theorems imply that all powers x^α preserve the TN_p functions for $p = 2, 3$, if $\alpha \in \mathbb{Z}^{\geq 0} \cup [p-2, \infty)$. We also saw in ♣ that $x^\alpha \mathbf{1}_{x \geq 0}$ is a TN_p function if and only if $\alpha \in \mathbb{Z}^{\geq 0} \cup [p-2, \infty)$. In light of this, a natural question involves studying the powers preserving total non-negativity of each degree:

Question 38.2. Given an integer $p \geq 2$, classify the powers x^α which preserve the class of TN_p functions. Note by ♣ that every such power is in $\mathbb{Z}^{\geq 0} \cup [p-2, \infty)$.

Coming to Pólya frequency sequences, the above classification results lead to additional, theoretical questions about related sequences, which are mentioned for the interested reader.

Question 38.3. Classify the preservers of one-sided Pólya frequency sequences: (i) that have finitely many diagonals, or (ii) generated by evaluating a polynomial at non-negative integers.

Like a question above, both of these classes of functions have nontrivial power-preservers. Indeed, x^n preserves both of these classes for all integers $n \geq 1$, by Maló's theorem 33.8 and a result of Wagner [357], respectively. Akin to the above discussion involving non-integer 'one-sided powers' which are TN_p, a related question is:

Question 38.4. Classify the power functions $x \mapsto x^\alpha$, such that if $\sum_j a_j x^j$ is a polynomial with positive coefficients and real roots, then so is $\sum_j a_j^\alpha x^j$. In particular, find the 'critical exponent' α_p for polynomials of a fixed degree p, such that every $\alpha \geq \alpha_p$ satisfies this property on polynomials of degree at most p.

Note again by Maló's theorem 33.8 that every integer $\alpha \geq 1$ satisfies this property. Recently, Wang and Zhang showed in [363] the existence of the threshold α_p. Notice by ♣ that this question is connected to entrywise powers preserving PF sequences with up to p non-zero diagonals, and hence refines the preceding question.

A final, Whitney-type density question is:

Question 38.5. Are the totally positive Pólya frequency sequences dense in the set of all Pólya frequency sequences?

Such a result could help obtain the preservers of the TP PF sequences from the preservers of all PF sequences. (Note that this goal is achieved above via alternate means.)

39. Preservers of TP Hankel kernels.

In this section, we change gears, and work with a setting studied above for matrices: Hankel kernels, now defined on a sub-interval of $\mathbb{R}$ instead of on the integers $\{0, 1, \dots\}$. As should be clear from the previous sections, working with kernels on intervals (e.g. $\mathbb{R} \times \mathbb{R}$) allows one to use a host of powerful, classical techniques and results from analysis, which help prove results even on discrete domains (e.g. $\mathbb{Z} \times \mathbb{Z}$).

We begin with terminology. Given subsets $X, Y \subset \mathbb{R}$, define their Minkowski sum $X+Y := \{x + y : x \in X, y \in Y\}$. Now a kernel $K : X \times Y \to \mathbb{R}$ is said to be *Hankel* if there is a function $f : X + Y \to \mathbb{R}$ such that $K(x, y) = f(x + y)$ for all $x \in X, y \in Y$. Note that if $X = Y$ then any such 'square' kernel is symmetric.

A natural class of such kernels in analysis is when X is an interval, and in this section we focus on this case. In keeping with the above sections, here is a typical example of a Hankel TN kernel: given finitely many positive scalars $c_1, \dots, c_n, u_1, \dots, u_n$, define

$$K_{\mathbf{c},\mathbf{u}} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}, \qquad K_{\mathbf{c},\mathbf{u}}(x, y) := \sum_{j=1}^{n} c_j u_j^{x+y}.$$

We show in Theorem 39.6 that this kernel is TN on $\mathbb{R} \times \mathbb{R}$ (whence on $X \times X$ for any sub-interval $X \subset \mathbb{R}$).

39.1. **Preservers of Hankel TN kernels on intervals.** In this section, we classify the preservers of Hankel TN and TP kernels, on $X \times X$ for X an interval that is always assumed to have positive measure. The first main result addresses Hankel TN kernels:

Theorem 39.1. *Let $X \subset \mathbb{R}$ be an interval with positive measure, and F a function $: [0, \infty) \to \mathbb{R}$. The following are equivalent:*

(1) The composition map $F \circ -$ preserves total non-negativity on the continuous Hankel TN kernels on $X \times X$.
(2) The composition map $F \circ -$ preserves positive semidefiniteness on the continuous Hankel TN kernels on $X \times X$.
(3) The function F is a power series with non-negative coefficients: $F(x) = \sum_{k=0}^{\infty} c_k x^k$ for $x > 0$, with all $c_k \geq 0$; and $F(0) \geq 0$.

The proof of this result uses a discretization technique that will also be useful later:

Lemma 39.2 (Discretization of Hankel kernels)**.** *Suppose $X \subset \mathbb{R}$ is an interval with positive measure, and $K : X \times X \to \mathbb{R}$. Then each of the following statements implies the next.*

(1) K is TN.
(2) All principal submatrices drawn from K are TN.
(3) All principal submatrices drawn from K, with arguments lying in an arithmetic progression, are TN.
(4) All principal submatrices drawn from K, with arguments lying in an arithmetic progression, are positive semidefinite.

Conversely, (2) $\implies$ (1) *for all kernels,* (3) $\implies$ (2) *for continuous kernels, and* (4) $\implies$ (3) *for continuous Hankel kernels.*

Proof. The forward implications are immediate from the definitions. Conversely, using the notation in Definition 25.1, if $\mathbf{x}, \mathbf{z} \in X^{n,\uparrow}$ for some $n \geq 1$, then the matrix $K[\mathbf{x}; \mathbf{z}]$ is a submatrix of $K[\mathbf{x} \cup \mathbf{z}; \mathbf{x} \cup \mathbf{z}]$, where $\mathbf{x} \cup \mathbf{z}$ denotes the union of the coordinates of $\mathbf{x}$ and $\mathbf{z}$, together arranged in increasing order. Thus (2) $\implies$ (1).

Now suppose K is continuous and (3) holds. We will show (2): given $x_1 < \cdots < x_n$ in X, let $\epsilon := \min_j (x_{j+1} - x_j)/2$, and approximate each x_j by a rational sequence $z_j^{(k)}$ with:

$$z_1^{(k)} \in [x_1, x_1 + \epsilon), \qquad z_n^{(k)} \in (x_n - \epsilon, x_n], \qquad z_j^{(k)} \in (x_j - \epsilon, x_j + \epsilon), \ j \in (1, n).$$

Choose an integer $N_k \geq 1$ such that $z_j^{(k)} \in \frac{1}{N_k}\mathbb{Z}$ for all k, and define:

$$\mathbf{z}^{(k)} := (z_1^{(k)}, z_1^{(k)} + \frac{1}{N_k}, z_1^{(k)} + \frac{2}{N_k}, \dots, z_n^{(k)}), \qquad \mathbf{z}_1^{(k)} := (z_1^{(k)}, z_2^{(k)}, \dots, z_n^{(k)}).$$

By assumption, the matrix $K[\mathbf{z}^{(k)}; \mathbf{z}^{(k)}]$ is TN, whence so is the submatrix $K[\mathbf{z}_1^{(k)}; \mathbf{z}_1^{(k)}]$ for all k. Let $k \to \infty$; since K is continuous, it follows that $K[\mathbf{x}; \mathbf{x}]$ is TN, where $\mathbf{x} = (x_1, \dots, x_n)$. This shows (2).

Finally, suppose K is continuous and Hankel, and (4) holds. Given an arithmetic progression $\mathbf{x} \in X^{n,\uparrow}$, let $A := K[\mathbf{x}; \mathbf{x}]$ be the corresponding positive semidefinite Hankel matrix. Now define the progression of running averages $\mathbf{y} \in X^{n-1,\uparrow}$ by: $y_j := (x_j + x_{j+1})/2$ for $1 \leq j \leq n-1$; and let $B := K[\mathbf{y}; \mathbf{y}]$. Since B is positive semidefinite, Hankel, as well as the truncation $A^{(1)}$ of A, it follows by Theorem 4.1 that A is TN, as desired. $\square$

Now just as Corollary 4.2 shows that Hankel matrices/kernels form a closed convex cone for $X = \{1, \dots, n\}$ (for any integer $n \geq 1$), the same immediately follows in the present setting:

Corollary 39.3. *Suppose $X \subset \mathbb{R}$ is an interval. The continuous Hankel TN kernels on $X \times X$ form a closed convex cone, which is further closed under taking pointwise/Schur products.*

Proof. This follows from the condition Lemma 39.2(4) being closed under addition, dilation, taking pointwise limits, and taking pointwise products (by the Schur product theorem). $\square$

Remark 39.4. The last two conditions of Lemma 39.2 can be further refined to ask for the arithmetic progressions in question to be rational (or to belong to a translate of any dense additive subgroup of $\mathbb{R}$). This does not affect either Lemma 39.2 or Corollary 39.3.

The next preliminary result identifies when a continuous Hankel TN (or TN_p for any $p \geq 2$) kernel vanishes at a point. Recall for an interval $X \subset \mathbb{R}$ that ∂X denotes its boundary, i.e., the set of endpoints of X.

Lemma 39.5. *Let $X \subset \mathbb{R}$ be an interval of positive length, and $K : X \times X \to \mathbb{R}$ a Hankel, TN_2 kernel. If $K(x, y) = 0$ for some $x, y \in X$, then K vanishes outside 'corners', i.e., on $X \times X \setminus \{(x_0, x_0) : x_0 \in \partial X\}$. If moreover K is continuous, then $K \equiv 0$ on $X \times X$.*

Proof. Suppose K is as given, and X has interior (a, b) for $-\infty \leq a < b \leq \infty$. Suppose some $K(x, y) = 0$; then so is $K(d_0, d_0)$ for $d_0 = (x + y)/2$, as K is Hankel. Again by this property, it suffices to show $K(d, d) = 0 \ \forall d \in X \setminus \partial X$. We show this for $d \in (a, d_0)$; the proof is similar for $d \in (d_0, b)$.

Let $d \in (a, d_0)$; the TN_2 property of $K[(d, d_0); (d, d_0)]$ gives

$$0 \leq K(d, d_0)^2 \leq K(d, d) K(d_0, d_0) = 0.$$

This shows $K((d + d_0)/2, (d + d_0)/2) = 0$. Now if $a = -\infty$ then running over all d we are done. Else say a is finite. Then the above argument shows that

$$K(d, d) = 0, \qquad \forall d \in ((a + d_0)/2, d_0).$$

Now define the sequence

$$d_{n+1} := (a + 3d_n)/4 \in ((a + d_n)/2, d_n), \qquad n \geq 0.$$

Clearly d_n decreases from d_0 to a^+. Now claim by induction that $K(d, d) = 0$ for all $d \in [d_{n+1}, d_n)$. The base case of $n = 0$ was shown above, and the same computations show the induction step as well. Finally, the last assertion is now immediate. $\square$

With these preliminaries in hand, we can show the above classification result.

Proof of Theorem 39.1. Clearly (1) $\implies$ (2); we next show (3) $\implies$ (1). Suppose (3) holds and $K : X \times X \to \mathbb{R}$ is Hankel and TN. There are two cases. First, K vanishes at a point and hence $K \equiv 0$ by Lemma 39.5, in which case $F(0) \geq 0$ gives: $(F \circ K)(0) = F(0)\mathbf{1}_{X \times X}$ is TN. Otherwise $F > 0$ on $X \times X$, in which case $F \circ K$ is again continuous, Hankel, and TN by Corollary 39.3.

It remains to show (2) $\implies$ (3). First if $K \equiv 0$ then $F(0)\mathbf{1}_{X \times X} = (F \circ K)$ is positive semidefinite, so $F(0) \geq 0$. Otherwise $K > 0$ by Lemma 39.5, We now appeal to Theorem 19.1 and Remark 19.18. Thus, it suffices to show that $F[-]$ preserves positivity on the matrices $(a + bu_0^{j+k})_{j,k=0}^n$ for all $a, b \geq 0$ with $a + b > 0$ and all n, as well as on all rank-1 and all Toeplitz matrices in $\mathbb{P}_2((0, \infty))$. For this, it suffices to produce continuous Hankel TN kernels on $\mathbb{R} \times \mathbb{R}$ which contain the given test set of matrices at equispaced arguments.

First, by the assumptions there exist linear maps $\varphi_n : \mathbb{R} \to \mathbb{R}, \ n \geq 1$ with positive slopes such that $[0, n] \subset \varphi_n(X)$. Now consider the continuous kernel

$$K_n(x, y) := a + bu_0^{\varphi_n(x)+\varphi_n(y)}, \qquad x, y \in \mathbb{R}.$$

This is a rank-2 kernel, and easily verified to be Hankel and TN on $\mathbb{R} \times \mathbb{R}$, whence on $X \times X$. Applying F, it follows that the matrix

$$(F \circ K_n)[\mathbf{x}; \mathbf{x}] = (F(a + bu_0^{j+k}))_{j,k=0}^n$$

is positive semidefinite as desired. Here $\mathbf{x} := (\varphi_n^{-1}(0), \dots, \varphi_n^{-1}(n)) \in X^{n+1,\uparrow}$ for all $n \geq 0$.

Next, if $A = \begin{pmatrix} p & q \\ q & q^2/p \end{pmatrix}$ has positive entries and rank one, then consider the continuous Hankel kernel associated to the measure $p\delta_{q/p}$, i.e., $K(x, y) := p(q/p)^{x+y}$ for $x, y \in \mathbb{R}$.

Finally, consider the Toeplitz matrix $\begin{pmatrix} b & a \\ a & b \end{pmatrix}$, with $0 < a < b$. It suffices to produce a continuous Hankel TN kernel containing this matrix. While one can use Theorem 7.4, we provide a direct proof as well. By rescaling by b, we may assume $b = 1$. Now choose any $\alpha \in (1 - a^2, 1)$ and consider the continuous Hankel kernel

$$K(x, y) := \alpha \left(a - \sqrt{\frac{(1 - \alpha)(1 - a^2)}{\alpha}} \right)^{x+y} + (1 - \alpha) \left(a + \sqrt{\frac{\alpha(1 - a^2)}{1 - \alpha}} \right)^{x+y},$$

for $x, y \in \mathbb{R}$. It is easy to check that $K(0, 0) = K(1, 1) = 1$ and $K(0, 1) = a$. Moreover, K is TN because we reduce to Theorem 4.1 via Lemma 39.2(4). $\square$

39.2. **Structure of Hankel TP kernels on intervals.** We next turn to Hankel TP kernels. For this, we need to understand both Hankel TN and TP kernels in greater detail:

Theorem 39.6. *Suppose $X \subset \mathbb{R}$ is an non-empty open interval.*

(1) *The following are equivalent for $K : X \times X \to \mathbb{R}$ a continuous Hankel kernel:*
   (a) *K is TN.*

(b) K is positive semidefinite.
(c) There exists a non-decreasing function $\sigma : \mathbb{R} \to \mathbb{R}$ such that $K(x,y) = \int_{\mathbb{R}} e^{-(x+y)u}\, d\sigma(u)$ for all $x, y \in X$.

Furthermore, the kernel K is TP if and only if the non-negative measure associated to σ has infinite support.

(2) The continuous Hankel TP kernels on $X \times X$ are dense in the continuous Hankel TN kernels on $X \times X$.
(3) The space of continuous Hankel TP kernels on $X \times X$ is a convex cone, which is further closed under taking pointwise/Schur products.

The first part is a representation theorem by Widder (1940) in *Bull. Amer. Math. Soc.*, which solves a moment problem in the spirit of Hamburger, Hausdorff, Stieltjes, and others, but now for 'exponential moments' of non-negative measures on the real line. These are termed 'exponentially convex functions' by Bernstein. The second part reveals a Whitney-type density result for Hankel kernels on an interval, in the spirit of Whitney's theorem 6.7 for matrices. In the next few sections, we will see similar variants for other structured kernels.

We now turn to Widder's proof of Theorem 39.6(1). This uses an intermediate notion of 'kernels of positive type', which are now introduced.

Definition 39.7. Given real numbers $a \leq b$, a continuous symmetric function $K : [a,b] \times [a,b] \to \mathbb{R}$ is said to be a *kernel of positive type on* $[a,b]^2$ if for all continuous functions $\xi : [a,b] \to \mathbb{R}$, we have

$$\int_a^b \int_a^b K(s,t)\xi(s)\xi(t)\ ds\ dt \geq 0.$$

If now $I \subset \mathbb{R}$ is a sub-interval, and $K : I \times I \to \mathbb{R}$ is continuous, we say K is of positive type if it is so on every closed sub-interval of I.

The following result relates positive semidefinite kernels with kernels of positive type.

Lemma 39.8 (Mercer, 1909). *Suppose $a \leq b$ are real numbers and $K : [a,b] \times [a,b] \to \mathbb{R}$ is continuous. Then K is a positive semidefinite kernel if and only if K is of positive type.*

Thus, Mercer's lemma provides an alternate equivalent condition to Theorem 39.6(1) for X compact. It was shown by Mercer in *Phil. Trans. Royal Soc. A* (1909), en route to showing the following famous result (which we do not use, nor pursue in this text):

Theorem 39.9 (Mercer, 1909). *Suppose $K : [a,b]^2 \to \mathbb{R}$ is a kernel of positive type. Then there exists an orthonormal basis $\{\mathbf{e}_j : j \geq 1\}$ of $L^2[a,b]$, such that: (a) each $\mathbf{e}_j$ is an eigenfunction of the integral operator $T_K\varphi(x) := \int_a^b K(x,s)\varphi(s)\ ds$; (b) the corresponding eigenvalue λ_j is non-negative; (c) if $\lambda_j > 0$ then $\mathbf{e}_j$ is continuous; and (d) K has the representation*

$$K(s,t) = \sum_{j \geq 1} \lambda_j \mathbf{e}_j(s)\mathbf{e}_j(t).$$

Proof of Mercer's lemma 39.8. If K is a positive semidefinite kernel, and $\xi : [a,b] \to \mathbb{R}$ is continuous, then the double integral

$$\int_a^b \int_a^b K(s,t)\xi(s)\xi(t)\ ds\ dt$$

can be expressed as a limit of Riemann sums:

$$\lim_{n\to\infty} \frac{(b-a)^2}{n^2} \sum_{j,k=0}^{n-1} K(s_j, s_k)\xi(s_j)\xi(s_k) = \lim_{n\to\infty} u_n^T K_n u_n,$$

where

$$s_j := a + j(b-a)/n, \qquad K_n := (K(s_j, s_k))_{j,k=0}^{n-1}, \qquad u_n = \frac{b-a}{n}(\xi(s_1), \dots, \xi(s_n))^T.$$

But the right-hand limit is non-negative since K is positive semidefinite.

Conversely, suppose there exist scalars $a \leq s_1 < \cdots < s_n \leq b$ such that the matrix $K_n := (K(s_j, s_k))_{j,k=0}^{n}$ is not positive semidefinite. Since K is continuous, a perturbation argument allows us to assume $s_0 > a$ and $s_n < b$. Thus there exists an eigenvector $u_n = (\xi_0, \dots, \xi_n)^T$ such that $u_n^T K_n u_n =: A < 0$.

We now construct a 'toothsaw' test function, as follows: we will choose $\epsilon, \eta > 0$, running only over values satisfying:

$$\epsilon + \eta < \min\{s_0 - a, (s_2 - s_1)/2, \dots, (s_n - s_{n-1})/2, b - s_n\}.$$

Define $\theta_{\epsilon,\eta}(x)$ to be the unique piecewise linear, continuous function on (a,b) such that

$$\theta_{\epsilon,\eta}(x) := \begin{cases} \xi_j, & \text{if } x \in (s_j - \eta, s_j + \eta), \\ 0, & \text{if } x \in (a, s_0 - \epsilon - \eta] \sqcup [s_n + \epsilon + \eta, b), \\ 0, & \text{if } x \in \bigsqcup_{j=1}^n [s_{j-1} + \epsilon + \eta, s_j - \epsilon - \eta], \end{cases}$$

and $\theta_{\epsilon,\eta}$ is linear on all remaining sub-intervals $[s_j - \epsilon - \eta, s_j - \eta]$ and $[s_j + \eta, s_j + \epsilon + \eta]$.

Now define $F_n : [-\eta, \eta]^2 \to \mathbb{R}$ via

$$F_n(x,y) := \sum_{j,k=0}^{n} K(s_j + x, s_k + y)\ \xi_j \xi_k.$$

A straightforward computation yields:

$$\begin{aligned} \int_a^b \int_a^b K(s,t)\theta_{\epsilon,\eta}(s)\theta_{\epsilon,\eta}(t)\ ds\ dt &= \sum_{j,k=0}^{n} \int_{x_j-\epsilon-\eta}^{x_j+\epsilon+\eta} \int_{x_k-\epsilon-\eta}^{x_k+\epsilon+\eta} K(s,t)\theta_{\epsilon,\eta}(s)\theta_{\epsilon,\eta}(t)\ ds\ dt \\ &= \int_{-\eta}^{\eta} \int_{-\eta}^{\eta} F_n(x,y)\ dx\ dy + J_1, \end{aligned}$$

where

$$J_1 = \sum_{j,k=0}^{n} \iint_{Q_{jk}} K(s,t)\theta_{\epsilon,\eta}(s)\theta_{\epsilon,\eta}(t)\ ds\ dt,$$

with Q_{jk} the 'annular' region between the squares

$$[x_j - \epsilon - \eta, x_j + \epsilon + \eta] \times [y_k - \epsilon - \eta, y_k + \epsilon + \eta] \quad \text{and} \quad [x_j - \eta, x_j + \eta] \times [y_k - \eta, y_k + \eta].$$

Now for $(s,t) \in Q_{jk}$ we have $|\theta_{\epsilon,\eta}(s)\theta_{\epsilon,\eta}(t)| \leq |\xi_j \xi_k|$, so by an easy computation:

$$|J_1| \leq 4\epsilon(2\eta + \epsilon) \left(\sum_{j=0}^{n} |\xi_j| \right)^2 \cdot \|K\|_\infty$$

where $\|K\|_\infty$ is the supnorm of K on the compact domain $[a,b]^2$. Note that $F_n(0,0) = A < 0$ (as above), so by continuity there exists a small $\eta > 0$ such that $F_n(x,y) < A/2$ when $|x|, |y| < \eta$. Fixing such an η,

$$\int_a^b \int_a^b K(s,t)\theta_{\epsilon,\eta}(s)\theta_{\epsilon,\eta}(t)\ ds\ dt \leq |J_1| + \int_{-\eta}^{\eta}\int_{-\eta}^{\eta} A/2\ dx\ dy$$
$$< 4\epsilon(2\eta+\epsilon)\left(\sum_{j=0}^n |\xi_j|\right)^2 \|K\|_\infty + 2A\eta^2,$$

for ϵ sufficiently small. Choose $\epsilon > 0$ small enough such that the right-hand side is negative (since $A < 0$); this yields the desired contradiction. $\square$

Another preliminary result required to show Theorem 39.6 is by Bernstein in his 1926 memoir. This is a 'dimension-free' analogue of the Boas–Widder theorem 18.10(2) – also recall the related Theorem 19.9 by Boas. In fact, Boas and Widder write that they were motivated to prove their Theorem 18.10(2) 'in an effort to make more accessible' the following result of Bernstein, which Widder used in proving Theorem 39.6(1).

Theorem 39.10 (Bernstein, 1926)**.** *Given a sub-interval $(a,b) \subset \mathbb{R}$ and a continuous function $f : (a,b) \to \mathbb{R}$, if the even-order forward differences*

$$(\Delta_\delta^{2n} f)(c) := \sum_{j=0}^{2n} \binom{2n}{j} (-1)^j f(c+j\delta), \qquad c \in (a,b),\ \delta \in (0,(b-c)/2n)$$

are all non-negative, then f is analytic in (a,b).

The final preliminary result is by Hamburger (1920) in *Math. Z.*:

Proposition 39.11 (Hamburger)**.** *If $f(x)$ is analytic in (a,b) and there exists $c \in (a,b)$ such that the semi-infinite Hankel matrix $(f^{(j+k)}(c))_{j,k\geq 0}$ is positive semi-definite, then*

$$f(x) = \int_{\mathbb{R}} e^{-xu}\ d\sigma(u)$$

for some non-decreasing function σ, with the integral converging on $x \in (a,b)$.

These two results are used without proofs.

Proof of Theorem 39.6.

(1) That (a) and (b) are equivalent follows by Lemma 39.2. Next assume (b) holds, let $[a,b] \subset X$, and let the continuous function $f : X + X \to \mathbb{R}$ be given by: $f(x+y) = K(x,y)$ for $x,y \in X$. Suppose $a \leq c < c + 2n\delta \leq b$ for some integer $n > 0$ and scalar $0 < \delta < \frac{b-c}{2n}$. The Hankel matrix $K_{n,c,\delta} := (f(c+j\delta+k\delta))_{j,k=0}^n$ is positive semidefinite by assumption; we evaluate it against the vector

$$\xi = (\xi_0, \dots, \xi_n)^T, \qquad \xi_j := \sum_{l=j}^n (-1)^{j+1}\binom{j}{l}\eta_j$$

for some scalars $\eta_0, \dots, \eta_n$. In the language of forward differences, this yields:

$$\xi^T K_{n,c,\delta}\xi = \sum_{j,k=0}^n \Delta_\delta^{j+k} f(c)\eta_j\eta_k \geq 0.$$

Now set $\eta_0 = \eta_1 = \cdots = \eta_{n-1} = 0$, $\eta_n = 1$, we have $(\Delta_\delta^{2n} f)(c) \geq 0$. By Theorem 39.10, since f is continuous, f is analytic in (a,b). Now replacing η_j in the preceding computation by η_j/δ^j, we obtain as $\delta \to 0^+$:

$$\sum_{j,k=0}^{n} f^{(j+k)}(c)\eta_j\eta_k \geq 0.$$

As this holds for all integers $n \geq 0$ and all $\eta_0, \dots, \eta_n \in \mathbb{R}$, Proposition 39.11 implies f is of the desired form.

Conversely, if (c) holds, then to show that a principal submatrix drawn from K at arguments $x_0 < x_1 < \cdots < x_n$ in X is positive semidefinite, it suffices to consider K restricted to $[a,b]^2$, where $a = x_0$ and $b = x_n$. Now by Mercer's lemma 39.8, it suffices to show that $K|_{[a,b]^2}$ is of positive type. But this is straightforward: given a continuous function $\xi : [a,b] \to \mathbb{R}$, we have

$$\begin{aligned}\int_{a/2}^{b/2}\int_{a/2}^{b/2} f(s+t)\xi(s)\xi(t)\ ds\ dt &= \int_{a/2}^{b/2}\int_{a/2}^{b/2} \xi(s)\xi(t)\int_{\mathbb{R}} e^{-(s+t)u}\ d\sigma(u)\ ds\ dt\\ &= \int_{\mathbb{R}} \left(\int_{a/2}^{b/2} e^{-su}\xi(s)\ ds\right)^2 d\sigma(u) \geq 0.\end{aligned}$$

The final step involves a change of order of integration, which is justified because the integral representation of f converges uniformly in $[a,b]$.

This ends the proof of the equivalence. Now suppose the measure associated to σ has infinite support. Then the kernel $K(x,y) = f(x+y)$ is TP, by Proposition 6.2. If needed, we can use the fact that a kernel $K(x,y)$ on $X \times Y$ (for $X, Y \subset \mathbb{R}$) is TN/TP if and only if the kernel $K(-x,-y)$ is so, because drawing square submatrices from one or the other kernel are equivalent, modulo applying the order reversing permutation to the rows as well as the columns.

Next, suppose the measure for σ has finite support, say with mass c_k at $u_k \in \mathbb{R}$ for $k = 1, \dots, r$. Given two n-tuples of points $\mathbf{x}, \mathbf{y} \in X^{n,\uparrow}$, we see that

$$K[\mathbf{x};\mathbf{y}] = \sum_{k=1}^{r} c_k (e^{-(x_i+y_j)u_k})_{i,j=1}^{n} = \sum_{k=1}^{r} c_k \mathbf{x}_0^{\circ u_k} {\mathbf{y}_0^{\circ u_k}}^T, \tag{39.12}$$

where $\mathbf{x}_0 = (e^{-x_i})_{i=1}^n, \mathbf{y}_0 = (e^{-y_j})_{j=1}^n$. Thus $K[\mathbf{x};\mathbf{y}]$ has rank $\leq r$, hence is singular if $n > r$. It follows that K is not TP.

(2) Given a continuous Hankel TN kernel $K(x,y) = f(x+y)$ on $X \times X$, with the function f as in the previous part, K is the limit as $\epsilon \to 0^+$, of the kernels

$$K_\epsilon(x,y) := K(x,y) + \epsilon \int_0^1 e^{-(x+y)u}\ du.$$

Since each underlying measure in K_ϵ has infinite support, K_ϵ is TP by the previous part.

(3) Note that TP kernels are closed under dilations. If now K, K' are TP, with underlying representative functions σ, σ', then the measures corresponding to these have infinite supports, whence the same holds for $\sigma + \sigma'$. Hence $K + K'$ is also TP. Finally, $K \cdot K'$ is TN by Corollary 39.3. If it is not TP, then the representative function τ has underlying measure of finite support, say of size r. Now choose an arithmetic progression $\mathbf{x} \in X^{r+1,\uparrow}$; then the principal submatrices $K[\mathbf{x};\mathbf{x}]$ and $K'[\mathbf{x};\mathbf{x}]$ are Hankel

by the choice of $\mathbf{x}$ and TP by assumption. Hence so is their Schur product, which is precisely $(K \cdot K')[\mathbf{x}; \mathbf{x}]$, by Theorem 4.1. In particular, this is a principal submatrix of $K \cdot K'$ of rank $r + 1$, which contradicts the choice of r, say by (39.12) for $K \cdot K'$. □

39.3. **Preservers of Hankel TP kernels.** To conclude, we classify the preservers of Hankel TP kernels, parallel to the TN version in Theorem 39.1.

Theorem 39.13. *Let $X \subsetneq \mathbb{R}$ be an open sub-interval with positive measure, and a function $F : (0, \infty) \to \mathbb{R}$. The following are equivalent:*

(1) The composition map $F \circ -$ preserves total positivity on the continuous Hankel TP kernels on $X \times X$.
(2) The composition map $F \circ -$ preserves positive definiteness on the continuous Hankel TP kernels on $X \times X$.
(3) The function F is a power series with non-negative coefficients: $F(x) = \sum_{k=0}^{\infty} c_k x^k$ for $x > 0$, with all $c_k \geq 0$; and F is non-constant.

Proof. Clearly (1) $\implies$ (2). Next, suppose (3) holds, with $c_{n_0} > 0$ for some $n_0 > 0$. Let $K : X \times X \to \mathbb{R}$ be a continuous TP Hankel kernel, then so is K^{n_0} by Theorem 39.6. Moreover, Corollary 39.3 shows $G \circ K$ is a continuous TN Hankel kernel, where $G(x) := F(x) - c_{n_0} x^{n_0}$. Now $G \circ K$ and K^{n_0} have integral representations as above, say with corresponding non-negative measures ν and μ respectively. Since μ has infinite support from above, so does $\nu + c_{n_0}\mu \geq 0$. Hence $F \circ K = G \circ K + c_{n_0} K^{n_0}$ is TP.

It remains to show (2) $\implies$ (3). First note by Theorem 7.4 that every 2×2 TP matrix occurs as a 'principal submatrix' of a continuous Hankel TP kernel on $\mathbb{R} \times \mathbb{R}$, drawn from a function evaluated at equispaced arguments. Hence by Lemma 12.14, f is continuous, positive, and strictly increasing on $(0, \infty)$. It follows that f preserves continuous Hankel TN kernels, hence is of the desired form by Theorem 39.6(2). Clearly f is non-constant, and the proof is complete. □

40. Total positivity preservers: All kernels.

We now bring together many of the techniques and results discussed above – not only in this part, but in the previous parts of this text – to solve the motivating problem in this part: *Classify the functions preserving totally non-negative/positive kernels on $X \times Y$, where X, Y are (arbitrary) non-empty totally ordered sets.*

Recall the characterizations in Theorems 12.11 and 12.13, which resolved this question for X, Y finite. Similarly, Theorem 12.15 and Corollary 12.17 respectively classified the preservers of TN and TP symmetric kernels, for $X = Y$ finite.

40.1. Finite-continuum kernels.

In this section, we answer the above question when at least one of X, Y is infinite. The first step is to resolve this when exactly one of X, Y is finite and the other is infinite; in this case we do not consider symmetric kernels.

The key result which is required to solve the classification question is a recent extension of Whitney's density theorem, which uses discretized Gaussian convolution.

Theorem 40.1. *Given an integer $p \geq 2$, and a bounded TN_p kernel K on $\mathbb{R} \times \mathbb{R}$, let $C \subset \mathbb{R}^2$ denote the points of continuity of K. Then there exists a sequence of TP_p kernels $(K_l)_{l \geq 1}$ that converge to K locally uniformly on C. If moreover K is 'symmetric', i.e., $K(x, y) = K(y, x)$ for all $x, y \in \mathbb{R}$, then the sequence K_l may also be taken to be symmetric for all $l \geq 1$.*

Theorem 40.1 is due to Belton, Guillot, Khare, and Putinar (2020), for arbitrary subsets $X, Y \subset \mathbb{R}$ – which is not more general because one can always extend such a kernel to one on $\mathbb{R} \times \mathbb{R}$, by padding by zeros. Notice also that the assumption that $X = Y = \mathbb{R}$ is itself not unnecessarily restrictive, given Lemma 7.2. (For $p = 1$, a TP_1 kernel is merely a positive-valued function, and so $K + \frac{1}{l}\mathbf{1}$ approximates any TN_1 kernel K.)

Bibliographic notes and references

Most of the material in this part is taken from Belton–Guillot–Khare–Putinar [32], and we discuss the remaining references.

Theorem 39.6(1) is a representation theorem for totally non-negative continuous Hankel kernels on an open interval, shown by Widder [369]. Mercer's lemma 39.8 and theorem 39.9 are from [261]. Bernstein's theorem 39.10 guaranteeing analyticity from the positivity of even-order forward differences is from [43]. Hamburger's theorem 39.11 is from [161]. (See also [44] and [8, Theorem 5.5.4].)

Part 6:

Entrywise polynomials preserving positivity in a fixed dimension

Part 6: Entrywise polynomials preserving positivity in a fixed dimension

41. Entrywise polynomial preservers. Horn–Loewner type necessary conditions. Classification of sign patterns.

In Part 3 of this text, we classified the entrywise functions preserving positivity in all dimensions; these are precisely the power series with non-negative coefficients. Earlier in Part 2, we had classified the entrywise powers preserving positivity (as well as total positivity and total non-negativity) in fixed dimension. In this final part of the text, we study polynomials that entrywise preserve positive semidefiniteness in fixed dimension.

Recall from the Schur product theorem 3.12 and its converse, the Schoenberg–Rudin theorem 16.2, that the only polynomials that entrywise preserve positivity in all dimensions are the ones with all non-negative coefficients. Thus, if one fixes the dimension $N \geq 3$ of the test set of positive matrices, then it is reasonable to expect that there should exist more polynomial preservers – in other words, polynomial preservers with negative coefficients. However, this problem remained completely open until very recently (~ 2016): *not a single* polynomial preserver was known with a negative coefficient, nor was a non-existence result proved!

In this final part, we answer this existence question as well as stronger variants of it. Namely, not only do we produce such polynomial preservers, we also fully resolve the more challenging question: *which coefficients of polynomial preservers on $N \times N$ matrices can be negative?* Looking ahead in this part:

- We classify the sign patterns of entrywise polynomial preservers on $\mathbb{P}_N$ for fixed N.
- We extend this to all power series; but also, countable sums of real powers, such as $\sum_{\alpha \in \mathbb{Q},\ \alpha \geq N-2} c_\alpha x^\alpha$. This case is more subtle than that of polynomial preservers.
- We will also completely classify the sign patterns of polynomials that entrywise preserve totally non-negative (TN) Hankel matrices of a fixed dimension. Recall from the discussions around Theorems 12.19 and 19.1 that this is expected to be very similar to (maybe even the same as) the classification for positivity preservers.

In what follows, we work with $\mathbb{P}_N((0,\rho))$ for $N > 0$ fixed and $0 < \rho < \infty$. Since we work with polynomials and power series, this is equivalent to working over $\mathbb{P}_N([0,\rho))$ by density and continuity. If $\rho = +\infty$, one can prove results that are similar to the ones shown in this part of the text; but for a first look at the proofs and techniques used, we restrict ourselves to $\mathbb{P}_N((0,\rho))$. For full details of the $\rho = +\infty$ case, as well as for the proofs, ramifications, and applications of the results below, we refer the reader to the paper "*On the sign patterns of entrywise positivity preservers in fixed dimension*" in *Amer. J. Math.* by Khare and Tao.

41.1. **Horn–Loewner-type necessary condition; matrices with negative entries.** In this section and beyond, we work with polynomials or power series

$$f(x) = c_{n_0} x^{n_0} + c_{n_1} x^{n_1} + \cdots, \qquad \text{with } n_0, n_1, \ldots \text{ pairwise distinct} \tag{41.1}$$

and $c_{n_j} \in \mathbb{R}$ typically non-zero. Recall the (stronger) Horn–Loewner theorem 17.1, which shows that if $f \in C^{N-1}(I)$ for $I = (0,\infty)$, and $f[-]$ preserves positivity on (rank-2 Hankel TN matrices in) $\mathbb{P}_N(I)$, then $f, f', \ldots, f^{(N-1)} \geq 0$ on I. In the special case that f is a polynomial or a power series, one can say more, and under weaker assumptions:

Lemma 41.2 (Horn–Loewner-type necessary condition). *Fix an integer $N > 0$. Let $\rho > 0$ and $f : (0,\rho) \to \mathbb{R}$ be a function of the form* (41.1) *satisfying:*

1. *f is absolutely convergent on $(0,\rho)$, i.e., $\sum_{j \geq 0} |c_{n_j}| x^{n_j} < \infty$ on $(0,\rho)$.*
2. *$f[-]$ preserves positivity on rank-1 Hankel TN matrices in $\mathbb{P}_N((0,\rho))$.*

If $c_{n_{j_0}} < 0$ for some $j_0 \geq 0$, then $c_{n_j} > 0$ for at least N values of j for which $n_j < n_{j_0}$.

Remark 41.3. In both (41.1) as well as Lemma 41.2 (and its proof), we have deliberately **not** insisted on the exponents n_j being non-negative integers. In fact, one can choose $\{n_j : j \geq 0\}$ to be an arbitrary sequence of pairwise distinct real numbers.

Proof of Lemma 41.2. By the properties of f, the function

$$g(x) := \sum_{j \neq j_0 : c_{n_j} < 0} |c_{n_j}| x^{n_j}$$

entrywise preserves positivity on rank-1 Hankel TN matrices in $\mathbb{P}_N((0,\rho))$. Hence, so does

$$f(x) + g(x) = \sum_{j : c_{n_j} > 0} c_{n_j} x^{n_j} + c_{n_{j_0}} x^{n_{j_0}}.$$

Now suppose the result is false. Then the preceding sum contains at most k terms n_j that lie in $(0, n_{j_0})$ (for some $0 \leq k < N$), and which we label by $n_0, \dots, n_{k-1}$. Also, set $m := n_{j_0}$. Choose any $u_0 \in (0,1)$ and define $\mathbf{u} := (1, u_0, \dots, u_0^{N-1})^T \in \mathbb{R}^N$. Then $\mathbf{u}^{\circ n_0}, \dots, \mathbf{u}^{\circ n_{k-1}}; \mathbf{u}^{\circ m}$ are linearly independent, forming (some of) the columns of a generalized Vandermonde matrix. Hence, there exists $\mathbf{v} \in \mathbb{R}^N$, such that

$$\mathbf{v} \perp \mathbf{u}^{\circ n_0}, \dots, \mathbf{u}^{\circ n_{k-1}} \quad \text{and} \quad \mathbf{v}^T \mathbf{u}^{\circ m} = 1.$$

For $0 < \epsilon < \rho$, we let $A_\epsilon := \epsilon \mathbf{u}\mathbf{u}^T$, which is a rank-1 Hankel moment matrix in $\mathbb{P}_N((0,\rho))$ (and hence TN). Now compute using the hypotheses:

$$\begin{aligned}
0 \leq \mathbf{v}^T (f+g)[A_\epsilon] \mathbf{v} &= \mathbf{v}^T \left(\sum_{j : c_{n_j} > 0} c_{n_j} \epsilon^{n_j} \mathbf{u}^{\circ n_j} (\mathbf{u}^{\circ n_j})^T + c_m \epsilon^m \mathbf{u}^{\circ m} (\mathbf{u}^{\circ m})^T \right) \mathbf{v} \\
&= c_m \epsilon^m (\mathbf{v}^T \mathbf{u}^{\circ m})^2 + \sum_{j : c_{n_j} > 0,\ n_j > n_{j_0}} c_{n_j} \epsilon^{n_j} (\mathbf{v}^T \mathbf{u}^{\circ n_j})^2 \\
&= c_m \epsilon^m + o(\epsilon^m).
\end{aligned}$$

Thus, $0 \leq \lim_{\epsilon \to 0^+} \dfrac{\mathbf{v}^T (f+g)[A_\epsilon] \mathbf{v}}{\epsilon^m} = c_m < 0$, which is a contradiction. Hence, $k \geq N$, proving the claim. $\square$

By Lemma 41.2, every polynomial that entrywise preserves positivity on $\mathbb{P}_N((0,\rho))$ must have its N non-zero Maclaurin coefficients of "lowest degree" to be positive. The obvious question is if any of the other terms can be negative, e.g., the immediate next coefficient.

We tackle this question in the remainder of this text, and show that, in fact, every other coefficient can indeed be negative. For now, we point out that working with positive matrices with other entries *cannot* provide such a structured answer (in the flavor of Lemma 41.2). As a simple example, consider the family of polynomials

$$p_{k,t}(x) := t(1 + x^2 + \cdots + x^{2k}) - x^{2k+1}, \qquad t > 0,$$

where $k \geq 0$ is an integer. Now claim that $p_{k,t}[-]$ can never preserve positivity on $\mathbb{P}_N((-\rho,\rho))$ for $N \geq 2$. Indeed, if $\mathbf{u} := (1, -1, 0, \dots, 0)^T$ and $A := (\rho/2)\mathbf{u}\mathbf{u}^T \in \mathbb{P}_N((-\rho,\rho))$, then

$$\mathbf{u}^T p_{k,t}[A] \mathbf{u} = -4(\rho/2)^{2k+1} < 0.$$

Therefore, $p_{k,t}[A]$ is not positive semidefinite for any $k \geq 0$. If one allows complex entries, similar examples with higher-order roots of unity can be constructed, in which such negative results (compared to Lemma 41.2) can be obtained.

41.2. **Classification of sign patterns for polynomials.** In light of the above discussion, henceforth we restrict ourselves to working with matrices in $\mathbb{P}_N((0,\rho))$ for $0 < \rho < \infty$. By Lemma 41.2, every polynomial preserver on $\mathbb{P}_N((0,\rho))$ must have its N lowest-degree Maclaurin coefficients (which are non-zero) to be positive.

We are interested in understanding if any (or every) other coefficient can be negative. If, say, the next lowest-degree coefficient could be negative, this would achieve two goals:

- It would provide (the first example of) a polynomial preserver in fixed dimension, which has a negative Maclaurin coefficient.
- It would provide (the first example of) a polynomial that preserves positivity on $\mathbb{P}_N((0,\rho))$, but necessarily *not* on $\mathbb{P}_{N+1}((0,\rho))$. In particular, this would show that the Horn–Loewner-type necessary condition in Lemma 41.2 is "best possible". (See Remark 17.3 in the parallel setting of entrywise power preservers for the original Horn condition.)

We show in this part of the text that these goals are indeed achieved:

Theorem 41.4 (Classification of sign patterns, fixed dimension)**.** *Fix integers $N > 0$ and $0 \leq n_0 < n_1 < \cdots < n_{N-1}$, as well as a sign $\varepsilon_M \in \{-1, 0, 1\}$ for each integer $M > n_{N-1}$. Given reals $\rho, c_{n_0}, c_{n_1}, \dots, c_{n_{N-1}} > 0$, there exists a power series*

$$f(x) = c_{n_0} x^{n_0} + \cdots + c_{n_{N-1}} x^{n_{N-1}} + \sum_{M > n_{N-1}} c_M x^M,$$

satisfying the following properties:

(1) f is convergent on $(0,\rho)$.
(2) $f[-] : \mathbb{P}_N((0,\rho)) \to \mathbb{P}_N$.
(3) $\operatorname{sgn}(c_M) = \varepsilon_M$ for each $M > n_{N-1}$.

This is slightly stronger than classifying the sign patterns, in that the "initial coefficients" are also specified. In fact, this result can be strengthened in two different ways, see (1) Theorem 41.7, in which the set of powers allowed is vastly more general; and (2) Theorem 44.14 and the discussion preceding it, in which the coefficients for $M > n_{N-1}$ are also specified.

Proof. Suppose we can prove the theorem in the special case when exactly one ε_M is negative. Then for each $M > n_{N-1}$, there exists $0 < \delta_M < \frac{1}{M!}$, such that

$$f_M(x) := \sum_{j=0}^{N_1} c_{n_j} x^{n_j} + c_M x^M$$

preserves positivity on $\mathbb{P}_N((0,\rho))$ whenever $|c_M| \leq \delta_M$. Set $c_M := \varepsilon_M \delta_M$ for each $M > n_{N-1}$ and define $f(x) := \sum_{M > n_{N-1}} 2^{n_{N-1} - M} f_M(x)$. If $x \in (0,\rho)$, then we have

$$\begin{aligned} |f(x)| &\leq \sum_{M > n_{N-1}} 2^{n_{N-1}-M} |f_M(x)| \leq \sum_{M > n_{N-1}} 2^{n_{N-1}-M} \left(\sum_{j=0}^{N-1} c_{n_j} x^{n_j} + \delta_M x^M \right) \\ &\leq \sum_{j=0}^{N-1} c_{n_j} x^{n_j} + e^x < \infty. \end{aligned}$$

Hence, f converges on $(0,\rho)$. As each $f_M[-]$ preserves positivity and $\mathbb{P}_N$ is a closed convex cone, $f[-]$ also preserves positivity. It therefore remains to show that the result holds when one coefficient is negative. But this follows from Theorem 41.5. $\square$

Thus, it remains to show the following "qualitative" result:

Theorem 41.5. *Let $N > 0$, $0 \leq n_0 < n_1 < \ldots < n_{N-1} < M$ be integers, and $\rho, c_{n_0}, c_{n_1}, \ldots, c_{n_{N-1}} > 0$ be real. Then the function $f(x) = \sum_{j=0}^{N-1} c_{n_j} x^{n_j} + c_M x^M$ entrywise preserves positivity on $\mathbb{P}_N((0,\rho))$, for some $c_M < 0$.*

We will show this result in the next two sections (Sections 42 and 43).

41.3. **Classification of sign patterns for sums of real powers.** After proving Theorem 41.5, we further strengthen it by proving a *quantitative* version – see Theorem 44.1 – which gives a sharp lower bound on c_M. For now, we list a special case of that result (without proof, as we show the more general Theorem 44.1). In the following result and beyond, the set $\mathbb{Z}^{\geq 0} \cup [N-2,\infty)$ comes from Theorem 9.3.

Theorem 41.6. *Theorem 41.5 holds even when the exponents $n_0, n_1, \ldots, n_{N-1}, M$ are real and lie in the set $\mathbb{Z}^{\geq 0} \cup [N-2,\infty)$.*

With Theorem 41.6 in hand, it is possible to classify the sign patterns of a more general family of preservers, of the form $f(x) = \sum_{j=0}^{\infty} c_{n_j} x^{n_j}$, where $n_j \in \mathbb{Z}^{\geq 0} \cup [N-2,\infty)$ are an arbitrary countable collection of pairwise distinct non-negative (real) exponents.

Theorem 41.7 (Classification of sign patterns of power series preservers, fixed dimension)**.** *Let $N \geq 2$ and let $n_0, n_1, \ldots$ be a sequence of pairwise distinct real numbers in $\mathbb{Z}^{\geq 0} \cup [N-2,\infty)$. For each $j \geq 0$, let $\varepsilon_j \in \{-1,0,1\}$ be a sign, such that whenever $\varepsilon_{j_0} = -1$, one has $\varepsilon_j = +1$ for at least N choices of j satisfying: $n_j < n_{j_0}$. Then for every $\rho > 0$, there exists a series with real exponents and real coefficients*

$$f(x) = \sum_{j=0}^{\infty} c_{n_j} x^{n_j}$$

which is convergent on $(0,\rho)$, which entrywise preserves positivity on $\mathbb{P}_N((0,\rho))$, and in which $\operatorname{sgn}(c_{n_j}) = \varepsilon_j$ *for all $j \geq 0$.*

That the sign patterns must satisfy the given hypotheses follows from Lemma 41.2. In particular, Theorem 41.7 shows that the Horn–Loewner-type necessary condition in Lemma 41.2 remains the best possible in this generality as well.

Remark 41.8. A key difference between the classifications in Theorems 41.4 and 41.7 is that the latter is more flexible, since the sequence $n_0, n_1, \ldots$ can now contain an infinite decreasing subsequence of exponents. This is more general than even the Hahn or Puiseux series, not just power series. For instance, the sum may be over all rational powers in $\mathbb{Z}^{\geq 0} \cup [N-2,\infty)$.

Proof of Theorem 41.7. Given any set $\{n_j : j \geqslant 0\}$ of (pairwise distinct) non-negative powers,

$$\sum_{j \geqslant 0} \frac{x^{n_j}}{j! \lceil n_j \rceil !} < \infty, \qquad \forall x > 0. \tag{41.9}$$

Indeed, if we partition $\mathbb{Z}^{\geqslant 0}$ into the disjoint union of $J_k := \{j \geqslant 0 : n_j \in (k-1,k]\}$, $k \geqslant 0$, then using Tonelli's theorem, we can estimate

$$\sum_{j \geqslant 0} \frac{x^{n_j}}{j! \lceil n_j \rceil !} = \sum_{k \geqslant 0} \frac{1}{k!} \sum_{j \in J_k} \frac{x^{n_j}}{j!} \leqslant e + \sum_{k \geqslant 1} \frac{1}{k!} \sum_{j \in J_k} \frac{x^k + x^{k-1}}{j!} < e + e(e^x + x^{-1} e^x) < \infty.$$

We now turn to the proof. Set $J := \{j : \varepsilon_j = -1\} \subset \mathbb{Z}^{\geq 0}$. By the hypotheses, for each $l \in J$ there exist $j_1(l), \dots, j_N(l)$, such that $\varepsilon_{j_k(l)} = 1$ and $n_{j_k(l)} < n_l$, for $k = 1, \dots, N$. Define

$$f_l(x) := \sum_{k=1}^{N} \frac{x^{n_{j_k(l)}}}{\lceil n_{j_k(l)} \rceil!} - \delta_l \frac{x^{n_l}}{\lceil n_l \rceil!},$$

where $\delta_l \in (0,1)$ is chosen, such that $f_l[-]$ preserves positivity on $\mathbb{P}_N((0,\rho))$ by Theorem 41.6. Let $J' \subset \mathbb{Z}^{\geq 0}$ consist of all $j \geqslant 0$, such that $\varepsilon_j = +1$ but $j \neq j_k(l)$ for any $l \in J, k \in [1,N]$. Finally, define

$$f(x) := \sum_{l \in J} \frac{f_l(x)}{l!} + \sum_{j \in J'} \frac{x^{n_j}}{j! \lceil n_j \rceil!}, \qquad x > 0.$$

Repeating the calculation in (41.9), one can verify that f converges absolutely on $(0,\infty)$ and hence on $(0,\rho)$. By the above hypotheses and the Schur product theorem, it follows that $f[-]$ preserves positivity on $\mathbb{P}_N((0,\rho))$. $\square$

42. Polynomial preservers for generic rank-one matrices. Schur polynomials.

The goal in this section and the next (Section 43) is to prove the "qualitative" Theorem 41.5 from the previous section (Section 41). Thus, we work with polynomials of the form

$$f(x) = \sum_{j=0}^{N-1} c_{n_j} x^{n_j} + c_M x^M,$$

where $N > 0$, $0 \leq n_0 < n_1 < \ldots < n_{N-1} < M$ are integers, and $\rho, c_{n_0}, c_{n_1}, \ldots, c_{n_{N-1}} > 0$ are real.

42.1. Basic properties of Schur polynomials.

In this section, we begin by defining the key tool required here and beyond: Schur polynomials. We then use these functions – via the Cauchy–Binet formula – to understand when polynomials of the above form entrywise preserve positivity on a generic rank-1 matrix in $\mathbb{P}_N((0,\rho))$.

Definition 42.1. Fix integers $m, N > 0$, and define $\mathbf{n}_{\min} := (0, 1, \ldots, N-1)$. Now suppose $0 \leq n'_0 \leq n'_1 \leq \cdots \leq n'_{N-1}$ are also integers.

(1) A *column-strict Young tableau*, with shape $\mathbf{n}' := (n'_0, n'_1, \ldots, n'_{N-1})$ and cell entries $1, 2, \ldots, m$, is a left-aligned two-dimensional rectangular array T of cells, with n'_0 cells in the bottom row, n'_1 cells in the second lowest row, and so on, such that:
- Each cell in T has integer entry j with $1 \leq j \leq m$.
- Entries weakly decrease in each row, from left to right.
- Entries strictly decrease in each column, from top to bottom.

(2) Given variables $u_1, u_2, \ldots, u_m$ and a column-strict Young tableau T as above, define its *weight* to be

$$\mathrm{wt}(T) := \prod_{j=1}^{m} u_j^{f_j},$$

where f_j equals the number of cells in T with entry j.

(3) Given an increasing sequence of integers $0 \leq n_0 < \cdots < n_{N-1}$, define the tuple $\mathbf{n} := (n_0, n_1, \ldots, n_{N-1})$, and the corresponding *Schur polynomial* over $\mathbf{u} := (u_1, u_2, \ldots, u_m)^T$ to be

$$s_{\mathbf{n}}(\mathbf{u}) := \sum_T \mathrm{wt}(T), \tag{42.2}$$

where T runs over all column-strict Young tableaux of shape $\mathbf{n}' := \mathbf{n} - \mathbf{n}_{\min}$ with cell entries $1, 2, \ldots, m$. (We will also abuse notation slightly and write $s_{\mathbf{n}}(\mathbf{u}) = s_{\mathbf{n}}(u_1, \ldots, u_m)$ on occasion.)

Example 42.3. Suppose $N = m = 3$ and $\mathbf{n} = (0, 2, 4)$. The column-strict Young tableaux with shape $\mathbf{n} - \mathbf{n}_{\min} = (0, 1, 2)$ and cell entries $(1, 2, 3)$ are

$$\begin{array}{|c|c|}\hline 3 & 3 \\ \hline 2 \\ \cline{1-1}\end{array} \quad \begin{array}{|c|c|}\hline 3 & 3 \\ \hline 1 \\ \cline{1-1}\end{array} \quad \begin{array}{|c|c|}\hline 3 & 2 \\ \hline 2 \\ \cline{1-1}\end{array} \quad \begin{array}{|c|c|}\hline 3 & 2 \\ \hline 1 \\ \cline{1-1}\end{array} \quad \begin{array}{|c|c|}\hline 3 & 1 \\ \hline 2 \\ \cline{1-1}\end{array} \quad \begin{array}{|c|c|}\hline 3 & 1 \\ \hline 1 \\ \cline{1-1}\end{array} \quad \begin{array}{|c|c|}\hline 2 & 2 \\ \hline 1 \\ \cline{1-1}\end{array} \quad \begin{array}{|c|c|}\hline 2 & 1 \\ \hline 1 \\ \cline{1-1}\end{array}$$

As a consequence,

$$\begin{aligned} s_{(0,2,4)}(u_1, u_2, u_3) &= u_3^2 u_2 + u_3^2 u_1 + u_3 u_2^2 + 2u_3 u_2 u_1 + u_3 u_1^2 + u_2^2 u_1 + u_2 u_1^2 \\ &= (u_1 + u_2)(u_2 + u_3)(u_3 + u_1). \end{aligned}$$

Remark 42.4. A visible notational distinction with the literature is that column-strict Young tableaux traditionally have entries that are *increasing* down columns, and weakly increasing as one moves across rows. Since we only work with sets of tableaux through the sums of their weights occurring in Schur polynomials, this distinction is unimportant in the text, for the following reason: define an involutive bijection $\iota : j \mapsto m+1-j$, where $\{1,\dots,m\}$ is the alphabet of possible cell entries. Then the column-strict Young tableaux in our notation bijectively correspond under ι – applied to each cell entry – to the "usual" column-strict Young tableaux (in the literature); and as Schur polynomials are symmetric under permuting the variables by ι (see Proposition 42.6), the sums of weights of the two sets of tableaux coincide.

Remark 42.5. Schur polynomials are fundamental objects in type A representation theory (of the general linear group, or the special linear Lie algebra), and are characters of irreducible finite-dimensional polynomial representations (over fields of characteristic zero). The above example 42.3 is a special case, corresponding to the adjoint representation for the Lie algebra of 3×3 traceless matrices. This interpretation will not be used in this text.

Schur polynomials are always homogeneous – and also symmetric, because they can be written as a quotient of two generalized Vandermonde determinants. This is Cauchy's definition; the definition (42.2) using Young tableaux is by Littlewood. One can show that these two definitions are equivalent, among other basic properties:

Proposition 42.6. *Fix integers $m = N > 0$ and $0 \leq n_0 < n_1 < \cdots < n_{N-1}$.*

(1) (Cauchy's definition.) If $\mathbb{F}$ is a field and $\mathbf{u} = (u_1,\dots,u_N)^T \in \mathbb{F}^N$, then

$$\det(\mathbf{u}^{\circ n_0} \mid \mathbf{u}^{\circ n_1} \mid \dots \mid \mathbf{u}^{\circ n_{N-1}})_{N\times N} = V(\mathbf{u}) s_{\mathbf{n}}(\mathbf{u}),$$

where for a (column) vector or (row) tuple $\mathbf{u}$, we denote by $V(\mathbf{u}) := \prod_{1\leq j<k\leq N}(u_k - u_j)$ the Vandermonde determinant as in (17.4). In particular, $s_{\mathbf{n}}(\mathbf{u})$ is symmetric and homogeneous of degree $\sum_{j=0}^{N-1}(n_j - j)$.

(2) (Principal specialization formula.) For any $q \in \mathbb{F}$ that is not a root of unity or else has order $\geq N$, we have

$$s_{\mathbf{n}}(1,q,q^2,\dots,q^{N-1}) = \prod_{0\leq j<k\leq N-1} \frac{q^{n_k} - q^{n_j}}{q^k - q^j}.$$

(3) (Weyl dimension formula.) Specialized to $q = 1$, we have

$$s_{\mathbf{n}}(1,1,\dots,1) = \frac{V(\mathbf{n})}{V(\mathbf{n}_{\min})} \in \mathbb{N}.$$

In particular, there are $V(\mathbf{n})/V(\mathbf{n}_{\min})$ column-strict tableaux of shape $\mathbf{n} - \mathbf{n}_{\min}$ and cell entries $1,\dots,N$. Here and below, we will mildly abuse notation and write $V(\mathbf{n})$ for a tuple/row vector $\mathbf{n}$ to denote $V(\mathbf{n}^T)$.

Proof. The first part is proved in Theorem 46.1 below. Using this, we show the second part. Set $\mathbf{u} := (1,q,q^2,\dots,q^{N-1})^T$ with q as given. Then it is easy to verify that

$$s_{\mathbf{n}}(\mathbf{u}) = \frac{\det(\mathbf{u}^{\circ n_0} \mid \mathbf{u}^{\circ n_1} \mid \dots \mid \mathbf{u}^{\circ n_{N-1}})}{V(\mathbf{u})} = \frac{V((q^{n_0},\dots,q^{n_{N-1}}))}{V((q^0,\dots,q^{N-1}))} = \prod_{0\leq j<k\leq N-1} \frac{q^{n_k} - q^{n_j}}{q^k - q^j},$$

as desired.

Finally, to prove the Weyl dimension formula, notice that by the first part, the Schur polynomial has integer coefficients and hence makes sense over $\mathbb{Z}$, and then specializes to $s_{\mathbf{n}}(\mathbf{u})$ over any ground field. Now work over the ground field $\mathbb{Q}$, and let $f_{\mathbf{n}}(T) := s_{\mathbf{n}}(1, T, \dots, T^{N-1}) \in \mathbb{Z}[T]$ be the corresponding "principally specialized" polynomial. Then,

$$V((q^0, \dots, q^{N-1})) f_{\mathbf{n}}(q) = V((q^{n_0}, \dots, q^{n_{N-1}})), \qquad \forall q \in \mathbb{Q}.$$

In particular, for every $q \neq 1$, dividing both sides by $(q-1)^{\binom{N}{2}}$, we obtain:

$$\prod_{0 \leq j < k \leq N-1} (q^{n_j} + q^{n_j+1} + \cdots + q^{n_k - 1}) - f_{\mathbf{n}}(q) \prod_{0 \leq j < k \leq N-1} (q^j + q^{j+1} + \cdots + q^{k-1}) = 0,$$

for all $q \in \mathbb{Q} \setminus \{1\}$. This means that the left-hand side is (the specialization of) a polynomial with infinitely many roots, hence the polynomial vanishes identically on $\mathbb{Q}$. Specializing this polynomial to $q = 1$ now yields the Weyl dimension formula

$$\frac{V(\mathbf{n})}{V(\mathbf{n}_{\min})} = \prod_{0 \leq j < k \leq N-1} \frac{n_k - n_j}{k - j} = f_{\mathbf{n}}(1) = s_{\mathbf{n}}(1, \dots, 1).$$

The final assertion now follows from Littlewood's definition (42.2) of $s_{\mathbf{n}}(\mathbf{u})$. $\square$

42.2. **Polynomials preserving positivity on individual rank-one positive matrices.** We return to proving Theorem 41.5, and hence Theorem 41.4 on sign patterns. As we have shown, it suffices to prove the theorem for one higher degree (leading) term with a negative coefficient. Before proving the result in full, we tackle the following (simpler) versions. Thus, we are given a real polynomial as above: $f(x) = \sum_{j=0}^{N-1} c_{n_j} x^{n_j} + c_M x^M$, where $c_{n_j} > 0 \; \forall j$.

(1) Does there exist $c_M < 0$, such that $f[-] : \mathbb{P}_N((0, \rho)) \to \mathbb{P}_N$?
Here is a reformulation: dividing the expression for $f(x)$ throughout by $|c_M| = 1/t > 0$, define

$$p_t(x) := t \sum_{j=0}^{N-1} c_{n_j} x^{n_j} - x^m, \quad \text{where } c_{n_j} > 0 \; \forall j. \tag{42.7}$$

Then it is enough to ask for which $t > 0$ (if any) does $p_t[-] : \mathbb{P}_N((0, \rho)) \to \mathbb{P}_N$?

(2) Here are two simplifications: Can we produce such a constant $t > 0$ for only the subset of rank-1 matrices in $\mathbb{P}_N((0, \rho))$? How about for a *single* rank-1 matrix $\mathbf{u}\mathbf{u}^T$?

(3) A further special case: let $\mathbf{u}$ be generic, in that $\mathbf{u} \in (0, \rho)^N$ has distinct coordinates, and p_t is as above. Can one now compute all $t > 0$, such that $p_t[\mathbf{u}\mathbf{u}^T] \in \mathbb{P}_N$? How about all $t > 0$, such that $\det p_t[\mathbf{u}\mathbf{u}^T] \geq 0$?

We begin by answering the last of these questions – the answer crucially uses Schur polynomials. The following result shows that, in fact, $\det p_t[\mathbf{u}\mathbf{u}^T] \geq 0$ implies $p_t[\mathbf{u}\mathbf{u}^T]$ is positive semidefinite!

Proposition 42.8. *With $N \geq 1$ and notation as in* (42.7)*, define the vectors*

$$\mathbf{n} := (n_0, \dots, n_{N-1}), \qquad \mathbf{n}_j := (n_0, \dots, n_{j-1}, \widehat{n_j}, n_{j+1}, \dots, n_{N-1}, M), \quad 0 \leq j < N, \tag{42.9}$$

where $0 \leq n_0 < \cdots < n_{N-1} < M$. Now if the n_j and M are integers, and a vector $\mathbf{u} \in (0, \infty)^N$ has pairwise distinct coordinates, then the following are equivalent:

(1) $p_t[\mathbf{u}\mathbf{u}^T]$ is positive semidefinite.
(2) $\det p_t[\mathbf{u}\mathbf{u}^T] \geq 0$.
(3) $t \geq \displaystyle\sum_{j=0}^{N-1} \frac{s_{\mathbf{n}_j}(\mathbf{u})^2}{c_{n_j} s_{\mathbf{n}}(\mathbf{u})^2}$.

In particular, at least for "most" rank-1 matrices, it is possible to find polynomial preservers of positivity (on that one matrix), with a negative coefficient.

The proof of Proposition 42.8 uses the following even more widely applicable equivalence between the non-negativity of the determinant and of the entire spectrum for "special" linear pencils of matrices:

Lemma 42.10. *Fix* $\mathbf{w} \in \mathbb{R}^N$ *and a positive definite matrix* H. *Define the linear pencil* $P_t := tH - \mathbf{w}\mathbf{w}^T$, *for* $t > 0$. *Then the following are equivalent:*

(1) P_t *is positive semidefinite.*
(2) $\det P_t \geq 0$.
(3) $t \geq \mathbf{w}^T H^{-1} \mathbf{w} = 1 - \dfrac{\det(H - \mathbf{w}\mathbf{w}^T)}{\det H}$.

This lemma is naturally connected to the theory of *(generalized) Rayleigh quotients*, although we do not pursue this further.

Proof. We show a cyclic chain of implications. That (1) $\implies$ (2) is immediate.

(2) $\implies$ (3) : Using the identity (2.33) from Section 2.4 on Schur complements, we obtain by taking determinants

$$\det \begin{pmatrix} A & B \\ B' & D \end{pmatrix} = \det D \cdot \det(A - BD^{-1}B')$$

whenever A, D are square matrices, with D invertible. Using this, we compute:

$$0 \leq \det(tH - \mathbf{w}\mathbf{w}^T) = \det \begin{pmatrix} tH & \mathbf{w} \\ \mathbf{w}^T & 1 \end{pmatrix} = \det \begin{pmatrix} 1 & \mathbf{w}^T \\ \mathbf{w} & tH \end{pmatrix} = \det(tH) \det(1 - \mathbf{w}^T (tH)^{-1} \mathbf{w}).$$

Since the last quantity is a scalar, and $\det(tH) > 0$ by assumption, it follows from (2) that

$$1 \geq t^{-1}(\mathbf{w}^T H^{-1} \mathbf{w}) \quad \implies \quad t \geq \mathbf{w}^T H^{-1} \mathbf{w}.$$

Now substitute $t = 1$ in the above computation, to obtain

$$\begin{aligned} \det(H - \mathbf{w}\mathbf{w}^T) &= \det(H) \det(1 - \mathbf{w}^T H^{-1} \mathbf{w}) \\ \implies \frac{\det(H - \mathbf{w}\mathbf{w}^T)}{\det H} &= 1 - \mathbf{w}^T H^{-1} \mathbf{w} \geq 1 - t, \end{aligned}$$

which implies (3).

(3) $\implies$ (1) : It suffices to show that $\mathbf{x}^T P_t \mathbf{x} \geq 0$ for all non-zero vectors $\mathbf{x} \in \mathbb{R}^N$. Using a change of variables $\mathbf{y} = \sqrt{H}\mathbf{x} \neq \mathbf{0}$, we compute:

$$\begin{aligned} \mathbf{x}^T P_t \mathbf{x} &= t\mathbf{y}^T\mathbf{y} - (\mathbf{y}^T \sqrt{H}^{-1} \mathbf{w})^2 \\ &= \|\mathbf{y}\|^2 (t - ((\mathbf{y}')^T \sqrt{H}^{-1} \mathbf{w})^2), \qquad \text{where } \mathbf{y}' := \frac{\mathbf{y}}{\|\mathbf{y}\|} \\ &\geq \|\mathbf{y}\|^2 (t - \|\mathbf{y}'\|^2 \left\| \sqrt{H^{-1}} \mathbf{w} \right\|^2) \qquad \text{(using Cauchy–Schwarz)} \\ &= \|\mathbf{y}\|^2 (t - \mathbf{w}^T H^{-1} \mathbf{w}) \geq 0 \qquad \text{(by assumption).} \end{aligned}$$

□

We can now answer the last of the above questions on positivity preservers, for generic rank-1 matrices.

Proof of Proposition 42.8. The result is easily shown for $N = 1$, so we now assume $N \geq 2$. We are interested in the following matrix and its determinant:

$$p_t[\mathbf{u}\mathbf{u}^T] = t \sum_{j=0}^{N-1} c_{n_j} (\mathbf{u}^{\circ n_j})(\mathbf{u}^{\circ n_j})^T - (\mathbf{u}^{\circ M})(\mathbf{u}^{\circ M})^T.$$

We first work more generally: over any field $\mathbb{F}$, and with matrices $\mathbf{u}\mathbf{v}^T$, where $\mathbf{u}, \mathbf{v} \in \mathbb{F}^N$. Thus, we study

$$p_t[\mathbf{u}\mathbf{v}^T] = t \sum_{j=0}^{N-1} c_{n_j} \mathbf{u}^{\circ n_j} (\mathbf{v}^{\circ n_j})^T - \mathbf{u}^{\circ M} (\mathbf{v}^{\circ M})^T,$$

where $t, c_{n_j} \in \mathbb{F}$, and $c_{n_j} \neq 0 \ \forall j$. Setting $D = \operatorname{diag}(tc_{n_0}, \dots, tc_{n_{N-1}}, -1)$, we have the decomposition

$$p_t[\mathbf{u}\mathbf{v}^T] = U(\mathbf{u}) D U(\mathbf{v})^T, \quad \text{where } U(\mathbf{u})_{N \times (N+1)} := (\mathbf{u}^{\circ n_0} \mid \dots \mid \mathbf{u}^{\circ n_{N-1}} \mid \mathbf{u}^{\circ M}).$$

Applying the Cauchy–Binet formula to $A = U(\mathbf{u}), B = DU(\mathbf{v})^T$, as well as Cauchy's definition in Proposition 42.6(1), we obtain the following general determinantal identity, valid over any field:

$$\det p_t[\mathbf{u}\mathbf{v}^T] = V(\mathbf{u})V(\mathbf{v}) t^{N-1} \prod_{j=0}^{N-1} c_{n_j} \cdot \left(s_{\mathbf{n}}(\mathbf{u}) s_{\mathbf{n}}(\mathbf{v}) t - \sum_{j=0}^{N-1} \frac{s_{\mathbf{n}_j}(\mathbf{u}) s_{\mathbf{n}_j}(\mathbf{v})}{c_{n_j}} \right). \tag{42.11}$$

Now specialize this identity to $\mathbb{F} = \mathbb{R}$, with $t, c_{n_j} > 0$ and $\mathbf{u} = \mathbf{v} \in (0, \infty)^N$ having distinct coordinates. From this we deduce the following consequences. First, set

$$H := \sum_{j=0}^{N-1} c_{n_j} (\mathbf{u}\mathbf{u}^T)^{\circ n_j} = U'(\mathbf{u}) D' U'(\mathbf{u})^T, \qquad \mathbf{w} := \mathbf{u}^{\circ M},$$

where $D' := \operatorname{diag}(c_{n_0}, \dots, c_{n_{N-1}})$ is a positive definite matrix and $U'(\mathbf{u}) := (\mathbf{u}^{\circ n_0} \mid \dots \mid \mathbf{u}^{\circ n_{N-1}})$ is a generalized Vandermonde matrix which has determinant $V(\mathbf{u}) s_{\mathbf{n}}(\mathbf{u}) \neq 0$. From this it follows that H is positive definite, so Lemma 42.10 applies. Moreover, $H - \mathbf{w}\mathbf{w}^T = p_1[\mathbf{u}\mathbf{u}^T]$, so using the above calculation (42.11) and the Cauchy–Binet formula respectively, we have

$$\det(H - \mathbf{w}\mathbf{w}^T) = V(\mathbf{u})^2 \prod_{j=0}^{N-1} c_{n_j} \cdot s_{\mathbf{n}}(\mathbf{u})^2 \left(1 - \sum_{j=0}^{N-1} \frac{s_{\mathbf{n}_j}(\mathbf{u})^2}{c_{n_j} s_{\mathbf{n}}(\mathbf{u})^2} \right),$$

$$\det H = V(\mathbf{u})^2 \prod_{j=0}^{N-1} c_{n_j} \cdot s_{\mathbf{n}}(\mathbf{u})^2.$$

In particular, from Lemma 42.10(3) we obtain

$$\mathbf{w}^T H^{-1} \mathbf{w} = \sum_{j=0}^{N-1} \frac{s_{\mathbf{n}_j}(\mathbf{u})^2}{c_{n_j} s_{\mathbf{n}}(\mathbf{u})^2}.$$

Now the proposition follows directly from Lemma 42.10, since $P_t = p_t[\mathbf{u}\mathbf{u}^T]$ for all $t > 0$. $\square$

43. First-order approximation / leading term of Schur polynomials. From rank-one matrices to all matrices.

In the previous section, we computed the exact threshold for the leading term of a polynomial

$$p_t(x) := t \sum_{j=0}^{N-1} c_{n_j} x^{n_j} - x^M, \quad \text{where } c_{n_j} > 0 \ \forall j$$

(and where $0 \leq n_0 < \cdots < n_{N-1} < M$ are integers), such that $p_t[\mathbf{u}\mathbf{u}^T] \in \mathbb{P}_N$ for a single vector $\mathbf{u} \in (0,\infty)^N$ with pairwise distinct coordinates. Recall that our (partial) goal is to find a threshold that works for all rank-1 matrices $\mathbf{u}\mathbf{u}^T \in \mathbb{P}_N((0,\rho))$ – i.e., for $\mathbf{u} \in (0,\sqrt{\rho})^N$. Thus, we need to show that the supremum of the threshold over all such $\mathbf{u}$ is bounded:

$$\sup_{\mathbf{u} \in (0,\sqrt{\rho})^N} \sum_{j=0}^{N-1} \frac{s_{\mathbf{n}_j}(\mathbf{u})^2}{c_{n_j} s_{\mathbf{n}}(\mathbf{u})^2} < \infty.$$

Since we only consider vectors $\mathbf{u}$ with positive coordinates, it suffices to bound $s_{\mathbf{n}_j}(\mathbf{u})/s_{\mathbf{n}}(\mathbf{u})$ from above, for each j. In turn, for this it suffices to find lower and upper bounds for every Schur polynomial evaluated at $\mathbf{u} \in (0,\infty)^N$. This is achieved by the following result:

Theorem 43.1 (First-order approximation/Leading term of Schur polynomials)**.** *Say $N \geq 1$ and $0 \leq n_0 < \cdots < n_{N-1}$ are integers. Then for all real numbers $0 < u_1 \leq u_2 \leq \cdots \leq u_N$, we have the bounds*

$$1 \times \mathbf{u}^{\mathbf{n}-\mathbf{n}_{\min}} \leq s_{\mathbf{n}}(\mathbf{u}) \leq \frac{V(\mathbf{n})}{V(\mathbf{n}_{\min})} \times \mathbf{u}^{\mathbf{n}-\mathbf{n}_{\min}},$$

where $\mathbf{u}^{\mathbf{n}-\mathbf{n}_{\min}} = u_1^{n_0-0} u_2^{n_1-1} \cdots u_N^{n_{N-1}-(N-1)}$ and $V(\mathbf{n})$ is as in (17.4). Moreover, the constants 1 and $\frac{V(\mathbf{n})}{V(\mathbf{n}_{\min})}$ cannot be improved.

Proof. Recall that $s_{\mathbf{n}}(\mathbf{u})$ is obtained by summing the weights of all column-strict Young tableaux of shape $\mathbf{n} - \mathbf{n}_{\min}$ with cell entries $1,\dots,N$. Moreover, by the Weyl dimension formula in Proposition 42.6(3), there are precisely $V(\mathbf{n})/V(\mathbf{n}_{\min})$ such tableaux. Now each such tableau can have weight at most $\mathbf{u}^{\mathbf{n}-\mathbf{n}_{\min}}$, as follows: the cells in the top row each have entries at most N; the cells in the next row at most $N-1$; and so on. The tableau $T_{\max}$ obtained in this fashion has weight precisely $\mathbf{u}^{\mathbf{n}-\mathbf{n}_{\min}}$. Hence, by definition, we have

$$\mathbf{u}^{\mathbf{n}-\mathbf{n}_{\min}} = \mathrm{wt}(T_{\max}) \leq \sum_T \mathrm{wt}(T) = s_{\mathbf{n}}(\mathbf{u}) \leq \sum_T \mathrm{wt}(T_{\max}) = \frac{V(\mathbf{n})}{V(\mathbf{n}_{\min})} \mathbf{u}^{\mathbf{n}-\mathbf{n}_{\min}}.$$

This proves the bounds; we claim that both bounds are sharp. If $\mathbf{n} = \mathbf{n}_{\min}$ then all terms in the claimed inequalities are 1, and we are done. Thus, assume henceforth that $\mathbf{n} \neq \mathbf{n}_{\min}$. Let $A > 1$ and define $\mathbf{u}(A) := (A, A^2, \dots, A^N)$. Then $\mathrm{wt}(T_{\max}) = A^M$ for some $M > 0$. Hence, for every column-strict Young tableau $T \neq T_{\max}$ as above, $\mathrm{wt}(T)$ is at most $\mathrm{wt}(T_{\max})/A$ and at least $1 = \mathrm{wt}(T_{\max})/A^M$. Now summing over all such tableaux T yields

$$s_{\mathbf{n}}(\mathbf{u}(A)) \leq \mathbf{u}(A)^{\mathbf{n}-\mathbf{n}_{\min}} \left(1 + \left(\frac{V(\mathbf{n})}{V(\mathbf{n}_{\min})} - 1\right)\frac{1}{A}\right),$$
$$s_{\mathbf{n}}(\mathbf{u}(A)) \geq \mathbf{u}(A)^{\mathbf{n}-\mathbf{n}_{\min}} \left(1 + \left(\frac{V(\mathbf{n})}{V(\mathbf{n}_{\min})} - 1\right)\frac{1}{A^M}\right).$$

Divide throughout by $\mathbf{u}(A)^{\mathbf{n}-\mathbf{n}_{\min}}$; now taking the limit as $A \to \infty$ yields the sharp lower bound 1 while taking the limit as $A \to 1^+$ yields the sharp upper bound $V(\mathbf{n})/V(\mathbf{n}_{\min})$. $\square$

We now use Theorem 43.1, and Proposition 42.8 in the previous section, to find a threshold for $t > 0$ beyond which $p_t[-]$ preserves positivity on all rank-1 matrices in $\mathbb{P}_N((0,\rho))$ – and, in fact, on *all* matrices in $\mathbb{P}_N((0,\rho))$.

Theorem 43.2. *Fix integers $N \geq 1$, $0 \leq n_0 < n_1 < \cdots < n_{N-1} < M$, and scalars $\rho, t, c_{n_0}, \dots, c_{n_{N-1}} > 0$. The polynomial $p_t(x) := t\sum_{j=0}^{N-1} c_{n_j} x^{n_j} - x^M$ entrywise preserves positivity on $\mathbb{P}_N((0,\rho))$, if $t \geq t_0 := \displaystyle\sum_{j=0}^{N-1} \frac{V(\mathbf{n}_j)^2}{c_{n_j} V(\mathbf{n}_{\min})^2} \rho^{M-n_j}$.*

The notation in Definition 25.1 is useful here and in the sequel. Specifically, $X^{N,\neq}$ for a set X and an integer $N \geq 1$ denotes the set of N-tuples from X with pairwise distinct entries.

Proof of Theorem 43.2. Given $\mathbf{u} \in (0,\sqrt{\rho})^{N,\neq}$, from Proposition 42.8 it follows that $p_t[\mathbf{u}\mathbf{u}^T] \in \mathbb{P}_N$ if and only if $t \geq \sum_{j=0}^{N-1} \frac{s_{\mathbf{n}_j}(\mathbf{u})^2}{c_{n_j} s_{\mathbf{n}}(\mathbf{u})^2}$. Now suppose $\mathbf{u} \in (0,\sqrt{\rho})^{N,\uparrow}$. Then by Theorem 43.1,

$$\sum_{j=0}^{N-1} \frac{s_{\mathbf{n}_j}(\mathbf{u})^2}{c_{n_j} s_{\mathbf{n}}(\mathbf{u})^2} \leq \sum_{j=0}^{N-1} \frac{\mathbf{u}^{2(\mathbf{n}_j - \mathbf{n}_{\min})} V(\mathbf{n}_j)^2 / V(\mathbf{n}_{\min})^2}{c_{n_j} \mathbf{u}^{2(\mathbf{n} - \mathbf{n}_{\min})}} = \sum_{j=0}^{N-1} \frac{V(\mathbf{n}_j)^2}{c_{n_j} V(\mathbf{n}_{\min})^2} \mathbf{u}^{2(\mathbf{n}_j - \mathbf{n})},$$

and this is bounded above by t_0, since if $\mathbf{v} := \sqrt{\rho}(1,\dots,1)^T$ then $\mathbf{u}^{2(\mathbf{n}_j-\mathbf{n})} \leq \mathbf{v}^{2(\mathbf{n}_j-\mathbf{n})} = \rho^{M-n_j}$ for all j. Thus, we conclude that

$$\begin{aligned} t \geq t_0 \quad \implies \quad p_t[\mathbf{u}\mathbf{u}^T] \in \mathbb{P}_N \ \forall \mathbf{u} \in (0,\sqrt{\rho})^{N,\uparrow} \quad &\implies \quad p_t[\mathbf{u}\mathbf{u}^T] \in \mathbb{P}_N \ \forall \mathbf{u} \in (0,\sqrt{\rho})^{N,\neq} \\ &\implies \quad p_t[\mathbf{u}\mathbf{u}^T] \in \mathbb{P}_N \ \forall \mathbf{u} \in (0,\sqrt{\rho})^{N}, \end{aligned}$$

where the first implication was proved above, the second follows by (the symmetric nature of Schur polynomials and by) relabeling the rows and columns of $\mathbf{u}\mathbf{u}^T$ to rearrange the entries of $\mathbf{u}$ in increasing order, and the third implication follows from the continuity of p_t and the density of $(0,\sqrt{\rho})^{N,\neq}$ in $(0,\sqrt{\rho})^N$.

This validates the claimed threshold t_0 for all rank-1 matrices. To prove the result on all of $\mathbb{P}_N((0,\rho))$, we use induction on $N \geq 1$, with the base case of $N = 1$ already done since 1×1 matrices have rank 1.

For the induction step, recall the extension principle (Theorem 9.12), which said that: *Suppose $I = (0,\rho)$ or $(-\rho,\rho)$ or its closure, for some $0 < \rho \leq \infty$. If $h \in C^1(I)$ is such that $h[-]$ preserves positivity on rank-1 matrices in $\mathbb{P}_N(I)$ and $h'[-] : \mathbb{P}_{N-1}(I) \to \mathbb{P}_{N-1}$, then $h[-] : \mathbb{P}_N(I) \to \mathbb{P}_N$.*

We will apply this result to $h(x) = p_{t_0}(x)$, with t_0 as above. By the extension principle, we need to show that $h'[-] : \mathbb{P}_{N-1}((0,\rho)) \to \mathbb{P}_{N-1}$. Note that

$$h'(x) = t_0 \sum_{j=0}^{N-1} n_j c_{n_j} x^{n_j - 1} - M x^{M-1} = M g(x) + t_0 n_0 c_{n_0} x^{n_0 - 1},$$

where we define

$$g(x) := \frac{t_0}{M} \sum_{j=1}^{N-1} n_j c_{n_j} x^{n_j - 1} - x^{M-1}.$$

We **claim** that the entrywise polynomial map $g[-] : \mathbb{P}_{N-1}((0,\rho)) \to \mathbb{P}_{N-1}$. If this holds, then by the Schur product theorem, the same property is satisfied by $Mg(x) + t_0 n_0 c_{n_0} x^{n_0-1}$ (regardless of whether $n_0 = 0$ or $n_0 > 0$). But this function is precisely h' and the theorem would follow.

It thus remains to prove the claim, and we do so via a series of reductions and simplifications – i.e., "working backward". By the induction hypothesis, Theorem 43.2 holds in dimension $N-1 \geq 1$, for the polynomials

$$q_t(x) := t\sum_{j=1}^{N-1} n_j c_{n_j} x^{n_j-1} - x^{M-1}.$$

For this family, the threshold is now given by

$$\sum_{j=1}^{N-1} \frac{V(\mathbf{n}'_j)^2}{n_j c_{n_j} V(\mathbf{n}'_{\min})^2} \rho^{M-1-(n_j-1)},$$

where

$$\mathbf{n}'_{\min} := (0,1,\dots,N-2), \quad \mathbf{n}'_j := (n_1,\dots,n_{j-1},\widehat{n_j},n_{j+1},\dots,n_{N-1},M) \ \ \forall j>0.$$

Thus, the proof is complete if we show that

$$\sum_{j=1}^{N-1} \frac{V(\mathbf{n}'_j)^2}{n_j c_{n_j} V(\mathbf{n}'_{\min})^2} \rho^{M-1-(n_j-1)} \leq \frac{t_0}{M} = \sum_{j=0}^{N-1} \frac{V(\mathbf{n}_j)^2}{M c_{n_j} V(\mathbf{n}_{\min})^2} \rho^{M-n_j}.$$

In turn, comparing just the jth summand for each $j>0$, it suffices to show that

$$\frac{V(\mathbf{n}'_j)}{\sqrt{n_j}\,V(\mathbf{n}'_{\min})} \leq \frac{V(\mathbf{n}_j)}{\sqrt{M}\,V(\mathbf{n}_{\min})}, \qquad \forall j>0.$$

Dividing the right-hand side by the left-hand side, and canceling common factors, we obtain the expression

$$\prod_{k=1}^{N-1} \frac{n_k-n_0}{k} \cdot \frac{\sqrt{n_j}}{\sqrt{M}} \cdot \frac{M-n_0}{n_j-n_0}.$$

Since every factor in the product term is at least 1, it remains to show that

$$\frac{M-n_0}{n_j-n_0} \geq \frac{\sqrt{M}}{\sqrt{n_j}}, \qquad \forall j>0.$$

But this follows from a straightforward calculation:

$$(M-n_0)^2 n_j - (n_j-n_0)^2 M = (M-n_j)(Mn_j - n_0^2) > 0,$$

and the proof is complete. $\square$

Finally, we recall our original goal of classifying the sign patterns of positivity preservers in a fixed dimension – see Theorem 41.4. We showed this result holds if one can prove its special case, Theorem 41.5. Now this latter result follows from Theorem 43.2, by setting $c_M := -t_0^{-1}$, where $t_0 = \sum_{j=0}^{N-1} \frac{V(\mathbf{n}_j)^2}{c_{n_j} V(\mathbf{n}_{\min})^2} \rho^{M-n_j}$ as in Theorem 43.2. $\square$

44. Exact quantitative bound: monotonicity of Schur ratios. Real powers and power series.

In the last two sections (Sections 42 and 43), we proved the existence of a negative threshold c_M for polynomials

$$f(x) = \sum_{j=0}^{N-1} c_{n_j} x^{n_j} + c_M x^M$$

to entrywise preserve positivity on $\mathbb{P}_N((0,\rho))$. (Here $N > 0$ and $0 \leq n_0 < \cdots < n_{N-1} < M$ are integers.) We now compute the exact value of this threshold, more generally for *real* powers; this has multiple consequences which are described after stating Theorem 44.1. Thus, our goal is to prove the following quantitative result, for real powers – including negative powers:

Theorem 44.1. *Fix an integer $N > 0$ and real powers $n_0 < \cdots < n_{N-1} < M$. Also fix real scalars $\rho > 0$ and $c_{n_0}, \dots, c_{n_{N-1}}, c_M$, and define*

$$f(x) := \sum_{j=0}^{N-1} c_{n_j} x^{n_j} + c_M x^M. \tag{44.2}$$

Then the following are equivalent:

1. *The entrywise map $f[-]$ preserves positivity on all rank-1 matrices in $\mathbb{P}_N((0,\rho))$.*
2. *The map $f[-]$ preserves positivity on rank-1 totally non-negative (TN) Hankel matrices in $\mathbb{P}_N((0,\rho))$.*
3. *Either all $c_{n_j}, c_M \geqslant 0$; or $c_{n_j} > 0\ \forall j$ and $c_M \geqslant -\mathcal{C}^{-1}$, where*

$$\mathcal{C} = \sum_{j=0}^{N-1} \frac{V(\mathbf{n}_j)^2}{c_{n_j} V(\mathbf{n})^2} \rho^{M-n_j}. \tag{44.3}$$

Here $V(\mathbf{u}), \mathbf{n}, \mathbf{n}_j$ are defined as in (17.4) and (42.9).

If, moreover, we assume that $n_j \in \mathbb{Z}^{\geqslant 0} \cup [N-2, \infty)$ for all j, then the above conditions are further equivalent to the "full-rank" version:

4. *The entrywise map $f[-]$ preserves positivity on $\mathbb{P}_N([0,\rho])$, where we set $0^0 := 1$.*

Theorem 44.1 is a powerful result. It has multiple applications; we now list some of them.

1. Suppose $M = N$ and $n_j = j$ for $0 \leq j \leq N-1$. Then the result provides a complete characterization of which polynomials of degree $\leq N$ entrywise preserve positivity on $\mathbb{P}_N((0,\rho))$ – or more generally, on any intermediate set between $\mathbb{P}_N((0,\rho))$ and the rank-1 Hankel TN matrices inside it.
2. In fact, a similar result to the previous characterization is implied, whenever one considers linear combinations of at most $N+1$ monomial powers.
3. The result provides information on positivity preservers beyond polynomials, since n_j, M are now allowed to be real, even negative if one works with rank-1 matrices.
4. In particular, the result implies Theorem 41.6, and hence Theorem 41.7 (see its proof). This latter theorem provides a full classification of the sign patterns of possible "countable sums of real powers" which entrywise preserve positivity on $\mathbb{P}_N((0,\rho))$.
5. The result also provides information on preservers of total non-negativity on Hankel matrices in fixed dimension; see Corollary 44.11.

(6) There are further applications, two of which are (i) to the matrix cube problem and to sharp linear matrix inequalities/spectrahedra involving entrywise powers; and (ii) to computing the simultaneous kernels of entrywise powers and a related "Schubert cell-type" stratification of the cone $\mathbb{P}_N(\mathbb{C})$. These are explained in the 2016 paper of Belton, Guillot, Khare, and Putinar in *Adv. in Math.*; see also the 2021 paper in *Amer. J. Math.* by Khare and Tao (mentioned a few lines above (41.1)).

(7) Theorem 44.1 is proved using a monotonicity phenomenon for ratios of Schur polynomials; see Theorem 44.6. This latter result is also useful in extending a 2011 conjecture by Cuttler–Greene–Skandera (and its proof). In fact, this line of attack ends up *characterizing* majorization and weak majorization – for *real* tuples – using Schur polynomials. See the aforementioned paper by Khare and Tao [217] for more details.

(8) One further application is Theorem 44.14, which finds a threshold for bounding by $\sum_{j=0}^{N-1} c_{n_j} A^{\circ n_j}$, any power series – and more general "Laplace transforms" – applied entrywise to a positive matrix A. This extends Theorem 44.1, where the power series is simply x^M, because Theorem 44.1 says in particular that $(x^M)[A] = A^{\circ M}$ is dominated by a multiple of $\sum_{j=0}^{N-1} c_{n_j} A^{\circ n_j}$.

(9) As mentioned in the remarks prior to Theorem 41.4, Theorem 44.1 also provides examples of power series preservers on $\mathbb{P}_N((0,\rho))$ with negative coefficients; and of such functions which preserve positivity on $\mathbb{P}_N((0,\rho))$ but not on $\mathbb{P}_{N+1}((0,\rho))$.

44.1. **Monotonicity of ratios of Schur polynomials.** The proof of Theorem 44.1 uses the same ingredients as developed in previous sections. A summary of what follows is now provided. In the rank-1 case, we use a variant of Proposition 42.8 for an individual matrix; the result does not apply as is, since the powers may now be real. Next, in order to find the sharp threshold for all rank-1 matrices, even for real powers we crucially appeal to the integer power case. Namely, we will first understand the behavior and supremum of the function $s_{\mathbf{n}_j}(\mathbf{u})/s_{\mathbf{n}}(\mathbf{u})$ over $\mathbf{u} \in (0,\sqrt{\rho})^N$ (and for each $0 \leq j \leq N-1$). One may hope that these suprema behave well enough that the sharp threshold can be computed for rank-1 matrices; the further hope would be that this threshold bound is tight enough to behave well with respect to the extension principle in Theorem 9.12, and hence to work for all matrices in $\mathbb{P}_N((0,\rho))$. Remarkably, these two hopes are indeed justified, proving the theorem.

We begin with the key result required to be able to take suprema over ratios of Schur polynomials $s_{\mathbf{m}}(\mathbf{u})/s_{\mathbf{n}}(\mathbf{u})$. To motivate the result, here is a special case.

Example 44.4. Suppose $N = 3, \mathbf{n} = (0,2,3)$, and $\mathbf{m} = (0,2,4)$. As above, we have $\mathbf{u} = (u_1,u_2,u_3)^T$ and $\mathbf{n}_{\min} = (0,1,2)$. Now let $f(\mathbf{u}) := \frac{s_{\mathbf{m}}(\mathbf{u})}{s_{\mathbf{n}}(\mathbf{u})} : (0,\infty)^N \to (0,\infty)$. This is a rational function, whose numerator sums weights over tableaux of shape $(0,1,2)$, and hence by Example 42.3 above, equals $(u_1+u_2)(u_2+u_3)(u_3+u_1)$. The denominator sums weights over tableaux of shape $(0,1,1)$; there are only three such tableaux

$$\begin{array}{|c|}\hline 3 \\ \hline 2 \\ \hline \end{array} \qquad \begin{array}{|c|}\hline 3 \\ \hline 1 \\ \hline \end{array} \qquad \begin{array}{|c|}\hline 2 \\ \hline 1 \\ \hline \end{array}$$

and hence,

$$f(\mathbf{u}) := \frac{(u_1+u_2)(u_2+u_3)(u_3+u_1)}{u_1u_2+u_2u_3+u_3u_1}, \qquad u_1,u_2,u_3 > 0.$$

Notice that the numerator and denominator are both Schur polynomials, hence positive combinations of monomials (this is called "monomial positivity"). In particular, they are both non-decreasing in each coordinate. One can verify that their ratio $f(\mathbf{u})$ is not a polynomial; moreover, it is not a priori clear if $f(\mathbf{u})$ shares the same coordinatewise monotonicity property. However, we claim that this does hold, i.e., $f(\mathbf{u})$ *is non-decreasing in each coordinate on* $\mathbf{u} \in (0,\infty)^N$.

To see why: by symmetry, it suffices to show that f is non-decreasing in u_3. Using the quotient rule of differentiation, we claim that the expression

$$s_{\mathbf{n}}(\mathbf{u})\partial_{u_3} s_{\mathbf{m}}(\mathbf{u}) - s_{\mathbf{m}}(\mathbf{u})\partial_{u_3} s_{\mathbf{n}}(\mathbf{u}) \tag{44.5}$$

is non-negative on $(0,\infty)^3$. Indeed, computing this expression yields

$$(u_1+u_2)(u_1u_3+2u_1u_2+u_2u_3)u_3,$$

and this is clearly non-negative, as desired. More strongly, the expression (44.5) turns out to be monomial positive, which implies non-negativity.

Here is the punchline: an even stronger phenomenon holds. Namely, when we write the expression (44.5) in the form $\sum_{j\geq 0} p_j(u_1,u_2)u_3^j$, each polynomial p_j is *Schur positive*! This means that it is a non-negative integer-linear combination of Schur polynomials:

$$p_0(u_1,u_2) = 0,$$

$$p_1(u_1,u_2) = 2u_1u_2^2 + 2u_1^2u_2 = 2\,\begin{array}{|c|c|}\hline 2 & 2 \\ \hline 1 & \multicolumn{1}{c}{} \\ \cline{1-1}\end{array} + 2\,\begin{array}{|c|c|}\hline 2 & 1 \\ \hline 1 & \multicolumn{1}{c}{} \\ \cline{1-1}\end{array} = 2s_{(1,3)}(u_1,u_2),$$

$$p_2(u_1,u_2) = (u_1+u_2)^2 = \begin{array}{|c|c|}\hline 2 & 2 \\ \hline\end{array} + \begin{array}{|c|c|}\hline 2 & 1 \\ \hline\end{array} + \begin{array}{|c|c|}\hline 1 & 1 \\ \hline\end{array} + \begin{array}{|c|}\hline 2 \\ \hline 1 \\ \hline\end{array} = s_{(0,3)}(u_1,u_2) + s_{(1,2)}(u_1,u_2),$$

modulo a mild abuse of notation. This yields the sought-for non-negativity, as each $s_{\mathbf{n}}(\mathbf{u})$ is monomial positive by definition. (See the discussion following (30.17) for the "original" occurrence of monomial positivity and its "upgrade" to (skew) Schur positivity.)

The remarkable fact is that the phenomena described in the above example also occur for every pair of Schur polynomials $s_{\mathbf{m}}(\mathbf{u}), s_{\mathbf{n}}(\mathbf{u})$ for which $\mathbf{m} \geq \mathbf{n}$ coordinatewise:

Theorem 44.6 (Monotonicity of Schur polynomial ratios)**.** *Suppose* $0 \leqslant n_0 < \cdots < n_{N-1}$ *and* $0 \leqslant m_0 < \cdots < m_{N-1}$ *are integers satisfying:* $n_j \leqslant m_j\ \forall j$. *Then the symmetric function*

$$f:(0,\infty)^N \to \mathbb{R}, \qquad f(\mathbf{u}) := \frac{s_{\mathbf{m}}(\mathbf{u})}{s_{\mathbf{n}}(\mathbf{u})}$$

is non-decreasing in each coordinate.

More strongly, viewing the expression

$$s_{\mathbf{n}}(\mathbf{u}) \cdot \partial_{u_N} s_{\mathbf{m}}(\mathbf{u}) - s_{\mathbf{m}}(\mathbf{u}) \cdot \partial_{u_N} s_{\mathbf{n}}(\mathbf{u})$$

as a polynomial in u_N, *the coefficient of each monomial* u_N^j *is a Schur positive polynomial in* $(u_1,u_2,\dots,u_{N-1})^T$.

Theorem 44.6 is an application of a deep result in representation theory/symmetric function theory, by Lam, Postnikov, and Pylyavskyy in *Amer. J. Math.* (2007). The proof of this latter result is beyond the scope of this text, and hence is not pursued further; but its usage means that in the spirit of the previous two sections, the proof of Theorem 44.6 once again combines analysis with symmetric function theory. Moreover, this 2007 result in [231] arose from the prior work of Skandera [339] in 2004, on determinant inequalities for minors of totally non-negative matrices.

To proceed further, we introduce the following notation:

Definition 44.7. Given a vector $\mathbf{u} = (u_1, \dots, u_m)^T \in (0,\infty)^m$ and a real tuple $\mathbf{n} = (n_0, \dots, n_{N-1})$ for integers $m, N \geq 1$, define

$$\mathbf{u}^{\circ \mathbf{n}} := (\mathbf{u}^{\circ n_0} \mid \dots \mid \mathbf{u}^{\circ n_{N-1}})_{m \times N} = (u_j^{n_{k-1}})_{j=1,k=1}^{m,N}.$$

We now extend Theorem 44.6 to arbitrary real powers (instead of non-negative integer powers). As one can no longer use Schur polynomials, the next result uses generalized Vandermonde determinants instead:

Theorem 44.8. *Fix an integer $N \geq 1$ and real tuples*

$$\mathbf{n} = (n_0 < n_1 < \cdots < n_{N-1}), \qquad \mathbf{m} = (m_0 < m_1 < \cdots < m_{N-1})$$

with $n_j \leq m_j \ \forall j$ and $\mathbf{n} \neq \mathbf{m}$. Then the symmetric function

$$f_{\neq}(\mathbf{u}) := \frac{\det(\mathbf{u}^{\circ \mathbf{m}})}{\det(\mathbf{u}^{\circ \mathbf{n}})}$$

is strictly increasing in each coordinate on $(0,\infty)^{N,\neq}$. (See Definition 25.1.) If $n_0 = m_0 = 0$, then $f_{\neq}(\mathbf{u})$ is strictly increasing in each coordinate on $[0,\infty)^{N,\neq}$.

While we only require $f_{\neq}$ to be non-decreasing in each coordinate, and only on $(0,\infty)^{N,\neq}$, we will show this stronger result.

Proof. The result is immediate for $N = 1$; henceforth suppose $N \geq 2$. For a fixed $t \in \mathbb{R}$, if for each j we multiply the jth row of the matrix $\mathbf{u}^{\circ \mathbf{m}}$ by u_j^t, we obtain a matrix $\mathbf{u}^{\circ \mathbf{m}'}$ where $\mathbf{m}'_j = m_j + t \ \forall j$. In particular, if we start with real powers n_j, m_j, then multiplying the numerator and denominator of $f_{\neq}$ by $(u_1 \cdots u_N)^{-n_0}$ reduces the situation to working with the non-negative real tuples $\mathbf{n}' := (n_j - n_0)_{j=0}^{N-1}$ and $\mathbf{m}' := (m_j - n_0)_{j=0}^{N-1}$. Thus, we suppose henceforth that $n_j, m_j \geq 0 \ \forall j$.

We first show that $f_{\neq}$ is non-decreasing in each coordinate. If n_j, m_j are all integers, then the result is an immediate reformulation of the first part of Theorem 44.6, via Proposition 42.6(1). Next suppose n_j, m_j are rational. Choose a (large) integer $L > 0$ such that $Ln_j, Lm_j \in \mathbb{Z} \ \forall j$ and define $y_j := u_j^{1/L}$. By the previous subcase, the symmetric function

$$f(\mathbf{y}) := \frac{\det(\mathbf{y}^{\circ L\mathbf{m}})}{\det(\mathbf{y}^{\circ L\mathbf{n}})} = \frac{\det(\mathbf{u}^{\circ \mathbf{m}})}{\det(\mathbf{u}^{\circ \mathbf{n}})}, \qquad \mathbf{y} := (y_1, \dots, y_N)^T \in (0,\infty)^{N,\neq}$$

is coordinatewise non-decreasing on $(0,\infty)^{N,\neq}$ in the y_j, and hence on $(0,\infty)^{N,\neq}$ in the u_j.

Finally, in the general case, given non-negative real powers n_j, m_j satisfying the hypotheses, choose sequences

$$0 \leq n_{0,k} < n_{1,k} < \cdots < n_{N-1,k}, \qquad 0 \leq m_{0,k} < m_{1,k} < \cdots < m_{N-1,k}$$

for $k = 1, 2, \dots$, which further satisfy:

(1) $n_{j,k}, m_{j,k}$ are rational for $0 \leq j \leq N-1, \ k \geq 1$;
(2) $n_{j,k} \leq m_{j,k} \ \forall j, k$; and
(3) $n_{j,k} \to n_j$ and $m_{j,k} \to m_j$ as $k \to \infty$, for each $j = 0, 1, \dots, N-1$.

By the rational case above, for each $k \geq 1$ the symmetric function

$$f_k(\mathbf{u}) := \frac{\det(\mathbf{u}^{\circ \mathbf{m}_k})}{\det(\mathbf{u}^{\circ \mathbf{n}_k})}$$

is coordinatewise non-decreasing, where $\mathbf{m}_k := (m_{0,k}, \dots, m_{N-1,k})$ and similarly for $\mathbf{n}_k$. But then their limit $\lim_{k\to\infty} f_k(\mathbf{u}) = f_{\neq}(\mathbf{u})$ is also coordinatewise non-decreasing, as claimed.

The next step is to show that $f_{\neq}$ is strictly increasing on $(0,\infty)^{N,\neq}$ in each coordinate, say in u_N by symmetry. Suppose instead that $f_{\neq}$ is constant on the intersection with $\mathbb{R}^{N,\neq}$ of a line segment

$$\{(u_1,\dots,u_{N-1},u_N) : u_N \in [x,x']\} \subset (0,\infty)^N,$$

where $u_1,\dots,u_{N-1} \in (0,\infty)$ are fixed and $0 < x < x'$. Here we may replace $[x,x']$ by a smaller sub-interval (still of positive length) that does not contain $u_1,\dots,u_{N-1}$; thus, without loss of generality the above segment is contained in $(0,\infty)^{N,\neq}$. Now evaluating $f_{\neq}$ as a function of u_N, we obtain a constant function of the form

$$h(u_N) := \frac{\sum_{j=0}^{N-1} u_N^{m_j} g_j}{\sum_{j=0}^{N-1} u_N^{n_j} g_j'}, \qquad u_N \in [x,x']$$

where g_j, g_j' are generalized Vandermonde determinants, hence all non-zero. Denoting the numerator and denominator by $h_1(u_N), h_2(u_N)$ respectively, if $h(\cdot) \equiv c$ on $[x,x']$ for some $c \in \mathbb{R}$, then $h_1 - ch_2$ has infinitely many zeros on $[x,x']$. Since $\mathbf{m} \neq \mathbf{n}$, this contradicts Descartes' rule of signs (Lemma 5.2).

Finally, we show that $f_{\neq}$ is strictly increasing in each coordinate u_j at $\mathbf{u}$, where one coordinate of $\mathbf{u}$, say u_1, equals zero – and $n_0 = m_0 = 0$. There are two cases: if $j > 1$, then both $\mathbf{u}^{\circ \mathbf{m}}$ and $\mathbf{u}^{\circ \mathbf{n}}$ are matrices with the first column $(1,\dots,1)^T$ and first row $\mathbf{e}_1^T = (1,0,\dots,0)$. But then $\det(\mathbf{u}^{\circ \mathbf{m}}) = \det(\mathbf{u}_1^{\circ \mathbf{m}_1})$, where $\mathbf{v}_1$ for a vector $\mathbf{v}$ is the sub-vector that removes the first coordinate. Hence,

$$f_{\neq}(\mathbf{u}) = \frac{\det(\mathbf{u}_1^{\circ \mathbf{m}_1})}{\det(\mathbf{u}_1^{\circ \mathbf{n}_1})},$$

and since $\mathbf{u}_1 \in (0,\infty)^{N-1,\neq}$, the right-hand side is strictly increasing in u_j for $j > 1$ by the above analysis – now on $(0,\infty)^{N-1,\neq}$.

The other case is if $j = 1$. Then we consider $\mathbf{v} := \mathbf{u} + \mu \mathbf{e}_1$ for some $\mu \in (0,\infty)$. If $c := \min(\mu, \min_{j>1} u_j)$, then $\mathbf{v}(\epsilon) := \mathbf{u} + \epsilon \mathbf{e}_1$ lies in $(0,\infty)^{N,\neq}$ for $\epsilon \in (0,c)$, so by above, $f_{\neq}(\mathbf{v}(\cdot))$ is strictly increasing as a function of $\epsilon \in (0,c)$. Hence, for $\epsilon \in (0,c/4)$, we have from above:

$$f_{\neq}(\mathbf{v}) = f_{\neq}(\mathbf{v}(\mu)) > f_{\neq}(\mathbf{v}(c/4)) > f_{\neq}(\mathbf{v}(\epsilon)).$$

Letting $\epsilon \to 0^+$, the proof is complete:

$$f_{\neq}(\mathbf{v}) > f_{\neq}(\mathbf{v}(c/4)) \geq \lim_{\epsilon \to 0^+} f_{\neq}(\mathbf{v}(\epsilon)) = f_{\neq}(\mathbf{v}(0)) = f_{\neq}(\mathbf{u}). \qquad \square$$

44.2. **Proof of the quantitative bound.** Using Theorem 44.8, we can now prove the main result in this section.

Proof of Theorem 44.1. We first work only with rank-1 matrices. Clearly, (1) $\implies$ (2), and we show that (2) $\implies$ (3) $\implies$ (1).

If all coefficients $c_{n_j}, c_M \geq 0$, then $f[-]$ preserves positivity on rank-1 matrices. Otherwise, by the Horn–Loewner-type necessary conditions in Lemma 41.2 (now for real powers, possibly negative!), it follows that $c_{n_0}, \dots, c_{n_{N-1}} > 0 > c_M$. In this case, the discussion that opens Section 42.2 allows us to reformulate the problem using

$$p_t(x) := t \sum_{j=0}^{N-1} c_{n_j} x^{n_j} - x^M, \qquad t > 0,$$

and the goal is to find a *sharp* positive lower bound for t, above which $p_t[-]$ preserves positivity on rank-1 Hankel TN matrices $\mathbf{u}\mathbf{u}^T \in \mathbb{P}_N((0,\rho))$.

But now one can play the same game as in Section 42.2. In other words, Lemma 42.10 shows that the "real powers analogue" of Proposition 42.8 holds: $p_t[\mathbf{u}\mathbf{u}^T] \geq 0$ if and only if

$$t \geq \sum_{j=0}^{N-1} \frac{\det(\mathbf{u}^{\circ \mathbf{n}_j})^2}{c_{n_j} \det(\mathbf{u}^{\circ \mathbf{n}})^2},$$

for all generic rank-1 matrices $\mathbf{u}\mathbf{u}^T$, with $\mathbf{u} \in (0, \sqrt{\rho})^{N,\neq}$. By the same reasoning as in the proof of Theorem 43.2 (see the previous section), $p_t[-]$ preserves positivity on a given test set of rank-1 matrices $\{\mathbf{u}\mathbf{u}^T : \mathbf{u} \in S \subset (0,\sqrt{\rho})^N\}$, if and only if (by density and continuity,) t exceeds the following supremum:

$$t \geq \sup_{\mathbf{u} \in S \cap (0,\sqrt{\rho})^{N,\neq}} \sum_{j=0}^{N-1} \frac{\det(\mathbf{u}^{\circ \mathbf{n}_j})^2}{c_{n_j} \det(\mathbf{u}^{\circ \mathbf{n}})^2}. \tag{44.9}$$

This is, of course, subject to $S \cap (0,\sqrt{\rho})^{N,\neq}$ being dense in the set S, which is indeed the case if $\{\mathbf{u}\mathbf{u}^T : \mathbf{u} \in S \cap (0,\sqrt{\rho})^N\}$ equals the set of rank-1 Hankel TN matrices as in assertion (2).

Thus, to prove (2) $\implies$ (3) $\implies$ (1) in the theorem, it suffices to prove: (i) the supremum (44.9) is bounded above by the value $\displaystyle\sum_{j=0}^{N-1} \frac{V(\mathbf{n}_j)^2}{c_{n_j} V(\mathbf{n})^2} \rho^{M-n_j}$; and (ii) this value is attained on (a countable set of) rank-1 Hankel TN matrices, hence it equals the supremum.

We now prove both of these assertions. By Theorem 44.8, each ratio $\det(\mathbf{u}^{\circ \mathbf{n}_j}) / \det(\mathbf{u}^{\circ \mathbf{n}})$ is coordinatewise non-decreasing, hence its supremum on $(0,\sqrt{\rho})^{N,\neq}$ is bounded above by (and in fact equals) its limit as $\mathbf{u} \to \sqrt{\rho}(1^-,\dots,1^-)$. To see why this limit exists, note that every vector $u \in (0,\rho)^N$ is bounded above – coordinatewise – by a vector of the form

$$\mathbf{u}(\epsilon) := \sqrt{\rho}(\epsilon, \epsilon^2, \dots, \epsilon^N)^T \in (0,\sqrt{\rho})^{N,\neq}, \qquad \epsilon \in (0,1).$$

In particular, by Theorem 44.8 the limit as $\mathbf{u} \to \sqrt{\rho}(1^-,\dots,1^-)$ exists and equals the limit by using the rank-1 Hankel TN family $\mathbf{u}(\epsilon)\mathbf{u}(\epsilon)^T$, for any sequence of $\epsilon \to 1^-$ – provided this latter limit exists. We show this presently; thus, we work with a countable sequence of $\epsilon \to 0^+$ in place of Lemma 41.2, and another countable sequence of $\epsilon \to 1^-$ in what follows. First observe:

Lemma 44.10 (Principal specialization formula for real powers)**.** *Suppose $q > 0$ and $n_0 < n_1 < \cdots < n_{N-1}$ are real exponents. If $\mathbf{n} := (n_0,\dots,n_{N-1})$ and $\mathbf{u} := (1, q, \dots, q^{N-1})^T$, then*

$$\det(\mathbf{u}^{\circ \mathbf{n}}) = \prod_{0 \leq j < k \leq N-1} (q^{n_k} - q^{n_j}) = V(q^{\circ \mathbf{n}}).$$

The proof is exactly the same as of Proposition 42.6(2), since the transpose of $\mathbf{u}^{\circ \mathbf{n}}$ is a *usual* Vandermonde matrix.

We can now complete the proof of Theorem 44.1. The above lemma immediately implies

$$\frac{\det(\mathbf{u}(\epsilon)^{\circ \mathbf{n}_j})}{\det(\mathbf{u}(\epsilon)^{\circ \mathbf{n}})} = \sqrt{\rho}^{M-n_j} \frac{V(\epsilon^{\circ \mathbf{n}_j})}{V(\epsilon^{\circ \mathbf{n}})}, \qquad \forall 0 \leq j \leq N-1.$$

Dividing the numerator and denominator by $(1-\epsilon)^{\binom{N}{2}}$ and taking the limit as $\epsilon \to 1^-$ using L'Hôpital's rule, we obtain the expression $\sqrt{\rho}^{M-n_j} \frac{V(\mathbf{n}_j)}{V(\mathbf{n})}$. Since all of these suprema/limits occur as $\epsilon \to 1^-$, we finally have

$$\sup_{\mathbf{u} \in (0,\sqrt{\rho})^{N,\neq}} \sum_{j=0}^{N-1} \frac{\det(\mathbf{u}^{\circ \mathbf{n}_j})^2}{c_{n_j} \det(\mathbf{u}^{\circ \mathbf{n}})^2} = \lim_{\epsilon \to 1^-} \sum_{j=0}^{N-1} \frac{\det(\mathbf{u}(\epsilon)^{\circ \mathbf{n}_j})^2}{c_{n_j} \det(\mathbf{u}(\epsilon)^{\circ \mathbf{n}})^2} = \sum_{j=0}^{N-1} \frac{V(\mathbf{n}_j)^2}{V(\mathbf{n})^2} \frac{\rho^{M-n_j}}{c_{n_j}}.$$

This proves the equivalence of assertions (1)–(3) in the theorem, for rank-1 matrices.

Finally, suppose all $n_j \in \mathbb{Z}^{\geq 0} \cup [N-2, \infty)$. In this case (4) $\implies$ (1) is immediate. Conversely, given that (1) holds, we prove (4) using once again the integration trick of FitzGerald and Horn, as isolated in Theorem 9.12. The proof and calculation are similar to that of Theorem 43.2 above and are left to the interested reader as an exercise. $\square$

44.3. **Applications to Hankel TN preservers in a fixed dimension and to power series preservers.** We conclude by discussing some applications of Theorem 44.1. First, the result implies in particular that $A^{\circ M}$ is bounded above by a multiple of $\sum_{j=0}^{N-1} c_{n_j} A^{\circ n_j}$. In particular, the proof of Theorem 41.7 above goes through; thus, we have classified the sign patterns of all entrywise power series preserving positivity on $\mathbb{P}_N((0, \rho))$.

Second, the equivalent conditions in Theorem 44.1 classifying the (entrywise) polynomial positivity preservers on $\mathbb{P}_N((0, \rho))$ – or on rank-1 matrices – also end up classifying the polynomial preservers of *total non-negativity* on the corresponding Hankel test sets:

Corollary 44.11. *With notation as in Theorem 44.1, if we restrict to all real powers and only rank-1 matrices, then assertions (1)–(3) in Theorem 44.1 are further equivalent to:*

(1′) $f[-]$ preserves total non-negativity on all rank-1 matrices in $\mathtt{HTN}_N$ with entries in $(0, \rho)$.

If, moreover, all n_j lie in $\mathbb{Z}^{\geq 0} \cup [N-2, \infty)$, then these conditions are further equivalent to:

(4′) $f[-]$ preserves total non-negativity on all matrices in $\mathtt{HTN}_N$ with entries in $[0, \rho]$.

Recall here that by Definition 12.18, $\mathtt{HTN}_N$ denotes the set of $N \times N$ Hankel totally non-negative matrices.

Proof. Clearly, (4′) implies (1′), which implies assertion (2) in Theorem 44.1. Conversely, we claim that assertion (1) in Theorem 44.1 implies (1′) via Theorem 4.1. Indeed, if $A \in \mathtt{HTN}_N$ has rank 1 and entries in $(0, \rho)$, then $f[A] \in \mathbb{P}_N$ by Theorem 44.1(1). Similarly, $A^{(1)} \oplus (0)_{1\times 1} \in \mathbb{P}_N((0, \rho))$ and has rank 1, so $f[A^{(1)}]$ is also positive semidefinite, and hence Theorem 4.1 applies, as desired. The same proof works to show that (4′) follows from Theorem 44.1(4). $\square$

The third and final application is to bounding $g[A]$, where $g(x)$ is a power series – or more generally, a linear combination of real powers – by a threshold times $\sum_{j=0}^{N-1} c_{n_j} A^{\circ n_j}$. This extends Theorem 44.1 in which $g(x) = x^M$. The idea is that if we fix exponents $0 \leq n_0 < \cdots < n_{N-1}$ and coefficients c_{n_j} for $j = 0, \dots, N-1$, then

$$A^{\circ M} \leq t_M \sum_{j=0}^{N-1} c_{n_j} A^{\circ n_j}, \qquad \text{where } t_M := \sum_{j=0}^{N-1} \frac{V(\mathbf{n}_j)^2}{c_{n_j} V(\mathbf{n})^2} \rho^{M-n_j}, \tag{44.12}$$

and this linear matrix inequality holds for all $A \in \mathbb{P}_N((0, \rho))$ – possibly of rank 1 if the n_j are allowed to be arbitrary non-negative real numbers, else of all ranks if all $n_j \in \mathbb{Z}^{\geq 0} \cup [N-2, \infty)$. Here the $\leq$ stands for the positive semidefinite ordering, or Loewner ordering – see, e.g., Definition 14.7. Moreover, the constant t_M depends on M through $\mathbf{n}_j$ and ρ^{M-n_j}.

If now we consider a power series $g(x) := \sum_{M \geq n_{N-1}+1} c_M x^M$, then by adding several linear matrix inequalities of the form (44.12), it follows that

$$g[A] \leq t_g \sum_{j=0}^{N-1} c_{n_j} A^{\circ n_j}, \qquad \text{where } t_g := \sum_{M \geq n_{N-1}+1} \max(c_M, 0) t_M,$$

and this is a valid linear matrix inequality, as long as the sum t_g is convergent. Thus, we now explore when this sum converges.

Even more generally: notice that a power series is the sum/integral of the power function, over a measure on the *powers* which is supported on the integers. Thus, given any real measure μ supported in $[n_{N-1} + \varepsilon, \infty)$, one can consider its corresponding "Laplace transform"

$$g_\mu(x) := \int_{n_{N-1}+\varepsilon}^{\infty} x^t \; d\mu(t). \tag{44.13}$$

The final application of Theorem 44.1 explores in this generality, when a finite threshold exists to bound $g_\mu[A]$ by a sum of N lower powers.

Theorem 44.14. *Fix $N \geq 2$ and real exponents $0 \leq n_0 < \cdots < n_{N-1}$ in the set $\mathbb{Z}^{\geq 0} \cup [N-2, \infty)$. Also fix scalars $\rho, c_{n_j} > 0$ for all j.*

Now suppose $\varepsilon, \varepsilon' > 0$ and μ is a real measure supported on $[n_{N-1} + \varepsilon, \infty)$ such that $g_\mu(x)$ – defined as in (44.13) – is absolutely convergent at $\rho(1 + \varepsilon')$. Then there exists a finite constant $t_\mu \in (0, \infty)$, such that the map

$$t_\mu \sum_{j=0}^{N-1} c_{n_j} x^{n_j} - g_\mu(x)$$

entrywise preserves positivity on $\mathbb{P}_N((0, \rho))$. Equivalently, $g_\mu[A] \leq t_\mu \sum_{j=0}^{N-1} c_{n_j} A^{\circ n_j}$, for all $A \in \mathbb{P}_N((0, \rho))$.

Proof. If $\mu = \mu_+ - \mu_-$ denotes the decomposition of μ into its positive and negative parts, then notice (e.g., by the FitzGerald–Horn Theorem 9.3) that

$$\int_{\mathbb{R}} A^{\circ M} \; d\mu_-(M) \in \mathbb{P}_N, \qquad \forall A \in \mathbb{P}_N((0, \rho)).$$

Hence, it suffices to show that

$$t_\mu := \int_{n_{N-1}+\varepsilon}^{\infty} t_M \; d\mu_+(M) = \int_{n_{N-1}+\varepsilon}^{\infty} \sum_{j=0}^{N-1} \frac{V(\mathbf{n}_j)^2}{c_{n_j} V(\mathbf{n})^2} \rho^{M - n_j} \; d\mu_+(M) < \infty, \tag{44.15}$$

since this would imply

$$t_\mu \sum_{j=0}^{N-1} c_{n_j} A^{\circ n_j} - g_\mu[A] = \int_{n_{N-1}+\varepsilon}^{\infty} \left(t_M \sum_{j=0}^{N-1} c_{n_j} A^{\circ n_j} - A^{\circ M} \right) d\mu_+(M) + \int_{n_{N-1}+\varepsilon}^{\infty} A^{\circ M} \; d\mu_-(M),$$

and both integrands and integrals are positive semidefinite.

In turn, isolating the terms in (44.15) that depend on M, it suffices to show for each j that

$$\int_{n_{N-1}+\varepsilon}^{\infty} \prod_{k=0, k \neq j}^{N-1} (M - n_k)^2 \rho^M \; d\mu_+(M) < \infty.$$

By linearity, it suffices to examine the finiteness of the integrals

$$\int_{n_{N-1}+\varepsilon}^{\infty} M^k \rho^M \ d\mu_+(M), \qquad k \geq 0.$$

But by assumption, $\int_{n_{N-1}+\varepsilon}^{\infty} \rho^M (1+\varepsilon')^M \ d\mu_+(M)$ is finite; and moreover, for any fixed $k \geq 0$ there is a threshold M_k beyond which $(1+\varepsilon')^M \geq M^k$. (Indeed, this happens when $\frac{\log M}{M} \leq \frac{\log(1+\varepsilon')}{k}$.) Therefore,

$$\int_{n_{N-1}+\varepsilon'}^{\infty} M^k \rho^M \ d\mu_+(M) \leq \int_{n_{N-1}+\varepsilon'}^{M_k} M^k \rho^M \ d\mu_+(M) + \int_{M_k}^{\infty} \rho^M (1+\varepsilon')^M \ d\mu_+(M) < \infty,$$

which concludes the proof. $\square$

45. Polynomial preservers on matrices with real or complex entries. Complete homogeneous symmetric polynomials.

Having discussed in detail the case of matrices with entries in $(0, \rho)$, we conclude this part of the text with a brief study of entrywise polynomials preserving positivity in fixed dimension – but now on matrices with possibly negative or even complex entries. The first observation is that non-integer powers can no longer be applied, so we restrict ourselves to polynomials. Second, as discussed following the proof of Lemma 41.2, it is not possible to obtain structured results along the same lines as above, for all matrices in $\mathbb{P}_N((-\rho, \rho))$, for every polynomial of the form

$$t(c_{n_0}x^{n_0} + \cdots + c_{n_{N-1}}x^{n_{N-1}}) - x^M$$

acting entrywise.

The way one now proceeds is as follows. Akin to the two previous sections, the analysis begins by bounding from above the ratio $s_{\mathbf{n}_j}(\mathbf{u})^2/s_{\mathbf{n}}(\mathbf{u})^2$ on the domain – in this case, on $[-\rho, \rho]^N$. Since the numerator and denominator both vanish at the origin, a sufficient condition to proceed would be that the zero locus of the denominator $s_{\mathbf{n}}(\cdot)$ is contained in the zero locus of $s_{\mathbf{n}_j}(\cdot)$ for every j. Since the choice of $M > n_{N-1}$ is arbitrary, we therefore try to seek the best possible solution: namely, that $s_{\mathbf{n}}(\cdot)$ *does not vanish on* $\mathbb{R}^N \setminus \{\mathbf{0}\}$. And indeed, it is possible to completely characterize all such tuples $\mathbf{n}$:

Theorem 45.1. *Fix integers $N \geq 2$ and $0 \leq n_0 < \cdots < n_{N-1}$. The following are equivalent:*

(1) The Schur polynomial $s_{\mathbf{n}}(\cdot) : \mathbb{R}^N \to \mathbb{R}$ is positive except possibly at the origin.
(2) The Schur polynomial $s_{\mathbf{n}}(\cdot) : \mathbb{R}^N \to \mathbb{R}$ is non-vanishing except possibly at the origin.
(3) The Schur polynomial $s_{\mathbf{n}}(\cdot)$ does not vanish at the two vectors $\mathbf{e}_1$ and $\mathbf{e}_1 - \mathbf{e}_2$.
(4) The tuple $\mathbf{n}$ satisfies: $n_0 = 0, \dots, n_{N-2} = N-2$, and $n_{N-1} - (N-1) = 2r \geq 0$ is an even integer.

Using Littlewood's definition (42.2), it is easy to see that such a polynomial is precisely the complete homogeneous symmetric polynomial (of even degree $k = 2r$)

$$h_k(u_1, u_2, \dots) := \sum_{1 \leq j_1 \leq j_2 \leq \cdots \leq j_k} u_{j_1} u_{j_2} \cdots u_{j_k}, \qquad \forall u_j \in \mathbb{R}$$

for $k \geq 0$, where we set $h_0(u_1, u_2, \dots) \equiv 1$.

In this section, we will prove Theorem 45.1 and apply it to study entrywise polynomial preservers of positivity over $\mathbb{P}_N((-\rho, \rho))$. We then study such preservers of $\mathbb{P}_N(D(0, \rho))$.

45.1. Complete homogeneous symmetric polynomials are always positive.

The major part of Theorem 45.1 is to show that the polynomials h_{2r} do not vanish outside the origin. This is a result by Hunter in 1977 in *Math. Proc. Camb. Phil. Soc.* More strongly, Hunter showed that these polynomials are always positive, with a strict lower bound:

Theorem 45.2 (Hunter, 1977, [187])**.** *Fix integers $r, N \geq 1$. Then we have*

$$h_{2r}(\mathbf{u}) \geq \frac{\|\mathbf{u}\|^{2r}}{2^r r!}, \qquad \mathbf{u} \in \mathbb{R}^N \tag{45.3}$$

with equality if and only if (a) $\min(r, N) = 1$ and (b) $\sum_{j=1}^N u_j = 0$.

The proof uses two observations, also made by Hunter in the same work.

Lemma 45.4 (Hunter, 1977, [187]). *Given integers $k, N \geq 1$, and $\mathbf{u} = (u_1, \dots, u_N)^T \in \mathbb{R}^N$,*

$$\frac{h_k(\mathbf{u}, \xi) - h_k(\mathbf{u}, \eta)}{\xi - \eta} = h_{k-1}(\mathbf{u}, \xi, \eta)$$

for all real $\xi \neq \eta$; and moreover,

$$\frac{\partial h_k}{\partial u_j}(\mathbf{u}) = h_{k-1}(\mathbf{u}, u_j).$$

Proof. Recall from the definition that

$$h_k(\mathbf{u}) = \sum_{s=0}^{k} h_{k-s}(u_1, \dots, u_{N-1}) u_N^s. \tag{45.5}$$

With (45.5) at hand, the first assertion follows immediately (in fact over any ground field):

$$\frac{h_k(\mathbf{u}, \xi) - h_k(\mathbf{u}, \eta)}{\xi - \eta} = \sum_{s=0}^{k} h_{k-s}(\mathbf{u}) \frac{\xi^s - \eta^s}{\xi - \eta} = \sum_{s=1}^{k} h_{k-s}(\mathbf{u}) \sum_{t=0}^{s-1} \xi^t \eta^{s-1-t} = h_{k-1}(\mathbf{u}, \xi, \eta).$$

We isolate the final equality here:

$$h_{k-1}(\mathbf{u}, \xi, \eta) = \sum_{s=1}^{k} h_{k-s}(\mathbf{u}) \sum_{t=0}^{s-1} \xi^t \eta^{s-1-t}, \tag{45.6}$$

noting that it holds at all $\xi, \eta \in \mathbb{R}$. Next, we show the second assertion. Since h_k is a symmetric polynomial, it suffices to work with $j = N$. Now compute using (45.5) and (45.6)

$$\begin{aligned} \frac{\partial h_k}{\partial u_j}(\mathbf{u}) &= \sum_{s=1}^{k} h_{k-s}(u_1, \dots, u_{N-1})(s u_N^{s-1}) \\ &= \sum_{s=1}^{k} h_{k-s}(u_1, \dots, u_{N-1}) \sum_{t=0}^{s-1} u_N^t u_N^{s-1-t} = h_{k-1}(\mathbf{u}, u_N). \qquad \square \end{aligned}$$

With Lemma 45.4 at hand, we proceed.

Proof of Theorem 45.2. If $N = 1 \leq r$, it is easy to see that (45.3) holds if and only if $u_1 = 0$. Similarly, if $r = 1$, then

$$h_2(\mathbf{u}) = \frac{1}{2} \left[\|\mathbf{u}\|^2 + (\sum_{j=1}^{N} u_j)^2 \right] \geq \frac{\|\mathbf{u}\|^2}{2},$$

with equality if and only if $\sum_{j=1}^N u_j = 0$, as desired.

Henceforth we suppose that $r, N \geq 2$, and claim by induction on r that (45.3) holds, with a strict inequality. To show the claim, note that since $h_{2r}(\mathbf{u})$ is homogeneous in $\mathbf{u}$ of total degree $2r$, it suffices to show (45.3) on the unit sphere

$$h_{2r}(\mathbf{u}) > \frac{1}{2^r r!}, \qquad \mathbf{u} \in S^{N-1}.$$

We are thus interested in optimizing (in fact minimizing) the smooth function $h_{2r}(\mathbf{u})$, subject to the constraint $\sum_{j=1}^N u_j^2 = 1$. This problem is amenable to the use of Lagrange multipliers, and we obtain that at any extreme point $\mathbf{y} \in S^{N-1}$, there exists $\lambda \in \mathbb{R}$ satisfying

$$\frac{\partial h_{2r}}{\partial u_j}(\mathbf{y}) + 2\lambda y_j = 0, \qquad j = 1, \dots, N.$$

Multiply this equation by y_j and sum over all j; since h_{2r} is homogeneous of total degree $2r$, Euler's equation yields

$$2rh_{2r}(\mathbf{y}) + 2\lambda\|\mathbf{y}\|^2 = 0 \quad \implies \quad \lambda = -rh_{2r}(\mathbf{y}).$$

With this at hand, compute using Lemma 45.4

$$h_{2r-1}(\mathbf{y}, y_j) = \frac{\partial h_{2r}}{\partial u_j}(\mathbf{y}) = -2\lambda y_j = 2rh_{2r}(\mathbf{y})y_j, \qquad j = 1, \dots, N. \tag{45.7}$$

We now show that at all points $\mathbf{y} \in S^{N-1}$ satisfying (45.7), one has (45.3) with a strict inequality. As one of these points is the global minimum, this would prove the result.

There are two cases. First, the vectors $\mathbf{y}_\pm := \frac{\pm 1}{\sqrt{N}}\mathbf{1}_{N\times 1} \in S^{N-1}$ satisfy (45.7); it may help here to observe that the number of terms/monomials in $h_k(u_1, \dots, u_N)$ is $\binom{N+k-1}{k}$. This observation also implies that at these points $\mathbf{y}_\pm$, we have

$$h_{2r}(\mathbf{y}_\pm) = \binom{N+2r-1}{2r}\frac{1}{N^{2r}} = \frac{(N+2r-1)(N+2r-2)\cdots N}{N^{2r}}\frac{1}{(2r)!} > \frac{1}{(2r)!} > \frac{1}{2^r r!},$$

and this yields (45.3). Otherwise, $\mathbf{y} \neq \mathbf{y}_\pm$ has at least two unequal coordinates, say $y_j \neq y_k$, and satisfies (45.7), hence

$$h_{2r-1}(\mathbf{y}, y_j) - h_{2r-1}(\mathbf{y}, y_k) = 2rh_{2r}(\mathbf{y})(y_j - y_k).$$

Rewriting this and using Lemma 45.4,

$$\begin{aligned} h_{2r}(\mathbf{y}) &= \frac{1}{2r}\frac{h_{2r-1}(\mathbf{y}, y_j) - h_{2r-1}(\mathbf{y}, y_k)}{y_j - y_k} = \frac{1}{2r}h_{2r-2}(\mathbf{y}, y_j, y_k) \\ &\geq \frac{1}{2r}\frac{\|\mathbf{y}\|^2 + |y_j|^2 + |y_k|^2}{2^{r-1}(r-1)!}, \end{aligned}$$

where the final inequality follows from the induction hypothesis. Now since $y_j \neq y_k$, the final numerator is strictly greater than 1, and this yields (45.3). $\square$

Theorem 45.2 allows us to prove the existence of polynomials with negative coefficients that entrywise preserve positivity in a fixed dimension. This is discussed presently; we first show for completeness that the polynomials h_{2r} are the only ones that vanish only at the origin.

Proof of Theorem 45.1. That (4) $\implies$ (1) follows directly from Theorem 45.2, and that (1) $\implies$ (2) $\implies$ (3) is immediate. Now suppose (3) holds. Using Littlewood's definition (42.2), if a tableau T of shape $\mathbf{n} - \mathbf{n}_{\min}$ has two nonempty rows, then in any semistandard filling of T, one is forced to use at least two different variables. Now evaluating the weight of T at $\mathbf{e}_1$ yields zero. This argument shows that (3) implies $\mathbf{n} - \mathbf{n}_{\min}$ has at most one row, so by (42.2), $s_{\mathbf{n}}(\mathbf{u}) = h_k(\mathbf{u})$ for some $k \geq 0$. Now $h_k(\mathbf{e}_1 - \mathbf{e}_2)$ is easily evaluated to be a geometric series (consisting of $k+1$ alternating entries 1 and -1). This vanishes if k is odd, so (3) implies k is even, proving (4). $\square$

45.2. **Application: entrywise polynomials preserving positivity.** With the above results at hand, we now prove:

Theorem 45.8. *Fix integers $N \geq 1$, $k, r \geq 0$, and $M \geq N + 2r$, as well as positive constants $\rho, c_0, \dots, c_{N-1}$. There exists a positive constant $t_0 > 0$ such that the polynomial*

$$p_t(x) := tx^k(c_0 + c_1 x + \cdots + c_{N-2}x^{N-2} + c_{N-1}x^{N-1+2r}) - x^{k+M}$$

entrywise preserves positivity on $\mathbb{P}_N([-\rho, \rho])$ whenever $t \geq t_0$.

Proof. The result for $k = 0$ implies that for arbitrary $k \geq 0$, by the Schur product theorem. Thus, we henceforth assume $k = 0$. We now prove the result by induction on $N \geq 1$ with the $N = 1$ case left to the reader as an exercise.

For the induction step, notice that the proof of Proposition 42.8 goes through for $\mathbf{u} \in \mathbb{R}^{N,\neq}$ as long as $s_{\mathbf{n}}(\mathbf{u}) \neq 0$. This is indeed the case if $\mathbf{n} = (0, 1, \dots, N-2, N-1+2r)$, by Theorem 45.2. Thus, to produce a threshold t_1 as in the theorem, which works for all rank-1 matrices, it suffices to show (by the discussion prior to Theorem 43.1, and using the density of $(-\sqrt{\rho}, \sqrt{\rho})^{N,\neq}$ in $[-\sqrt{\rho}, \sqrt{\rho}]^N$) that

$$\sup_{\mathbf{u} \in (-\sqrt{\rho}, \sqrt{\rho})^{N,\neq}} \sum_{j=0}^{N-1} \frac{s_{\mathbf{n}_j}(\mathbf{u})^2}{h_{2r}(\mathbf{u})^2} \frac{\rho^{M-n_j}}{c_{n_j}} < \infty.$$

In turn, using Theorem 45.2, it suffices to show

$$\sup_{\mathbf{u} \in (-\sqrt{\rho}, \sqrt{\rho})^{N,\neq}} \frac{s_{\mathbf{n}_j}(\mathbf{u})^2}{\|\mathbf{u}\|^{4r}} < \infty, \qquad j = 0, 1, \dots, N-1.$$

Now since the polynomial $s_{\mathbf{n}_j}$ is homogeneous of total degree $2r + M - n_j$,

$$\frac{s_{\mathbf{n}_j}(\mathbf{u})^2}{\|\mathbf{u}\|^{4r}} = s_{\mathbf{n}_j}(\mathbf{u}/\|\mathbf{u}\|)^2 \|\mathbf{u}\|^{2(M-n_j)} \leq K_{\mathbf{n}_j}^2 (N\rho)^{M-n_j}$$

for $\mathbf{u} \in (-\sqrt{\rho}, \sqrt{\rho})^{N,\neq}$, where $K_{\mathbf{n}_j}$ is the maximum of the Schur polynomial $s_{\mathbf{n}_j}(\cdot)$ on the unit sphere S^{N-1}.

This shows the existence of a threshold t_1 that proves the theorem for all rank-1 matrices in $\mathbb{P}_N([-\rho, \rho])$. We will prove the result for all matrices in $\mathbb{P}_N([-\rho, \rho])$ by applying the extension principle (Theorem 9.12); for this, we first note that

$$M^{-1} p_t'(x) = t \left(\sum_{j=1}^{N-2} \frac{j c_j}{M} x^{j-1} + \frac{(N-1+2r) c_{N-1}}{M} x^{N-2+2r} \right) - x^{M-1}$$

is again of the same form as in the theorem. Hence, by the induction hypothesis, there exists a threshold t_2 such that $p_t'[-]$ preserves positivity on $\mathbb{P}_{N-1}([-\rho, \rho])$ for $t \geq t_2$. The induction step is now complete by taking $t_0 := \max(t_1, t_2)$. $\square$

A natural question that remains, in parallel to the study of polynomial positivity preservers of matrices in $\mathbb{P}_N([0, \rho])$, is as follows:

Question 45.9. Given the data as in the preceding theorem, find the sharp constant t_0.

A first step toward this goal is the related question in rank 1, which can essentially be rephrased as follows:

Question 45.10. Given integers $r \geq 0$, $N \geq 1$, and

$$m_0 \geq 0, \quad m_1 \geq 1, \quad \dots, \quad m_{N-2} \geq N-2, \quad m_{N-1} \geq N-1+2r,$$

maximize the ratio $\dfrac{s_{\mathbf{m}}(\mathbf{u})^2}{h_{2r}(\mathbf{u})^2}$ over the punctured unit cube $[-1, 1]^N \setminus \{\mathbf{0}\}$.

Remark 45.11. Notice by homogeneity that this ratio of squares increases as one travels radially from the origin. Thus, the maximization on the punctured solid cube is equivalent to the same question on the boundary of this cube.

45.3. **Matrices with complex entries.** The final topic along this theme is to explore matrices with complex entries, say in the open disk $D(0,\rho)$ (or its closure) for some $0 < \rho < \infty$. In this case, the set of admissible "initial sequences of powers" $0 \leq n_0 < \cdots < n_{N-1}$ turns out to be far more limited – and (the same) tight threshold bound is available in all such cases:

Theorem 45.12. *Fix integers $M \geq N \geq 2$ and $k \geq 0$, and let $n_j = j + k$ for $0 \leq j \leq N-1$ – i.e., N consecutive integers. Also fix real scalars $\rho > 0$, $c_0, \dots, c_{N-1}$, and define*

$$f(z) := z^k(c_0 + c_1 z + \cdots + c_{N-1} z^{N-1}) + c_M z^{k+M}, \qquad z \in \mathbb{C}.$$

Then the following are equivalent:

1. *The entrywise map $f[-]$ preserves positivity on $\mathbb{P}_N(D(0,\rho))$.*
2. *The map $f[-]$ preserves positivity on rank-1 totally non-negative (TN) Hankel matrices in $\mathbb{P}_N((0,\rho))$.*
3. *Either all $c_j, c_M \geq 0$; or $c_j > 0$ for all $j < N$ and $c_M \geq -\mathcal{C}^{-1}$, where*

$$\mathcal{C} = \sum_{j=0}^{N-1} \frac{V(\mathbf{n}_j)^2}{c_j V(\mathbf{n}_{\min})^2} \rho^{M-j},$$

where $\mathbf{n}_{\min} := (0,1,\dots,N-1)$ and $\mathbf{n}_j := (0,1,\dots,j-1,j+1,\dots,N-1,M)$ for $0 \leq j \leq N-1$.

A chronological remark: this result was the first instance of entrywise polynomial positivity preservers with negative coefficients to be discovered, in 2016. The more refined and challenging sharp bound for arbitrary polynomials (or tuples of real powers $\mathbf{n}$) operating on $\mathbb{P}_N((0,\rho))$, as well as the existence of a tight threshold for the leading term of a polynomial preserver operating on $\mathbb{P}_N((-\rho,\rho))$, were worked out later – though in this text, we have already proved those results.

Remark 45.13. After proving Theorem 45.12, we will also show that if the initial sequence $\mathbf{n}$ of non-negative integer powers is non-consecutive (i.e., not of the form in Theorem 45.12), then such a "structured" result does not hold for infinitely many powers $M > n_{N-1}$.

Proof of Theorem 45.12. Clearly, (1) $\implies$ (2). Next, notice that the constant $\mathcal{C}$ in (3) remains unchanged under a simultaneous shift of all exponents by the same amount k. Thus, (2) $\implies$ (3) by Theorem 44.1 (and Lemma 41.2).

It remains to show that (3) $\implies$ (1). Since the $k \geq 0$ case follows from the $k = 0$ case of (1) by the Schur product theorem, we assume henceforth that $k = 0$. Now we proceed as in the previous two sections, by first showing the result for rank-1 matrices, and then using an analogue of the extension principle (Theorem 9.12) to extend to all ranks via induction on N. The first step here involves extending Lemma 42.10 to complex matrices:

Lemma 45.14. *Fix $\mathbf{w} \in \mathbb{C}^N$ and a positive definite (Hermitian) matrix $H \in \mathbb{C}^{N\times N}$. Define the linear pencil $P_t := tH - \mathbf{w}\mathbf{w}^*$, for $t > 0$. Then the following are equivalent:*

1. *P_t is positive semidefinite.*
2. *$\det P_t \geq 0$.*
3. *$t \geq \mathbf{w}^* H^{-1} \mathbf{w} = 1 - \dfrac{\det(H - \mathbf{w}\mathbf{w}^*)}{\det H}$.*

The proof is virtually identical to that of Lemma 42.10, and is hence omitted.

Next, using that $s_{\mathbf{n}}(\mathbf{u}^*) = \overline{s_{\mathbf{n}}(\mathbf{u})}$ for all integer tuples $\mathbf{n}$ and all vectors $\mathbf{u} \in \mathbb{C}^N$ (and $s_{\mathbf{n}_{\min}}(\mathbf{u}) \equiv 1$), we apply Lemma 45.14 to computing the sharp threshold bound for a single "generic" rank-1 complex matrix, parallel to how Proposition 42.8 is an adaptation of Lemma 42.10:

Proposition 45.15. *With the given positive scalars c_j, and integers $M \geq N \geq 2$ and $n_j = j - 1$, define*

$$p_t(z) := t \sum_{j=0}^{N-1} c_j z^j - z^M, \qquad t \in (0,\infty),\ z \in \mathbb{C}.$$

Then the following are equivalent for $\mathbf{u} \in \mathbb{C}^{N,\neq}$:

(1) $p_t[\mathbf{uu}^*]$ is positive semidefinite.

(2) $\det p_t[\mathbf{uu}^*] \geq 0$.

(3) $t \geq \displaystyle\sum_{j=0}^{N-1} \frac{|s_{\mathbf{n}_j}(\mathbf{u})|^2}{c_j}$.

Once again, the proof is omitted.

We continue to repeat the approach for $\mathbb{P}_N((0,\rho))$ in previous sections. By the discussion prior to Theorem 43.1, and using the density of $D(0,\sqrt{\rho})^{N,\neq}$ in $\overline{D(0,\sqrt{\rho})}^N$, we next compute

$$\sup_{\mathbf{u} \in D(0,\sqrt{\rho})^{N,\neq}} |s_{\mathbf{n}_j}(\mathbf{u})^2|, \qquad 0 \leq j \leq N-1.$$

Use Littlewood's definition (42.2) of $s_{\mathbf{n}}(\cdot)$, and the triangle inequality, to conclude that

$$|s_{\mathbf{n}}(\mathbf{u})| = \left| \sum_T \prod_{j=1}^N u_j^{f_j(T)} \right| \leq \sum_T \prod_{j=1}^N |u_j|^{f_j(T)} = s_{\mathbf{n}}(|\mathbf{u}|), \quad \text{where } |\mathbf{u}| := (|u_1|, \dots, |u_N|).$$

Thus, equality is indeed attained here if one works with a vector $\mathbf{u} \in (0,\sqrt{\rho})^{N,\neq}$. For this reason, and since $s_{\mathbf{n}}(\mathbf{u})$ is coordinatewise non-decreasing on $(0,\infty)^N$,

$$\sup_{\mathbf{u} \in D(0,\sqrt{\rho})^{N,\neq}} |s_{\mathbf{n}_j}(\mathbf{u})^2| = s_{\mathbf{n}_j}(\sqrt{\rho}(1,\dots,1))^2 = \frac{V(\mathbf{n}_j)^2}{V(\mathbf{n}_{\min})^2} \rho^{M-j}, \qquad \forall j.$$

Akin to $\mathbb{P}_N((0,\rho))$, we conclude that $p_t[\mathbf{uu}^*] \in \mathbb{P}_N$ for all $\mathbf{u} \in D(0,\sqrt{\rho})^N$, if and only if

$$t \geq \mathcal{C} = \sum_{j=0}^{N-1} \frac{V(\mathbf{n}_j)^2}{c_j V(\mathbf{n}_{\min})^2} \rho^{M-j}.$$

The final step is to prove the result for all matrices in $\mathbb{P}_N(D(0,\rho))$, not just those of rank 1. For this we work by induction on $N \geq 1$, with the base case following from above. For the induction step, we will apply the extension principle (Theorem 9.12); to do so, we first extend that result as follows, with essentially the same proof.

Lemma 45.16. *Theorem 9.12 holds if $h(z)$ is a polynomial and $I = D(0,\rho)$ or its closure.*

To apply this result, first note that

$$M^{-1} p_t'(z) = t \sum_{j=1}^{N-1} M^{-1} j c_j z^{j-1} - z^{M-1},$$

and so it suffices to show that this preserves positivity on $\mathbb{P}_{N-1}(D(0,\rho))$ if $t \geq \mathcal{C}$. By the induction hypothesis, it suffices to show that $\mathcal{C} \geq \mathcal{C}'$, where $\mathcal{C}'$ is the constant obtained from $M^{-1}p'_t$:

$$\mathcal{C}' = \sum_{j=1}^{N-1} \frac{MV(\mathbf{n}'_j)^2}{jc_j V(\mathbf{n}'_{\min})^2}\rho^{M-j}, \qquad \text{where } \mathbf{n}'_j := (0,1,\dots,j-2,j,\dots,N-2,M-1),$$

and $\mathbf{n}'_{\min} := (0,1,\dots,N-2)$. Thus, to show that $\mathcal{C} \geq \mathcal{C}'$, it suffices to show that

$$\frac{V(\mathbf{n}_j)^2}{V(\mathbf{n}_{\min})^2} \geq \frac{MV(\mathbf{n}'_j)^2}{jV(\mathbf{n}'_{\min})^2}$$

for $j = 1,\dots,N-1$. This is not hard to show; e.g., for "most" cases of j, a straightforward computation yields

$$\left(\frac{V(\mathbf{n}_j)/V(\mathbf{n}'_j)}{V(\mathbf{n}_{\min})/V(\mathbf{n}'_{\min})}\right)^2 = \left(\frac{(N-1)!M/j}{(N-1)!}\right)^2 = \frac{M^2}{j^2} > \frac{M}{j}. \qquad \square$$

As promised above, we conclude by showing that for every other tuple of "initial powers", i.e., non-consecutive powers $\mathbf{n}$, one cannot always have a positivity preserver with a negative coefficient – even on generic one-parameter families of rank-1 matrices.

Theorem 45.17. *Fix integers $N \geq 2$ and $0 \leq n_0 < \cdots < n_{N-1}$, where the n_j are not all consecutive. Also fix $N-1$ distinct numbers $u_1,\dots,u_{N-1} > 0$, and set*

$$\mathbf{u}(z) := (u_1,\dots,u_{N-1},z)^T \in \mathbb{C}^N, \qquad z \in \mathbb{C}.$$

Then there exists $z_0 \in \mathbb{C}$ and infinitely many integers $M > n_{N-1}$, such that for all choices of (a) scalar $\epsilon > 0$ and (b) coefficients $c_{n_0},\dots,c_{n_{N-1}} > 0 > c' \in \mathbb{R}$, the polynomial

$$f(z) := c_{n_0}z^{n_0} + \cdots + c_{n_{N-1}}z^{n_{N-1}} + c'z^M$$

does not preserve positivity on the rank-1 matrix $\epsilon\mathbf{u}(z_0)\mathbf{u}(z_0)^*$ when applied entrywise.

Note that if instead all $c_{n_j}, c_M \geq 0$, then $f[-]$ preserves positivity by the Schur product theorem; while if some $c_{n_j} < 0$, then the FitzGerald–Horn argument from Theorem 9.3 can be adapted to show that $f[\epsilon\mathbf{u}(u_N)\mathbf{u}(u_N)^*] \notin \mathbb{P}_N$ for all sufficiently small $\epsilon > 0$, where $u_N \in \mathbb{C}$ is such that the non-zero polynomial $s_{\mathbf{n}}(u_1,\dots,u_{N-1},u_N) \neq 0$.

Proof. Since the n_j are not all consecutive, the tableau-shape corresponding to $\mathbf{n} - \mathbf{n}_{\min}$ has at least one row with two cells. It follows by Littlewood's definition (42.2) that $s_{\mathbf{n}}(\mathbf{u})$ has at least two monomials. Now consider $s_{\mathbf{n}}(\mathbf{u}(z))$ as a function only of z, say $g(z)$. Then $g(z)$ is a polynomial that is not a constant multiple of a monomial, so it has a non-zero complex root $z_0 \in \mathbb{C}^\times$. Notice that z_0 is also not in $(0,\infty)$ because the Schur polynomial evaluated at $(u_1,\dots,u_{N-1},u_N)$ is positive for every $u_N \in (0,\infty)$. Thus, $z_0 \in \mathbb{C} \setminus [0,\infty)$.

By choice of z_0 and Cauchy's definition of $s_{\mathbf{n}}(\mathbf{u}(z_0))$ (see Proposition 42.6),

$$\mathbf{u}(z_0)^{\circ\mathbf{n}} = [\mathbf{u}(z_0)^{\circ n_0}|\cdots|\mathbf{u}(z_0)^{\circ n_{N-1}}]$$

is a singular matrix. That said, this matrix has rank $N-1$ by the properties of generalized Vandermonde determinants (see Theorem 5.1); in fact, every subset of $N-1$ columns here is linearly independent. Let V_0 denote the span of these columns; then the ortho-complement $V_0^\perp \subset \mathbb{C}^N$ is one-dimensional, i.e., there exists unique $\mathbf{v} \in \mathbb{C}^N$ up to rescaling, such that $\mathbf{v}^*\mathbf{u}(z_0)^{\circ n_j} = 0$ for all j.

Now given any N consecutive integers $l+1,\dots,l+N$ with $l \geq n_{N-1}$, we claim there exists an integer $M \in [l+1, l+N]$ such that $\mathbf{v}^*\mathbf{u}(z_0)^{\circ M} \neq 0$. Indeed, the *usual* Vandermonde matrix

$$[\mathbf{u}(z_0)^{\circ(l+1)}|\cdots|\mathbf{u}(z_0)^{\circ(l+N)}]$$

is non-singular (since no coordinate in $\mathbf{u}(z_0)$ is zero), so at least one column $\mathbf{u}(z_0)^{\circ M} \notin V_0$. In particular, $\mathbf{v}^*\mathbf{u}(z_0)^{\circ M} \neq 0$, proving the claim.

Finally, choose arbitrary $\epsilon, c_{n_j} > 0 > c'$ as in the theorem. We then assert that $f[\epsilon\mathbf{u}(z_0)\mathbf{u}(z_0)^*]$, where f is defined using this value of M, is not positive semidefinite. Indeed,

$$\mathbf{v}^* f[\epsilon\mathbf{u}(z_0)\mathbf{u}(z_0)^*]\mathbf{v} = \sum_{j=0}^{N-1} c_{n_j}\epsilon^{n_j}|\mathbf{v}^*\mathbf{u}(z_0)^{\circ n_j}|^2 + c'\epsilon^M|\mathbf{v}^*\mathbf{u}(z_0)^{\circ M}|^2 = c'\epsilon^M|\mathbf{v}^*\mathbf{u}(z_0)^{\circ M}|^2,$$

and this is negative, proving that $f[\epsilon\mathbf{u}(z_0)\mathbf{u}(z_0)^*] \notin \mathbb{P}_N$. $\square$

46. Appendix F: Cauchy's and Littlewood's definitions of Schur polynomials. The (dual) Jacobi–Trudi identity and Lindström–Gessel–Viennot bijections.

For completeness, in this section we show the equivalence of *four* definitions of Schur polynomials, two of which are named identities. To proceed, first recall two other families of symmetric polynomials: the *elementary symmetric polynomials* are simply

$$e_1(u_1, u_2, \dots) := u_1 + u_2 + \cdots, \qquad e_2(u_1, u_2, \dots) := u_1u_2 + u_1u_3 + u_2u_3 + \cdots,$$

and in general,

$$e_k(u_1, u_2, \dots) := \sum_{1 \leq j_1 < j_2 < \cdots < j_k} u_{j_1} u_{j_2} \cdots u_{j_k}$$

These symmetric functions crucially feature while decomposing polynomials into linear factors.

We also recall the *complete homogeneous symmetric polynomials*

$$h_k(u_1, u_2, \dots) := \sum_{1 \leq j_1 \leq j_2 \leq \cdots \leq j_k} u_{j_1} u_{j_2} \cdots u_{j_k}.$$

By convention, we set $e_0 = h_0 = 1$, and $e_k = h_k = 0$ for $k < 0$. Now we have:

Theorem 46.1. *Fix an integer $N \geq 1$ and any unital commutative ground ring. Given a partition of N – i.e., an N-tuple of non-increasing non-negative integers $\lambda = (\lambda_1 \geq \cdots \geq \lambda_N)$ with $\sum_j \lambda_j = N$ – the following four definitions give the same expression $s_{\overline{\lambda+\delta}}(u_1, u_2, \dots, u_N)$, where $\delta := (N-1, N-2, \dots, 0)$ and $\overline{\lambda+\delta} = (\lambda_N, \lambda_{N-1}+1, \dots, \lambda_1+N-1)$ in our convention.*

1. *(Littlewood's definition.) The sum of weights over all column-strict Young tableaux of shape λ with cell entries $u_1, \dots, u_N$.*
2. *(Cauchy's definition, aka the type A Weyl character formula.) The ratio of the (generalized) Vandermonde determinants $a_{\lambda+\delta}/a_\delta$, where $a_\lambda := \det(u_j^{\lambda_k+N-k})$.*
3. *(The Jacobi–Trudi identity.) The determinant $\det(h_{\lambda_j-j+k})_{j,k=1}^N$.*
4. *(The dual Jacobi–Trudi identity, or von Nägelsbach–Kostka identity.) The determinant $\det(e_{\lambda'_j-j+k})$, where λ' is the dual partition, meaning $\lambda'_k := \#\{j : \lambda_j \geq k\}$.*

From this result, we deduce the equivalence of these definitions of the Schur polynomial for fewer numbers of variables $u_1, \dots, u_n$, where $n \leq N$.

Corollary 46.2. *Suppose $1 \leq r < N$ and $\lambda_{r+1} = \cdots = \lambda_N = 0$. Then the four definitions in Theorem 46.1 agree for the smaller set of variables $u_1, \dots, u_r$.*

Proof. Using fewer numbers of variables in definitions (3) and (4) amounts to specializing the remaining variables $u_{r+1}, \dots, u_N$ to zero. The same holds for definition (1) since weights involving the extra variables $u_{r+1}, \dots, u_N$ now get set to zero. It follows that definitions (1), (3), and (4) agree for fewer numbers of variables.

We will show that Cauchy's definition (2) in Theorem 46.1 has the same property. In this case the definitions are different: Given $u_1, \dots, u_r$ for $1 \leq r \leq N$, the corresponding ratio of alternating polynomials would only involve $\lambda_1 \geq \cdots \geq \lambda_r$, and would equal $\det(u_j^{\lambda_k+r-k})_{j,k=1}^r / \det(u_j^{r-k})_{j,k=1}^r$. Now claim that this equals the ratio in (2), by downward induction on $r \leq N$. Note that it suffices to show the claim for $r = N-1$. But here, if we set $u_N := 0$, then both generalized Vandermonde matrices have last column $(0, \dots, 0, 1)^T$. In particular, we may expand along their last columns. Now canceling the common factors of $u_1 \cdots u_{N-1}$ from each of the previous columns reduces to the case of $r = N-1$, and the proof is completed by similarly continuing inductively. $\square$

The remainder of this section is devoted to proving Theorem 46.1. We will show that (4) $\Longleftrightarrow$ (1) $\Longleftrightarrow$ (3) $\Longleftrightarrow$ (2), and over the ground ring $\mathbb{Z}$, which then carries over to arbitrary ground rings. To do so, we use an idea due to Karlin–Macgregor (1959), Lindström (1973), and Gessel–Viennot (1985), which interprets determinants in terms of tuples of weighted lattice paths. The approach below is taken from the work of Bressoud–Wei (1993).

Proposition 46.3. *The definitions (1) and (3) are equivalent.*

Proof. The proof is divided into steps, for ease of exposition.

Step 1: In this step we define the formalism of lattice paths and their weights. Define points in the plane

$$P_k := (N-k+1, N), \quad Q_k := (N-k+1+\lambda_k, 1), \quad k = 1, 2, \dots, N,$$

and consider (ordered) N-tuples $\mathbf{p}$ of (directed) lattice paths satisfying the following properties:

1. The kth path starts at some P_j and ends at Q_k, for each k.
2. No two paths start at the same point P_j.
3. From P_j, and at each point (a, b), a path can go either east or south. Weight each east step at height (a, b) by u_{N+1-b}.

Notice that one can assign a unique permutation $\sigma = \sigma_{\mathbf{p}} \in S_N$ to each tuple of paths $\mathbf{p}$, so that paths go from $P_{\sigma(k)}$ to Q_k for each k.

We now assign a weight to each tuple $\mathbf{p}$, defined to be $(-1)^{\sigma_{\mathbf{p}}}$ times the product of the weights at all east steps in $\mathbf{p}$. For instance, if $\lambda = (3, 1, 1, 0, 0)$ partitions $N = 5$, then here is a typical tuple of paths:

- For $k = 4, 5$, P_k and Q_k are each connected by vertical straight lines (i.e., four south steps each).
- P_2 and Q_3 are connected by a vertical straight line (i.e., four south steps).
- The steps from P_3 to Q_2 are $SESESS$.
- The steps from P_1 to Q_1 are $SEESSES$.

This tuple $\mathbf{p}$ corresponds to the permutation $\sigma_{\mathbf{p}} = (13245)$, and has weight $-u_2^3 u_3 u_4$.

Step 2: The next goal is to examine the *generating function* of the tuples, i.e., $\sum_{\mathbf{p}} \mathrm{wt}(\mathbf{p})$. Note that given σ, among all tuples $\mathbf{p}$ with $\sigma_{\mathbf{p}} = \sigma$, the kth path contributes a monomial of total degree $\lambda_k - k + \sigma(k)$, which can be any monomial in $u_1, \dots, u_N$ of this total degree. It follows that the generating function equals

$$\sum_{\mathbf{p}} \mathrm{wt}(\mathbf{p}) = \sum_{\sigma \in S_N} (-1)^\sigma \prod_{k=1}^N h_{\lambda_k - k + \sigma(k)} = \det(h_{\lambda_k - k + j})_{j,k=1}^N.$$

Step 3: We next rewrite the above generating function to obtain $\sum_T \mathrm{wt}(T)$ (the sum of weights over all column-strict Young tableaux of shape λ with cell entries $u_1, \dots, u_N$), which is precisely $s_{\overline{\lambda+\delta}}(u_1, \dots, u_N)$ by definition. To do so, we will pair off the tuples $\mathbf{p}$ of *intersecting* paths into pairs, whose weights cancel one another.

Suppose $\mathbf{p}$ consists of intersecting paths. Define the *final intersection point* of $\mathbf{p}$ to be the lattice point with maximum x-coordinate where at least two paths intersect, and if there are more than one such points, then the one with minimal y-coordinate. Now claim that exactly two paths in $\mathbf{p}$ intersect at this point. Indeed, if three paths intersect at any point, then all of them have to go either east or south at the next step. By the pigeonhole principle, there are at least two paths that proceed in the same direction. It follows that a point common to three paths in $\mathbf{p}$ cannot be the final intersection point, as desired.

Define the *tail* of $\mathbf{p}$ to be the two paths to the east and south of the final intersection point in $\mathbf{p}$. Given an intersecting tuple of paths $\mathbf{p}$, there exists a unique other tuple $\mathbf{p}'$ with the same final intersection point between the same two paths, but with the tails swapped. It is easy to see that the paths $\mathbf{p}$ and $\mathbf{p}'$ satisfy have opposite signs (for their permutations $\sigma_{\mathbf{p}}, \sigma_{\mathbf{p}'}$), but the same monomials in their weights. Therefore, $\mathrm{wt}(\mathbf{p}) = -\mathrm{wt}(\mathbf{p}')$, and the intersecting paths pair off, as desired.

Step 4: From Step 3, the generating function $\sum_{\mathbf{p}} \mathrm{wt}(\mathbf{p})$ equals the sum over only tuples of non-intersecting paths. Each of these tuples necessarily has $\sigma_{\mathbf{p}} = \mathrm{id}$, so all signs are positive. In such a tuple, the monomial weight for the kth path naturally corresponds to a weakly increasing sequence of λ_k integers in $[1, N]$. That the paths do not intersect corresponds to the entries in the kth sequence being strictly smaller than the corresponding entries in the $(k+1)$st sequence. This yields a natural weight-preserving bijection from the tuples of non-intersecting paths to the column-strict Young tableaux of shape λ with cell entries $1, \dots, N$. (Notice that these tableaux are in direct bijection to the column-strict Young tableaux studied earlier in this part, by switching the cell entries $j \longleftrightarrow N+1-j$.) This concludes the proof. $\square$

Proposition 46.4. *The definitions (1) and (4) are equivalent.*

Proof. The proof is a variant of that of Proposition 46.3. Now we consider all tuples of paths such that the kth path goes from $P_{\sigma(k)}$, to the point

$$Q'_k := (N - k + 1 + \lambda'_k, 1),$$

and, moreover, each of these paths has at most one east step at each fixed height – i.e., no two east steps are consecutive.

Once again, in summing to obtain the generating function, given a permutation $\sigma = \sigma_{\mathbf{p}}$, the kth path in $\mathbf{p}$ contributes a monomial of total degree $\lambda'_k - k + \sigma(k)$, but now runs over all monomials with individual variables of degree at most 1 – i.e., all monomials in $e_{\lambda_k - k + \sigma(k)}$. It follows that

$$\sum_{\mathbf{p}} \mathrm{wt}(\mathbf{p}) = \sum_{\sigma \in S_N} (-1)^\sigma \prod_{k=1}^N e_{\lambda'_k - k + \sigma(k)} = \det(e_{\lambda'_k - k + j})_{j,k=1}^N.$$

On the other side, we once again pair off tuples – this time, leaving the ones that do not *overlap*. In other words, paths in tuples may intersect at a point, but do not share an east/south line segment. Now given a tuple containing two overlapping paths, define the *final overlap segment* similarly as in Proposition 46.3; as in the previous proof, notice that exactly two paths overlap on this segment. Then for every tuple of paths $\mathbf{p}$ that overlaps, there exists a unique other tuple $\mathbf{p}'$ with the same final overlap segment between the same two paths, but with the (new version of) tails swapped. It is easy to see that $\mathbf{p}$ and $\mathbf{p}'$ have the same monomials as weights, but with opposite signs, so they pair off and cancel weights.

This leaves us with tuples of non-overlapping paths, all of which again corresponding to $\sigma_{\mathbf{p}} = \mathrm{id}$. In such a tuple, from the kth path we obtain a strictly increasing sequence of λ'_k integers in $[1, N]$. That the paths do not overlap corresponds to the entries in the kth sequence being at most as large as the corresponding entries in the $(k+1)$st sequence. This gives a bijection to the *conjugates* of column-strict Young tableaux of shape λ, and hence we once again have $\sum_{\mathbf{p}} \mathrm{wt}(\mathbf{p}) = \sum_T \mathrm{wt}(T)$ in this setting. $\square$

Corollary 46.5. *Schur polynomials are symmetric and homogeneous.*

Proof. This follows because definition (4) is symmetric and homogeneous in the variables u_j. $\square$

Finally, we show:

Proposition 46.6. *The definitions (2) and (3) are equivalent.*

Proof. Once again, this proof is split into steps, for ease of exposition. In the proof below, we use the above results and assume that the definitions (1), (3), and (4) are all equivalent. Thus, our goal is to show that

$$\det(u_j^{N-k})_{j,k=1}^N \cdot \det(h_{\lambda_j - j + k})_{j,k=1}^N = \det(u_j^{\lambda_k + N - k})_{j,k=1}^N.$$

Step 1: We explain the formalism, which is a refinement of the one in the proof of Proposition 46.3. Thus, we return to the setting of paths between $P_k = (N - k + 1, N)$ and $Q_k = (N - k + 1 + \lambda_k, 1)$ for $k = 1, \dots, N$, but now *equipped also with a permutation* $\tau \in S_N$. The weight of an east step now depends on its height: at height $N + 1 - b$ an east step has weight $u_{\tau(b)}$ instead of u_b. Now consider tuples of paths over all τ; let us write their weights as $\mathrm{wt}_\tau(\mathbf{p})$ for notational clarity. In what follows, we also use $\mathbf{p}$ or $(\mathbf{p}, \tau)$ depending on the need to specify and work with $\tau \in S_N$.

For each fixed $\tau \in S_N$, notice first that the generating function $\sum_{\mathbf{p}} \mathrm{wt}_\tau(\mathbf{p})$ of the τ-permuted paths is independent of τ, by Corollary 46.5.

Now we define a new weight for these τ-permuted paths $\mathbf{p}$. Namely, given $\mathbf{p} = (\mathbf{p}, \tau)$, recall there exists a unique permutation $\sigma_{\mathbf{p}} \in S_N$; now define

$$\mathrm{wt}'_\tau(\mathbf{p}) := (-1)^\tau \mu(\tau) \cdot \mathrm{wt}_\tau(\mathbf{p}), \qquad \text{where} \qquad \mu(\tau) := u_{\tau(1)}^{N-1} u_{\tau(2)}^{N-2} \cdots u_{\tau(N-1)}.$$

The new generating function is

$$\sum_{\tau \in S_N} \sum_{\mathbf{p}} \mathrm{wt}'_\tau(\mathbf{p}) = \sum_{\tau \in S_N} (-1)^\tau \mu(\tau) \sum_{\mathbf{p}} \mathrm{wt}_\tau(\mathbf{p}) = \det(h_{\lambda_k - k + j})_{j,k=1}^N \cdot \det(u_j^{N-k})_{j,k=1}^N,$$

where the final equality follows from the above propositions, given that the inner sum is independent of τ from above.

Step 2: Say that a tuple $\mathbf{p} = (P_{\sigma_{\mathbf{p}}(k)} \to Q_k)_k$ is *high enough* if for every $1 \le k \le N$, the kth path has no east steps below height $N + 1 - k$. Now **claim** that (summing over all $\tau \in S_N$,) the τ-tuples that are not high enough once again pair up, with canceling weights.

Modulo the claim, we prove the theorem. The first reduction is that for a fixed τ, we may further restrict to the τ-tuples that are high enough *and* are non-intersecting (as in the proof of Proposition 46.3). Indeed, defining the final intersection point and the tail of $\mathbf{p}$ as in that proof, it follows that switching tails in tuples $\mathbf{p}$ of intersecting paths changes neither the monomial part of the weight, nor the high-enough property; and it induces the opposite sign to that of $\mathbf{p}$.

Thus, the generating function of all τ-tuples (over all τ) equals that of all non-intersecting, high-enough τ-tuples (also summed over all $\tau \in S_N$). But each such tuple corresponds to $\sigma_{\mathbf{p}} = \mathrm{id}$, and in it, all east steps in the first path must occur in the topmost row/height/y-coordinate of N. Hence, all east steps in the second path must occur in the next highest row, and so on. It follows that the non-intersecting, high-enough τ-tuples $\mathbf{p} = (\mathbf{p}, \tau)$ are in bijection with $\tau \in S_N$; moreover, each such tuple has weight $(-1)^\tau \mu(\tau) u_{\tau(1)}^{\lambda_1} u_{\tau(2)}^{\lambda_2} \cdots u_{\tau(N)}^{\lambda_N}$. Thus, the above generating function is shown to equal

$$\det(u_j^{\lambda_k+N-k})_{j,k=1}^N,$$

and the proof is complete.

Step 3: It thus remains to show the claim in Step 2 above. Given parameters

$$\sigma \in S_N, \quad k \in [1,N], \quad j \in [1,N-k],$$

let $NH_{\sigma,k,j}$ denote the τ-tuples of paths $\mathbf{p} = (\mathbf{p},\tau)$ (with τ running over S_N), which satisfy the following properties:

(1) $\mathbf{p}$ is not high (NH) enough.
(2) In $\mathbf{p}$, the kth path has an east step at most by height $N-k$, but the paths labeled $1,\dots,k-1$ are all high enough.
(3) Moreover, j is the height of the lowest east step in the kth path; thus, $j \in [1,N-k]$.
(4) The permutation associated to the start and end points of the paths in the tuple is $\sigma_{\mathbf{p}} = \sigma \in S_N$.

Note that the set NH of tuples of paths that are not high enough can be partitioned as

$$NH = \bigsqcup_{\sigma \in S_N,\ k\in[1,N],\ j\in[1,N-k]} NH_{\sigma,k,j}.$$

We now construct an involution of sets $\iota : NH \to NH$ which permutes each subset $NH_{\sigma,k,j}$, and such that $\mathbf{p}$ and $\iota(\mathbf{p})$ have the same monomial attached to them but different τ and τ', leading to canceling signs $(-1)^\tau \neq (-1)^{\tau'}$.

Thus, suppose $\mathbf{p}$ is a τ-tuple in $NH_{\sigma,k,j}$. Now define $\tau' := \tau \circ (N-j, N+1-j)$; in other words,

$$\tau'(i) := \begin{cases} \tau(i+1), & \text{if } i = N-j; \\ \tau(i-1), & \text{if } i = N-j+1; \\ \tau(i), & \text{otherwise.} \end{cases}$$

In particular,

$$(-1)^{\tau'} = -(-1)^\tau \qquad \text{and} \qquad \mu(\tau') = \mu(\tau) u_{\tau(N+1-j)} u_{\tau(N-j)}^{-1}.$$

With τ' in hand, we can define the tuple $\iota(\mathbf{p}) = (\iota(\mathbf{p}), \tau') \in NH_{\sigma,k,j}$. First, change the weight of each east step at height $N+1-b$, from $u_{\tau(b)}$ to $u_{\tau'(b)}$. Next, we keep unchanged the paths labeled $1,\dots,k-1$, and in the remaining paths we do not change the source and target nodes either (since σ is fixed). Notice that weights change at only two heights j and $j+1$; hence the first $k-1$ paths do not see any weights change.

The changes in the (other) paths are now described. In the kth path, change only the numbers n_l of east steps at height $l = j, j+1$, via: $(n_j, n_{j+1}) \mapsto (n_{j+1}+1, n_j-1)$. Note, the product of weights of all east steps in this path changes by a multiplicative factor of $u_{\tau(N+1-j)}^{-1} u_{\tau(N-j)}$ – which cancels the above change from $\mu(\tau)$ to $\mu(\tau')$. Finally, in the mth path for each $m > k$, if n_l again denotes the number of east steps at height l, then we swap $n_j \longleftrightarrow n_{j+1}$ steps in the mth path. This leaves unchanged the weight of those paths, and hence of the tuple $\mathbf{p}$ overall.

It is now straightforward to verify that the map ι is an involution that preserves each of the sets $NH_{\sigma,k,j}$. Since $\mathrm{wt}(\iota(\mathbf{p})) = -\mathrm{wt}(\mathbf{p})$ for all $\mathbf{p} \in NH$, the claim in Step 2 is true, and the proof of the theorem is complete. □

Bibliographic notes and references

Most of the material in this part is taken from two papers: one by Belton–Guillot–Khare–Putinar [23] (see also [24] and [25] for summaries), and the other by Khare and Tao [217] (see also its summary [216]). We list the remaining references. For preliminaries on Schur polynomials, the standard reference is Macdonald's monograph [247]. Theorem 44.6 on the coordinatewise monotonicity of Schur polynomial ratios is proved using a deep result of Lam, Postnikov, and Pylyavskyy [231], following previous work by Skandera [339]. There are other ways to show this result, e.g., using Chebyshev blossoming as shown by Ait-Haddou in joint works [6, 7], or by Dodgson condensation (see [217]). Theorem 44.8 is taken from [217] (the coordinatewise non-decreasing property on $(0,\infty)^N_{\neq}$). We also remark that Equation (42.11), like Proposition 17.5, was recently extended to arbitrary polynomials and (formal) power series by the author in [214].

Theorem 45.2 and its proof are due to Hunter [187]. Appendix F follows Bressoud and Wei [71], relying on the works of Karlin and McGregor [201, 202], Lindström [238], and Gessel and Viennot [142].

References

[1] Ivo Adan, Michael van Eenige, and Jacques Resing. Fitting discrete distributions on the first two moments. *Probab. Engrg. Inform. Sci.*, 9(4):623–632, 1995.

[2] Jim Agler and John Edward McCarthy. What can Hilbert spaces tell us about bounded functions in the bidisk? In: *A Glimpse at Hilbert Space Operators*, Oper. Theory Adv. Appl., Vol. 207 (S. Axler, P. Rosenthal, D. Sarason, Eds.), pp. 81–97, 2010.

[3] Jim Agler, John Edward McCarthy, and Nicholas John Young. *Operator Analysis: Hilbert space methods in complex analysis.* Cambridge Tracts in Mathematics 219, Cambridge University Press, Cambridge, 2020.

[4] Michael Aissen, Albert Edrei, Isaac Jacob Schoenberg, and Anne M. Whitney. On the generating functions of totally positive sequences. *Proc. Natl. Acad. Sci. USA*, 37(5): 303–307, 1951.

[5] Michael Aissen, Isaac Jacob Schoenberg, and Anne M. Whitney. On the generating functions of totally positive sequences I. *J. d'Analyse Math.*, 2:93–103, 1952.

[6] Rachid Ait-Haddou and Marie-Laurence Mazure. The fundamental blossoming inequality in Chebyshev spaces–I: Applications to Schur functions. *Found. Comput. Math.*, 18(1):135–158, 2018.

[7] Rachid Ait-Haddou, Yusuke Sakane, and Taishin Nomura. Chebyshev blossoming in Müntz spaces: Toward shaping with Young diagrams. *J. Comput. Appl. Math.*, 247:172–208, 2013.

[8] Naum Ilyich Akhiezer. *The classical moment problem and some related questions in analysis.* Translated by N. Kemmer. Hafner Publishing Co., New York, 1965.

[9] Arthur Albert. Conditions for positive and nonnegative definiteness in terms of pseudoinverses. *SIAM J. Appl. Math.*, 17(2), 434–440, 1969.

[10] Alexandru Aleman, Dmitry Beliaev, and Haakan Hedenmalm. Real zero polynomials and Pólya–Schur type theorems. *J. d'Analyse Math.*, 94:49–60, 2004.

[11] Suprasad V. Amari and Ravindra Babu Misra. Closed-form expressions for distribution of sum of exponential random variables. *IEEE Trans. Reliab.*, 46(4):519–522, 1997.

[12] Tsuyoshi Ando. Totally positive matrices. *Linear Algebra Appl.*, 90:165–219, 1987.

[13] Konstantin Alekseevich Andreev. Note sur une relation entre les intégrales définies des produits des fonctions. *Mém. Soc. Sci. Phys. Nat. Bordeaux*, 3(2):1–14, 1886.

[14] Bernard A. Asner, Jr. On the total non-negativity of the Hurwitz matrix. *SIAM J. Appl. Math.*, 18(2):407–414, 1970.

[15] Zhi Dong Bai and Li-Xin Zhang. Semicircle law for Hadamard products. *SIAM J. Matrix Anal. Appl.*, 29(2):473–495, 2007.

[16] Mihály Bakonyi and Hugo J. Woerdeman. *Matrix completions, moments, and sums of Hermitian squares.* Princeton Series in Applied Mathematics, Princeton University Press, Princeton, 2011.

[17] Stefan Banach. Sur l'équation fonctionelle $f(x+y) = f(x) + f(y)$. *Fund. Math.*, 1(1):123–124, 1920.

[18] Ravindra B. Bapat and T.E.S. Raghavan. *Nonnegative matrices and applications.* Encyclopedia of Mathematics and its Applications, Vol. 64, Cambridge University Press, Cambridge, 1997.

[19] Victor Simões Barbosa and Valdir Antonio Menegatto. Strict positive definiteness on products of compact two-point homogeneous spaces. *Integral Trans. Spec. Funct.*, 28(1):56–73, 2017.

[20] Alexander I. Barvinok. Low rank approximations of symmetric polynomials and asymptotic counting of contingency tables. *Preprint*, arXiv:math.CO/0503170, 2005.

[21] Richard Keith Beatson and Wolfgang zu Castell. Dimension hopping and families of strictly positive definite zonal basis functions on spheres. *J. Approx. Theory*, 221:22–37, 2017.

[22] Richard Keith Beatson, Wolfgang zu Castell, and Yuan Xu. A Pólya criterion for (strict) positive-definiteness on the sphere. *IMA J. Numer. Anal.*, 34(2):550–568, 2014.

[23] Alexander Belton, Dominique Guillot, Apoorva Khare, and Mihai Putinar. Matrix positivity preservers in fixed dimension. I. *Adv. Math.*, 298:325–368, 2016.

[24] Alexander Belton, Dominique Guillot, Apoorva Khare, and Mihai Putinar. Matrix positivity preservers in fixed dimension. *Comptes Rendus Math.*, 354(2):143–148, 2016.

[25] Alexander Belton, Dominique Guillot, Apoorva Khare, and Mihai Putinar. Schur polynomials and positivity preservers. Extended abstract, FPSAC 2016 Proceedings, *Disc. Math. Theor. Comp. Sc.*, Vol. BC, pp. 155–166, 2016.

[26] Alexander Belton, Dominique Guillot, Apoorva Khare, and Mihai Putinar. Simultaneous kernels of matrix Hadamard powers. *Linear Algebra Appl.*, 576:142–157, 2019.

[27] Alexander Belton, Dominique Guillot, Apoorva Khare, and Mihai Putinar. A panorama of positivity. Part I: Dimension free. In: *Analysis of Operators on Function Spaces* (The Serguei Shimorin Memorial

Volume; A. Aleman, H. Hedenmalm, D. Khavinson, M. Putinar, Eds.), pp. 117–165, Trends in Math., Birkhauser, 2019. Parts 1 and 2 (unified) available at arXiv:math.CA/1812.05482.

[28] Alexander Belton, Dominique Guillot, Apoorva Khare, and Mihai Putinar. A panorama of positivity. Part II: Fixed dimension. In: *Complex Analysis and Spectral Theory*, Proceedings of the CRM Workshop held at Laval University, QC, May 21–25, 2018 (G. Dales, D. Khavinson, J. Mashreghi, Eds.). CRM Proceedings – AMS Contemporary Mathematics 743, pp. 109–150, American Mathematical Society, 2020. Parts 1 and 2 (unified) available at arXiv:math.CA/1812.05482.

[29] Alexander Belton, Dominique Guillot, Apoorva Khare, and Mihai Putinar. Hirschman–Widder densities. *Appl. Comput. Harmon. Anal.*, 60:396–425, 2022.

[30] Alexander Belton, Dominique Guillot, Apoorva Khare, and Mihai Putinar. Moment-sequence transforms. *J. Eur. Math. Soc.*, 24(9):3109–3160, 2022.

[31] Alexander Belton, Dominique Guillot, Apoorva Khare, and Mihai Putinar. Matrix compression along isogenic blocks. *Acta Sci. Math. (Szeged)*, 88(1–2) – *100th anniversary special volume*, 417–448, 2022.

[32] Alexander Belton, Dominique Guillot, Apoorva Khare, and Mihai Putinar. Totally positive kernels, Pólya frequency functions, and their transforms. *J. d'Analyse Math.*, in press; arXiv:math.FA/2006.16213.

[33] Alexander Belton, Dominique Guillot, Apoorva Khare, and Mihai Putinar. *Work in progress*, 2022.

[34] Eduard W. Benz. Über lineare, verschiebungstreue Funktionaloperatoren und die Nullstellen ganzer Funktionen. *Comment. Math. Helv.*, 7:243–289, 1934.

[35] Arkady Berenstein, Sergey Fomin, and Andrei Zelevinsky. Parametrizations of canonical bases and totally positive matrices. *Adv. Math.*, 122:49–149, 1996.

[36] Arkady Berenstein and Andrei Zelevinsky. Total positivity in Schubert varieties. *Comment. Math. Helv.*, 72(1):128–166, 1997.

[37] Arkady Berenstein and Andrei Zelevinsky. Tensor product multiplicities, canonical bases and totally positive varieties. *Invent. Math.*, 143:77–128, 2001.

[38] Felix Alexandrovich Berezin. Quantization in complex symmetric spaces. *Izv. Akad. Nauk SSSR Ser. Mat.*, 39(2):363–402, 1975.

[39] Christian Berg, Jens Peter Reus Christensen, and Paul Ressel. Positive definite functions on abelian semigroups. *Math. Ann.*, 223(3):253–274, 1976.

[40] Christian Berg, Jens Peter Reus Christensen, and Paul Ressel. *Harmonic Analysis on Semigroups: Theory of Positive Definite and Related Functions.* Graduate Texts in Mathematics, Vol. 100, Springer, New York, 1984.

[41] Christian Berg, Ana Paula Peron, and Emilio Porcu. Schoenberg's theorem for real and complex Hilbert spheres revisited. *J. Approx. Theory*, 228:58–78, 2018.

[42] Christian Berg and Emilio Porcu. From Schoenberg coefficients to Schoenberg functions. *Constr. Approx.*, 45:217–241, 2017.

[43] Serge Bernstein. *Reçons sur les Propriétés Extrémales et la Meilleure Approximation des Fonctions Analytiques d'une Variable Réelle.* Borel series of monographs, Gauthier-Villars, Paris, 1926.

[44] Serge Bernstein. Sur les fonctions absolument monotones. *Acta Math.*, 52(1):1–66, 1929.

[45] Rajendra Bhatia. *Matrix analysis.* Graduate Texts in Mathematics, Vol. 169, Springer, New York, 1997.

[46] Rajendra Bhatia. *Positive definite matrices.* Princeton Series in Applied Mathematics, Princeton University Press, Princeton, 2015.

[47] Rajendra Bhatia and Ludwig Elsner. Positivity preserving Hadamard matrix functions. *Positivity*, 11(4):583–588, 2007.

[48] Peter John Bickel and Elizaveta Levina. Covariance regularization by thresholding. *Ann. Statist.*, 36(6):2577–2604, 2008.

[49] Charles Biehler. Sur une classe d'équations algébriques dont toutes les racines sont réelles. *J. reine angew. Math.*, 87:350–352, 1879.

[50] Garrett Birkhoff. Metric foundations of geometry. I. *Trans. Amer. Math. Soc.*, 55:465–492, 1944.

[51] Henry Blumberg. On convex functions. *Trans. Amer. Math. Soc.*, 20:40–44, 1919.

[52] Leonard Mascot Blumenthal. New theorems and methods in determinant theory. *Duke Math. J.*, 2(2):396–404, 1936.

[53] Ralph Philip Boas, Jr. The Stieltjes moment problem for functions of bounded variation. *Bull. Amer. Math. Soc.*, 45:399–404, 1939.

[54] Ralph Philip Boas, Jr. Functions with positive derivatives. *Duke Math. J.*, 8(1):163–172, 1941.

[55] Ralph Philip Boas, Jr. and David Vernon Widder. Functions with positive differences. *Duke Math. J.*, 7:496–503, 1940.

[56] Salomon Bochner. Monotone funktionen, Stieltjessche integrale und harmonische analyse. *Math. Ann.*, 108(1):378–410, 1933.

[57] Salomon Bochner. Hilbert distances and positive definite functions. *Ann. of Math. (2)*, 42:647–656, 1941.

[58] Rafaela Neves Bonfim, Jean Carlo Guella, and Valdir Antonio Menegatto. Strictly positive definite functions on compact two-point homogeneous spaces: the product alternative. *Symm., Integr. Geom.: Meth. Appl. (SIGMA)*, 14:112, 14 pp., 2018.

[59] Carl de Boor. On calculating with B-splines. *J. Approx. Theory*, 6(1):50–62, 1972.

[60] Julius Borcea and Petter Brändén. Applications of stable polynomials to mixed determinants: Johnson's conjectures, unimodality, and symmetrized Fischer products. *Duke Math. J.*, 143(2):205–223, 2008.

[61] Julius Borcea and Petter Brändén. Pólya–Schur master theorems for circular domains and their boundaries. *Ann. of Math. (2)*, 170(1):465–492, 2009.

[62] Julius Borcea and Petter Brändén. The Lee–Yang and Pólya–Schur programs. I. Linear operators preserving stability. *Invent. Math.*, 177(3):541–569, 2009.

[63] Julius Borcea and Petter Brändén. The Lee–Yang and Pólya–Schur programs. II. Theory of stable polynomials and applications. *Comm. Pure Appl. Math.*, 62(12):1595–1631, 2009.

[64] Julius Borcea and Petter Brändén. Multivariate Pólya–Schur classification problems in the Weyl algebra. *Proc. London Math. Soc.*, 101(1):73–104, 2016.

[65] Julius Borcea, Petter Brändén, and Thomas M. Liggett. Negative dependence and the geometry of polynomials. *J. Amer. Math. Soc.*, 22(2):521–567, 2009.

[66] James Borger. Witt vectors, semirings, and total positivity. In: *Absolute Arithmetic and* $\mathbb{F}_1$*-Geometry* (K. Thas, Ed.), pp. 273–331, Eur. Math. Soc., 2016.

[67] James Borger and Darij Grinberg. Boolean Witt vectors and an integral Edrei–Thoma theorem. *Selecta Math. (NS)*, 22:595–629, 2016.

[68] Petter Brändén. On linear transformations preserving the Pólya frequency property. *Trans. Amer. Math. Soc.*, 358(8):3697–3716, 2006.

[69] Francesco Brenti. *Unimodal, log-concave, and Pólya frequency sequences in combinatorics*. Mem. Amer. Math. Soc., Vol. 413, American Mathematical Society, Providence, 1989.

[70] Francesco Brenti. Combinatorics and total positivity. *J. Combin. Theory Ser. A*, 71(2):175–218, 1995.

[71] David Marius Bressoud and Shi-Yuan Wei. Combinatorial equivalence of definitions of the Schur function. *AMS Contemp. Math.*, 143:59–64 (A Tribute to Emil Grosswald: Number Theory and Related Analysis), 1993.

[72] Lawrence D. Brown, Iain M. Johnstone, and K. Brenda Macgibbon. Variation diminishing convolution kernels associated with second-order differential operators. *J. Amer. Statist. Assoc.*, 76(376):824–832, 1981.

[73] Nicolaas Govert de Bruijn. Some theorems on the roots of polynomials. *Nieuw Arch. Wiskunde*, 23:66–68, 1949.

[74] Victor Matveevich Buchstaber and Aleksei Antonovich Glutsyuk. Total positivity, Grassmannian and modified Bessel functions. *AMS Contemp. Math.*, 733:97–108 (Functional Analysis and Geometry: Selim Grigorievich Krein Centennial), 2019.

[75] François Budan de Boislaurent. Nouvelle méthode pour la résolution des équations numériques. Courcier, Paris, 1807.

[76] David A. Cardon, Tamás Forgács, Andrzej Piotrowski, Evan Sorensen, and Jason C. White. On zero-sector reducing operations. *J. Math. Anal. Appl.*, 468(1):480–490, 2018.

[77] Augustin-Louis Cauchy. *Mémoire sur les intégrales définies, prises entre des limites imaginaires*. De Bure frères, Paris, 1825.

[78] Arthur Cayley. On a theorem in the geometry of position. *Cambridge Math. J.*, II:267–271, 1841.

[79] Debao Chen, Valdir Antonio Menegatto, and Xingping Sun. A necessary and sufficient condition for strictly positive definite functions on spheres. *Proc. Amer. Math. Soc.*, 131(9):2733–2740, 2003.

[80] Projesh Nath Choudhury. Characterizing total positivity: single vector tests via Linear Complementarity, sign non-reversal, and variation diminution. *Bull. London Math. Soc.*, 54(2):791–811, 2022.

[81] Projesh Nath Choudhury, M. Rajesh Kannan, and Apoorva Khare. Sign non-reversal property for totally non-negative and totally positive matrices and testing total positivity of their interval hull. *Bull. London Math. Soc.*, 53(4):981–990, 2021.

[82] Jens Peter Reus Christensen and Paul Ressel. Functions operating on positive definite matrices and a theorem of Schoenberg. *Trans. Amer. Math. Soc.*, 243:89–95, 1978.

[83] Jens Peter Reus Christensen and Paul Ressel. Positive definite kernels on the complex Hilbert sphere. *Math. Z.*, 180(2):193–201, 1982.

[84] Henry Cohn and Matthew de Courcy-Ireland. The Gaussian core model in high dimensions. *Duke Math. J.*, 167(13):2417–2455, 2018.

[85] Henry Cohn and Abhinav Kumar. Universally optimal distribution of points on spheres. *J. Amer. Math. Soc.*, 20(1):99–148, 2007.

[86] Henry Cohn and Jeechul Woo. Three-point bounds for energy minimization. *J. Amer. Math. Soc.*, 25(4):929–958, 2012.

[87] Henry Cohn and Yufei Zhao. Sphere packing bounds via spherical codes. *Duke Math. J.*, 163(10):1965–2002, 2014.

[88] John Bligh Conway. *Functions of one complex variable.* Graduate Texts in Mathematics, Vol. 11, Springer, New York, 1978.

[89] Vincent Frederick Cowling and Wolfgang Joseph Thron. Zero-free regions of polynomials. *Amer. Math. Monthly*, 61(10):682–687, 1954.

[90] Trevor F. Cox and Michael A.A. Cox. Multidimensional scaling. In: *Handbook of Data Visualization* (C.-h. Chen, W. Härdle, and A. Unwin, Eds.), pp. 315–347, Springer, Berlin–Heidelberg, 2008.

[91] Thomas Craven and George Csordas. Composition theorems, multiplier Sequences and complex zero decreasing sequences. In: Value distribution theory and related topics (G. Barsegian, I. Laine, C.C. Yang (Eds.), pp. 131–166, Advances in Complex Analysis and Its Applications, Vol. 3, Springer, Boston, 2014.

[92] Thomas Craven, George Csordas, and Wayne Smith. The zeros of derivatives of entire functions and the Pólya–Wiman Conjecture. *Ann. of Math. (2)*, 125(2):405–431, 1987.

[93] Colin Walker Cryer. The LU-factorization of totally positive matrices. *Linear Algebra Appl.*, 7(1):83–92, 1973.

[94] Colin Walker Cryer. Some properties of totally positive matrices. *Linear Algebra Appl.*, 15(1):1–25, 1976.

[95] Haskell Brooks Curry and Isaac Jacob Schoenberg. On spline distributions and their limits: The Pólya distribution functions. *Bull. Amer. Math. Soc.*, 53(11):1114, 1947.

[96] Haskell Brooks Curry and Isaac Jacob Schoenberg. On Pólya frequency functions IV: The fundamental spline functions and their limits. *J. d'Analyse Math.*, 17:71–107, 1966.

[97] David Raymond Curtiss. Recent extensions of Descartes' Rule of Signs. *Ann. of Math. (2)*, 19(4):251–278, 1918.

[98] Yuri L'vovich Dalestkii and Selim Grigorievich Krein. Integration and differentiation of functions of Hermitian operators and applications to the theory of perturbations (Russian). *Voronoež. Gos. Univ. Trudy Sem. Funkcional. Anal.*, vol. 1, pp. 81–105 (1956); AMS Translations: Ser. 2, vol. 47, pp. 1–30, 1965.

[99] Aleksey Aleksandrovich Davydov. Totally positive sequences and R-matrix quadratic algebras. *J. Math. Sci.*, 100(1):1871–1876, 2000.

[100] Réne Descartes. Le Géométrie. Appendix to *Discours de la méthode*, 1637.

[101] Dimitar Kolev Dimitrov and Juan Manuel Peña. Almost strict total positivity and a class of Hurwitz polynomials. *Linear Algebra Appl.*, 132(2):212–223, 2005.

[102] Dimitar Kolev Dimitrov and Yuan Xu. Wronskians of Fourier and Laplace transforms. *Trans. Amer. Math. Soc.*, 372(6):4107–4125, 2019.

[103] William Francis Donoghue Jr. *Monotone matrix functions and analytic continuation.* Grundlehren der mathematischen Wissenschaften, Vol. 207, Springer-Verlag, Berlin, 1974.

[104] Alexander Dyachenko. Total nonnegativity of infinite Hurwitz matrices of entire and meromorphic functions. *Complex Anal. Oper. Theory*, 8:1097–1127, 2014.

[105] Harry Dym and Victor Emmanuilovich Katsnelson. Contributions of Issai Schur to analysis. In: *Studies in Memory of Issai Schur* (A. Joseph, A. Melnikov, and R. Rentschler, Eds.), pp. xci–clxxxiii, Progress in Math., Vol. 210, Birkhäuser, Basel, 2003.

[106] Albert Edrei. On the generating functions of totally positive sequences II. *J. d'Analyse Math.*, 2:104–109, 1952.

[107] Albert Edrei. Proof of a conjecture of Schoenberg on the generating function of a totally positive sequence. *Canad. J. Math.*, 5:86–94, 1953.

[108] Albert Edrei. On the generation function of a doubly infinite, totally positive sequence. *Trans. Amer. Math. Soc.*, 74:367–383, 1953.

[109] Bradley Efron. Increasing properties of Pólya frequency functions. *Ann. Math. Statist.*, 36(1):272–279, 1965.

[110] Noureddine El Karoui. Operator norm consistent estimation of large-dimensional sparse covariance matrices. *Ann. Statist.*, 36(6):2717–2756, 2008.

[111] Jan Emonds and Hartmut Führ. Strictly positive definite functions on compact abelian groups. *Proc. Amer. Math. Soc.*, 139(3):1105–1113, 2011.

[112] Shaun Michael Fallat and Charles Royal Johnson. *Totally nonnegative matrices.* Princeton Series in Applied Mathematics, Princeton University Press, Princeton, 2011.

[113] Shaun Michael Fallat, Charles Royal Johnson, and Alan David Sokal. Total positivity of sums, Hadamard products and Hadamard powers: Results and counterexamples. *Linear Algebra Appl.*, 520:242–259, 2017; Corrigendum, *Linear Algebra Appl.*, 613:393–396, 2021.

[114] Ky Fan. Some matrix inequalities. *Abh. Math. Sem. Univ. Hamburg*, 29:185–196, 1966.

[115] Jacques Faraut. Formule du binôme généralisée. In: *Harmonic Analysis*, Lecture Notes in Math., Vol. 1359, pp. 170–180, Springer, Berlin, 1988.

[116] Jacques Faraut and Ádám Korányi. Function spaces and reproducing kernels on bounded symmetric domains. *J. Funct. Anal.*, 88(1):64–89, 1990.

[117] Jacques Faraut and Ádám Korányi. *Analysis on symmetric cones.* Oxford Mathematical Monographs. The Clarendon Press, Oxford University Press, New York, second edition, 1994. Oxford Science Publications.

[118] Mihály Fekete and Georg Pólya. Über ein Problem von Laguerre. *Rend. Circ. Mat. Palermo*, 34:89–120, 1912.

[119] William Feller. An introduction to probability theory and its applications (1st ed.). Wiley & Sons, New York, xviii+636 pp., 1966.

[120] Miroslav Fiedler and Thomas Markham. An observation on the Hadamard product of Hermitian matrices. *Linear Algebra Appl.*, 215:179–182, 1995.

[121] Miroslav Fiedler and Vlastimil Pták. On matrices with non-positive off-diagonal elements and positive principal minors. *Czech. Math. J.*, 12(3):382–400, 1962.

[122] Steve Fisk. *Polynomials, roots, and interlacing.* Version 2, arXiv:math.CA/0612833, 2008.

[123] Carl Hanson FitzGerald and Roger Alan Horn. On fractional Hadamard powers of positive definite matrices. *J. Math. Anal. Appl.*, 61(3):633–642, 1977.

[124] Carl Hanson FitzGerald, Charles Anthony Micchelli, and Allan Pinkus. Functions that preserve families of positive semidefinite matrices. *Linear Algebra Appl.*, 221:83–102, 1995.

[125] Sergey Fomin and Andrei Zelevinsky. Double Bruhat cells and total positivity. *J. Amer. Math. Soc.*, 12(2):335–380, 1999.

[126] Sergey Fomin and Andrei Zelevinsky. Total positivity: Tests and parametrizations. *Math. Intelligencer*, 22(1):23–33, 2000.

[127] Sergey Fomin and Andrei Zelevinsky. Cluster algebras. I. Foundations. *J. Amer. Math. Soc.*, 15(2):497–529, 2002.

[128] Yanik-Pascal Förster, Mario Kieburg, and Holger Kösters. Polynomial ensembles and Pólya frequency functions. *J. Theor. Probab.*, 34(4):1917–1950, 2021.

[129] Jean-Baptiste Joseph Fourier. Sur l'usage du théorème de Descartes dans la recherche des limites des racines. *Bull. Sci. Soc. Phil. Paris*, 156–165 and 181–187, Gauthier–Villars, Paris, 1820.

[130] Maurice René Fréchet. Les dimensions d'un ensemble abstrait. *Math. Ann.*, 68:145–168, 1910.

[131] Maurice René Fréchet. Sur la définition axiomatique d'une classe d'espaces vectoriels distanciés applicables vectoriellement sur l'espace de Hilbert. *Ann. of Math. (2)*, 36(3):705–718, 1935.

[132] Kurt Otto Friedrichs. The identity of weak and strong extensions of differential operators. *Trans. Amer. Math. Soc.*, 55:132–151, 1944.

[133] Matsusaborô Fujiwara. Einige Bemerkungen über die elementare Theorie der algebraischen Gleichungen. *Tohoku Math. J.*, 9:102–108, 1916.

[134] David Gale and Hukukane Nikaido. The Jacobian matrix and global univalence of mappings. *Math. Ann.*, 159:81–93, 1965.

[135] Feliks Ruvimovich Gantmacher. Sur les noyaux de Kellogg non symmétriques. Comptes Rendus (Doklady) Acad. Sci. de l'URSS, 1(10):3–5, 1936.

[136] Feliks Ruvimovich Gantmacher. *The theory of matrices. Vols. 1, 2.* Translated by K.A. Hirsch. Chelsea Publishing Co., New York, 1959.

[137] Feliks Ruvimovich Gantmacher and Mark Grigor'evich Krein. Sur les matrices complètement non négatives et oscillatoires. *Compositio Math.*, 4:445–476, 1937.

[138] Feliks Ruvimovich Gantmacher and Mark Grigor'evich Krein. *Oscillation matrices and kernels and small vibrations of mechanical systems.* Translated by A. Eremenko. Chelsea Publishing Co., New York, 2002.

[139] Jürgen Garloff. Intervals of almost totally positive matrices. *Linear Algebra Appl.*, 363:103–108, 2003.

[140] Jürgen Garloff and David G. Wagner. Hadamard products of stable polynomials are stable. *J. Math. Anal. Appl.*, 202(3):797–809, 1996.

[141] *Total positivity and its applications* (M. Gasca and C.A. Micchelli, Eds.). Proceedings of Jaca Meeting (1994), Mathematics and its Applications, Vol. 359, Kluwer, Dordrecht, 1996.

[142] Ira Gessel and Xavier Viennot. Binomial determinants, paths, and hook length formulae. *Adv. Math.*, 58(3):300–321, 1985.

[143] Simon Grigorevich Gindikin. Invariant generalized functions in homogeneous domains. *Funct. Anal. Appl.*, 9:50–52, 1975.

[144] Tillmann Gneiting. Strictly and non-strictly positive definite functions on spheres. *Bernoulli*, 19(4):1327–1349, 2013.

[145] Piotr Graczyk, Jacek Małecki, and Eberhard Mayerhofer. A characterization of Wishart processes and Wishart distributions. *Stoch. Proc. Appl.*, 128(4):1386–1404, 2018.

[146] Kenneth Ralph Goodearl, Stephane Launois, and Tom H. Lenagan. Totally nonnegative cells and matrix Poisson varieties. *Adv. in Math.*, 226(1):779–826, 2011.

[147] Karlheinz Gröchenig. Schoenberg's theory of totally positive functions and the Riemann zeta function. *Preprint*, arXiv:math.NT/2007.12889, 2020.

[148] Karlheinz Gröchenig, José Luis Romero, and Joachim Stöckler. Sampling theorems for shift-invariant spaces, Gabor frames, and totally positive functions. *Invent. Math.*, 211:1119–1148, 2018.

[149] Karlheinz Gröchenig and Joachim Stöckler. Gabor frames and totally positive functions. *Duke Math. J.*, 162(6):1003–1031, 2013.

[150] Jean Carlo Guella, Valdir Antonio Menegatto, and Ana Paula Peron. Strictly positive definite kernels on a product of circles. *Positivity*, 21:329–342, 2017.

[151] Dominique Guillot, Apoorva Khare, and Bala Rajaratnam. Complete characterization of Hadamard powers preserving Loewner positivity, monotonicity, and convexity. *J. Math. Anal. Appl.*, 425(1):489–507, 2015.

[152] Dominique Guillot, Apoorva Khare, and Bala Rajaratnam. Preserving positivity for matrices with sparsity constraints. *Trans. Amer. Math. Soc.*, 368(12):8929–8953, 2016.

[153] Dominique Guillot, Apoorva Khare, and Bala Rajaratnam. Critical exponents of graphs. *J. Combin. Theory Ser. A*, 139:30–58, 2016.

[154] Dominique Guillot, Apoorva Khare, and Bala Rajaratnam. The critical exponent: a novel graph invariant. *Sem. Lothar. Combin.*, 78B, article 62, 12 pp., 2017.

[155] Dominique Guillot, Apoorva Khare, and Bala Rajaratnam. Preserving positivity for rank-constrained matrices. *Trans. Amer. Math. Soc.*, 369(9):6105–6145, 2017.

[156] Dominique Guillot and Bala Rajaratnam. Retaining positive definiteness in thresholded matrices. *Linear Algebra Appl.*, 436(11):4143–4160, 2012.

[157] Dominique Guillot and Bala Rajaratnam. Functions preserving positive definiteness for sparse matrices. *Trans. Amer. Math. Soc.*, 367(1):627–649, 2015.

[158] Jacques Salomon Hadamard. Essai sur l'étude des fonctions données par leur développement de Taylor. *J. Math. Pures Appl.*, 8:101–186, 1892.

[159] Jacques Salomon Hadamard. Étude sur les propriétés des fonctions entières et an particulier d'une fonction considérée par Riemann. *J. Math. Pures Appl.*, 171–215, 1893.

[160] Jacques Salomon Hadamard. Théorème sur les séries entières. *Acta Math.*, 22:55–63, 1899.

[161] Hans Ludwig Hamburger. Bemerkungen zu einer Fragestellung des Herrn Pólya. *Math. Z.*, 7:302–322, 1920.

[162] Hans Ludwig Hamburger. Über eine Erweiterung des Stieltjesschen Momentenproblems. *Math. Ann.*, Part I, Vol. 81, pp. 235–319, 1920; Part II, Vol. 82, pp. 120–164, 1921; Part III, Vol. 82, pp. 168–187, 1921.

[163] Berthold Heiligers. Totally nonnegative moment matrices. *Linear Algebra Appl.*, 199 (suppl. 1):213–227, 1994.

[164] Henry Helson, Jean-Pierre Kahane, Yitzhak Katznelson, and Walter Rudin. The functions which operate on Fourier transforms. *Acta Math.*, 102(1):135–157, 1959.

[165] Charles Hermite. Extrait d'une lettre de Mr. Ch. Hermite de Paris à Mr. Borchardt de Berlin sur le nombre des racines d'une équation algébrique comprises entre des limites données. *J. reine angew. Math.*, 52:39–51, 1856.

[166] Charles Hermite. Questions 777, 778, 779. *Nouv. Ann. Math.*, 5:pp. 432, 479, 1866.

[167] Alfred Hero and Bala Rajaratnam. Large-scale correlation screening. *J. Amer. Statist. Assoc.*, 106(496):1540–1552, 2011.

[168] Alfred Hero and Bala Rajaratnam. Hub discovery in partial correlation graphs. *IEEE Trans. Inform. Theory*, 58(9):6064–6078, 2012.

[169] Daniel Hershkowitz, Michael Neumann, and Hans Schneider. Hermitian positive semidefinite matrices whose entries are 0 or 1 in modulus. *Linear Multilinear Algebra*, 46(4):259–264, 1999.

[170] Carl S. Herz. Fonctions opérant sur les fonctions définies-positives. *Ann. Inst. Fourier (Grenoble)*, 13:161–180, 1963.

[171] Fumio Hiai. Monotonicity for entrywise functions of matrices. *Linear Algebra Appl.*, 431(8):1125–1146, 2009.

[172] Fumio Hiai and Denes Petz. *Introduction to matrix analysis and applications.* Universitext, Springer, Cham, 2014, viii+332 pp.

[173] David Hilbert. Über die Darstellung definiter Formen als Summe von Formenquadraten. *Math. Ann.*, 32:342–350, 1888.

[174] David Hilbert. Über ternäre definite Formen. *Acta Math.*, 17:169–197, 1893.

[175] Isidore Isaac Hirschman Jr. and David Vernon Widder. The inversion of a general class of convolution transforms. *Trans. Amer. Math. Soc.*, 66(1):135–201, 1949.

[176] Isidore Isaac Hirschman Jr. and David Vernon Widder. *The convolution transform.* Princeton Legacy Library, Princeton University Press, Princeton, 1955.

[177] Phùng Hô Hai. Poincaré series of quantum spaces associated to Hecke operators. *Acta Math. Vietnam.*, 24(2):235–246, 1999.

[178] Thomas Hofmann, Bernhard Schölkopf, and Alexander J. Smola. Kernel methods in machine learning. *Ann. Statist.*, 36(3):1171–1220, 2008.

[179] Olga Holtz. Hermite–Biehler, Routh–Hurwitz, and total positivity. *Linear Algebra Appl.*, 372:105–110, 2003.

[180] Olga Holtz, Sergey Khrushchev, and Olga Kushel. Generalized Hurwitz matrices, generalized Euclidean algorithm, and forbidden sectors of the complex plane. *Comput. Methods Funct. Theory*, 16:395–431, 2016.

[181] Olga Holtz and Mikhail Tyaglov. Structured matrices, continued fractions, and root localization of polynomials. *SIAM Rev.*, 54(3):421–509, 2012.

[182] Roger Alan Horn. The theory of infinitely divisible matrices and kernels. *Trans. Amer. Math. Soc.*, 136:269–286, 1969.

[183] Roger Alan Horn and Charles Royal Johnson. *Matrix analysis.* Cambridge University Press, Cambridge, 1985.

[184] Roger Alan Horn and Charles Royal Johnson. *Topics in matrix analysis.* Cambridge University Press, Cambridge, 1991.

[185] Loo-Keng Hua. Harmonic analysis of functions of several complex variables in the classical domains. Translations of Mathematical Monographs, Vol. 6, American Mathematical Society, Providence, 1963.

[186] Hsin-Yuan Huang and Chih-Jen Lin. Linear and kernel classification: When to use which? *Proc. 2016 SIAM Conf. Data Min.*, 216–224, 2016.

[187] David Boss Hunter. The positive-definiteness of the complete symmetric functions of even order. *Math. Proc. Camb. Phil. Soc.*, 82(2):255–258, 1977.

[188] Adolf Hurwitz. Über die Bedingungen, unter welchen eine Gleichung nur Wurzeln mit negativen reellen Theilen besitzt. *Math. Ann.*, 46:273–284, 1895.

[189] Adolf Hurwitz. Über den Satz von Budan–Fourier. *Math. Ann.*, 71:584–591, 1911.

[190] Carl Gustav Jacob Jacobi. De formatione et proprietatibus Determinatium. *J. reine angew. Math.*, 22:285–318, 1841.

[191] Tanvi Jain. Hadamard powers of some positive matrices. *Linear Algebra Appl.*, 528:147–158, 2017.

[192] Tanvi Jain. Hadamard powers of rank two, doubly nonnegative matrices. *Adv. in Oper. Theory*, 5:839–849, 2020 (Rajendra Bhatia volume).

[193] Graham James Oscar Jameson. Counting zeros of generalised polynomials: Descartes' rule of signs and Laguerre's extensions. *The Mathematical Gazette*, 90(518):223–234, 2006.

[194] Johan Ludwig William Valdemar Jensen. Recherches sur la théorie des équations. *Acta. Math.*, 36:181–195, 1913.

[195] Robert Jentzsch. Über Integralgleichungen mit positivem Kern. *J. reine angew. Math.*, 141:235–244, 1912.

[196] Charles Royal Johnson and Ronald L. Smith. Line insertions in totally positive matrices. *J. Approx. Theory*, 105:305–312, 2000.

[197] Charles Royal Johnson and Olivia Walch. Critical exponents: Old and new. *Electr. J. Linear Alg.*, 25:72–83, 2012.

[198] Richard V. Kadison and Isadore Manuel Singer. Extensions of pure states. *Amer. J. Math.*, 81(2):383–400, 1959.

[199] Samuel Karlin. Total positivity, absorption probabilities and applications. *Trans. Amer. Math. Soc.*, 111:33–107, 1964.

[200] Samuel Karlin. *Total positivity. Vol. 1.* Stanford University Press, Stanford, 1968.

[201] Samuel Karlin and James McGregor. Coincidence probabilities of birth and death processes. *Pacific J. Math.*, 9(4):1109–1140, 1959.

[202] Samuel Karlin and James McGregor. Coincidence probabilities. *Pacific J. Math.*, 9(4):1141–1164, 1959.

[203] Samuel Karlin, Frank Proschan, and Richard Eugene Barlow. Moment inequalities of Pólya frequency functions. *Pacific J. Math.*, 11(3):1023–1033, 1961.

[204] Samuel Karlin and Yosef Rinott. Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions. *J. Multivariate Anal.*, 10(4):467–498, 1980.

[205] Samuel Karlin and Yosef Rinott. Total positivity properties of absolute value multinormal variables with applications to confidence interval estimates and related probabilistic inequalities. *Ann. Statist.*, 9(5):1035–1049, 1981.

[206] Samuel Karlin and Zvi Ziegler. Chebyshevian spline functions. *SIAM J. Numer. Anal.*, 3(3):514–543, 1966.

[207] Olga M. Katkova. A certain method for constructing Pólya frequency sequences. *J. Soviet Math.*, 52:3539–3545, 1990.

[208] Olga M. Katkova. Multiple positivity and the Riemann zeta-function. *Comput. Methods Funct. Theory*, 7:13–31, 2007.

[209] Olga M. Katkova and Anna M. Vishnyakova. On sufficient conditions for the total positivity and for the multiple positivity of matrices. *Linear Algebra Appl.*, 16(2–3):1083–1097, 2006.

[210] Olga M. Katkova and Anna M. Vishnyakova. A sufficient condition for a polynomial to be stable. *J. Math. Anal. Appl.*, 347(1):81–89, 2008.

[211] Oliver Dimon Kellogg. Orthogonal function sets arising from integral equations. *Amer. J. Math.*, 40(2):145–154, 1918.

[212] Johannes Henricus Bernardus Kemperman. A Hurwitz matrix is totally positive. *SIAM J. Math. Anal.*, 13(2):331–341, 1982.

[213] Apoorva Khare. Sharp nonzero lower bounds for the Schur product theorem. *Proc. Amer. Math. Soc.*, 149(12):5049–5063, 2021.

[214] Apoorva Khare. Smooth entrywise positivity preservers, a Horn–Loewner master theorem, and symmetric function identities. *Trans. Amer. Math. Soc.*, 375(3):2217–2236, 2022.

[215] Apoorva Khare. Critical exponents for total positivity, individual kernel encoders, and the Jain–Karlin–Schoenberg kernel. *Preprint*, arXiv:math.FA/2008.05121, 2020.

[216] Apoorva Khare and Terence Tao. Schur polynomials, entrywise positivity preservers, and weak majorization. *Sem. Lothar. Combin.*, 80B, article 14, 12 pp., 2018.

[217] Apoorva Khare and Terence Tao. On the sign patterns of entrywise positivity preservers in fixed dimension. *Amer. J. Math.*, 143(6):1863–1929, 2021.

[218] Jee Soo Kim and Frank Proschan. Total positivity. In: *Encyclopedia of Statistical Sciences* (S. Kotz et al, Eds.), Vol. 14, Wiley, New York, pp. 8665–8672, 2006.

[219] Jiae Kim, Yoonkyung Lee, and Zhiyu Liang. The geometry of nonlinear embeddings in kernel discriminant analysis. *Preprint*, http://arxiv.org/abs/2005.05546, 2020.

[220] Tinne Hoff Kjeldsen. The early history of the moment problem. *Historia Math.*, 20(1):19–44, 1993.

[221] Yuji Kodama and Lauren Williams. KP solitons, total positivity and cluster algebras. *Proc. Natl. Acad. Sci. USA*, 108:8984–8989, 2011.

[222] Yuji Kodama and Lauren Williams. KP solitons and total positivity for the Grassmannian. *Invent. Math.*, 198(3):637–699, 2014.

[223] Andrey Nikolaevich Kolmogorov. Kurven in Hilbertschen Raum, die gegenüber einer einparametrigen Gruppe von Bewegungen invariant sind. *C.R. Doklady Acad. Sci. U.R.S.S.* 26:6–9, 1940.

[224] Jacob Korevaar. The zeros of approximating polynomials and the canonical representation of an entire function. *Duke Math. J.*, 18(2):573–592, 1951.

[225] Jacob Korevaar. Entire functions as limits of polynomials. *Duke Math. J.*, 21(2):533–548, 1954.

[226] Bertram Kostant and Siddhartha Sahi. The Capelli identity, tube domains, and the generalized Laplace transform. *Adv. Math.*, 87(1):71–92, 1991.

[227] Steven George Krantz and Harold Raymond Parks. *A primer of real analytic functions.* Birkhäuser Advanced Texts, Birkhäuser, Boston, xiii+209 pp., 2002.

[228] Edmond Laguerre. Sur les fonctions du genre zéro et du genre un. *C. R. Acad. Sci. Paris*, 95:828–831, 1882.

[229] Edmond Laguerre. Mémoire sur la théorie des équations numériques. *J. Math. Pures Appl.*, 9:9–146, 1883.

[230] Edmond Laguerre. Sur quelques points de la théorie des équations numériques. *Acta Math.*, 4:97–120, 1884.

[231] Thomas Lam, Alexander E. Postnikov, and Pavlo Pylyavskyy. Schur positivity and Schur log-concavity. *Amer. J. Math.*, 129(6):1611–1622, 2007.

[232] Michel Lassalle. Algèbre de Jordan et ensemble de Wallach. *Invent. Math.*, 89(2):375–393, 1987.

[233] Tsung-Dao Lee and Chen Ning Yang. Statistical theory of equations of state and phase transitions. II. Lattice gas and Ising model. *Phys. Rev.*, 87:410–419, 1952.

[234] Gérard Letac and Hélène Massam. The Laplace transform $(\det s)^{-p} \exp \operatorname{tr}(s^{-1}w)$ and the existence of non-central Wishart distributions. *J. Multivar. Anal.*, 163:96–110, 2018.

[235] Boris Yakovlevich Levin. *Distribution of zeros of entire functions.* Translations of Mathematical Monographs, Vol. 5, Translated from the Russian by R.P. Boas Jr., J.M. Danskin, F.M. Goodspeed, J. Korevaar, A.L. Shields, and H.P. Thielman, American Mathematical Society, Providence, 1980.

[236] Ai Li and Steve Horvath. Network neighborhood analysis with the multi-node topological overlap measure. *Bioinformatics*, 23(2):222–231, 2007.

[237] Leo Liberti and Carlile Lavor. Six mathematical gems from the history of distance geometry. *Int. Trans. Oper. Res.*, 23(5):897–920, 2016.

[238] Bernt Lindström. On the vector representations of induced matroids. *Bull. London Math. Soc.*, 5(1):85–90, 1973.

[239] Egon Lindwart and Georg Pólya. Über einen zusammenhang zwischen der konvergenz von polynomfolgen und der verteilung ihrer wurzeln. *Rend. Circ. Mat. Palermo*, 37:279–304, 1914.

[240] Charles Loewner. Über monotone Matrixfunktionen. *Math. Z.*, 38(1):177–216, 1934.

[241] Charles Loewner. On totally positive matrices. *Math. Z.*, 63:338–340, 1955 (Issai Schur memorial volume).

[242] Charles Loewner. Determination of the critical exponent of the Green's function. *1965 Symposium on Function Theory*, Erevan, USSR, 1966.

[243] Charles Loewner. On schlicht-monotonic functions of higher order. *J. Math. Anal. Appl.*, 14(2):320–325, 1966.

[244] Félix Lucas. Propriétés géométriques des fractionnes rationnelles. *C. R. Acad. Sci. Paris*, 77:431–433, 1874.

[245] George Lusztig. Total positivity in reductive groups, Lie theory and geometry. *Progr. Math.*, Vol. 123, Birkhäuser Boston, Boston, MA, 1994, pp. 531–568.

[246] George Lusztig. Total positivity and canonical bases. In: *Algebraic groups and Lie groups* (G.I. Lehrer, Ed.), Austral. Math. Soc. Lect. Ser., Vol. 9, Cambridge University Press, Cambridge, 1997, pp. 281–295.

[247] Ian Grant Macdonald. *Symmetric functions and Hall polynomials.* Oxford Mathematical Monographs. The Clarendon Press, Oxford University Press, New York, second edition, 1995. With contributions by A. Zelevinsky, Oxford Science Publications.

[248] Nathaniel Macon and Abraham Spitzbart. Inverses of Vandermonde matrices. *Amer. Math. Monthly*, 65(2):95–100, 1958.

[249] Ernest Maló. Note sur les équations algébriques dont toutes les racines sont réelles. *J. Math. Spéc.* (Ser. 4), 4:7–10, 1895.

[250] Adam W. Marcus, Daniel A. Spielman, and Nikhil Srivastava. Interlacing families I: Bipartite Ramanujan graphs of all degrees. *Ann. of Math. (2)*, 182(1):307–325, 2015.

[251] Adam W. Marcus, Daniel A. Spielman, and Nikhil Srivastava. Interlacing families II. Mixed characteristic polynomials and the Kadison-Singer problem. *Ann. of Math. (2)*, 182(1):327–350, 2015.

[252] Adam W. Marcus, Daniel A. Spielman, and Nikhil Srivastava. Interlacing families IV. Bipartite Ramanujan graphs of all sizes. *SIAM J. Comput.*, 47(6):2488–2509, 2018.

[253] Marvin David Marcus and Nisar A. Khan. A note on the Hadamard product. *Canad. Math. Bull.*, 2(2):81–83, 1959.

[254] Michael Margaliot and Eduardo D. Sontag. Revisiting totally positive differential systems: A tutorial and new results. *Automatica*, 101:1–14, 2019.

[255] Eberhard Mayerhofer. On Wishart and noncentral Wishart distributions on symmetric cones. *Trans. Amer. Math. Soc.*, 371(10):7093–7109, 2019.

[256] Valdir Antonio Menegatto and Claudemir Pinheiro Oliveira. Positive definiteness on products of compact two-point homogeneous spaces and locally compact abelian groups. *Canad. Math. Bull.*, 63(4):705–715, 2020.

[257] Valdir Antonio Menegatto and Ana Paula Peron. Positive definite kernels on complex spheres. *J. Math. Anal. Appl.*, 254(1):219–232, 2001.

[258] Karl Menger. Die Metrik des Hilbertschen Raumes. *Anz. Akad. Wissen. Wien, Math. Nat. Kl.*, 65:159–160, 1928.

[259] Karl Menger. Untersuchungen über allgemeine Metrik. *Math. Ann.*: *Part I*: Vol. 100:75–163, 1928; *Part II*: Vol. 103:466–501, 1930.

[260] Karl Menger. New foundation of Euclidean geometry. *Amer. J. Math.*, 53(4):721–745, 1931.

[261] James Mercer. Functions of positive and negative type and their connection with the theory of integral equations. *Phil. Trans. Royal Soc. A*, 209:415–446, 1909.

[262] Charles Anthony Micchelli. Cardinal L-splines. In: *Studies in spline functions and approximation theory* (S. Karlin, C.A. Micchelli, A. Pinkus, and I.J. Schoenberg, Eds.), pp. 203–250, Academic Press, New York, 1976.

[263] Theodore Samuel Motzkin. *Beiträge zur Theorie der linearen Ungleichungen.* PhD dissertation, Basel, 1933 and Jerusalem, 1936.

[264] Theodore Samuel Motzkin. Relations between hypersurface cross ratios, and a combinatorial formula for partitions of a polygon, for permanent preponderance, and for non-associative products. *Bull. Amer. Math. Soc.*, 54(4):352–360, 1948.

[265] Theodore Samuel Motzkin. The Euclidean algorithm. *Bull. Amer. Math. Soc.*, 55(12):1142–1146, 1949.

[266] Theodore Samuel Motzkin. A proof of Hilbert's Nullstellensatz. *Math. Z.*, 63:341–344, 1955 (Issai Schur memorial volume).

[267] Theodore Samuel Motzkin. The arithmetic-geometric inequality. In: *Inequalities* (Proc. Sympos. Wright–Patterson Air Force Base, Ohio, 1965; O. Shisha, Ed.), pp. 205–224, Academic Press, New York, 1967.

[268] Oleg R. Musin. The kissing number in four dimensions. *Ann. of Math. (2)*, 168(1):1–32, 2008.

[269] Arnold Neumaier. Introduction to coherent spaces. *Preprint*, arXiv:math-ph/1804.01402, 2018.

[270] Rolf Herman Nevanlinna. *Le théorème de Picard–Borel et la théorie des fonctions méromorphes.* Borel series of monographs, Gauthier-Villars, Paris, 1929.

[271] Nikola Obrechkoff. Quelques classes de fonctions entières limites de polynomes et de fonctions méromorphes limites de fractions rationnelles. Actualités Sci. Ind., Vol. 891, Hermann, Paris, 1941.

[272] Nikola Obrechkoff. Sur les fonctions entières limites de polynomes dont les zéros sont réels et entrelacés. *Mat. Sb.*, 9(51):421–428, 1941.

[273] Nikola Obrechkoff. *Verteilung und Berechnung der Nullstellen reeller Polynome.* VEB Deutscher Verlag der Wissenschafter, 1963.

[274] Nikola Obrechkoff. *Zeros of polynomials.* Marin Drinov Acad. Publ., Sofia, 2003.

[275] Alexander Markowich Ostrowski. Über die Funktionalgleichung der Exponentialfunktion und verwandte Funktionalgleichung. *Jber. Deut. Math. Ver.*, 38:54–62, 1929.

[276] James Eldred Pascoe. Noncommutative Schur-type products and their Schoenberg theorem. *Preprint*, arXiv:math.FA/1907.04480, 2019.

[277] Shyamal Das Peddada and Donald St. P. Richards. Proof of a conjecture of M.L. Eaton on the characteristic function of the Wishart distribution. *Ann. Probab.*, 19(2):868–874, 1991. See also *Ann. Probab.*, 20(2):1107, 1992.

[278] Oskar Perron. Zur theorie der matrices. *Math. Ann.*, 64:248–263, 1907.

[279] Allan Pinkus. Spectral Properties of Totally Positive Kernels and Matrices. Available online.

[280] Allan Pinkus. Some remarks on zero-increasing transformations. In: *Approximation Theory X: Abstract and Classical Analysis* (C.K. Chui, L.L. Schumaker, J. Stöckler, Eds.), pp. 333–352, Vanderbilt Univ. Press, Nashville, 2002.

[281] Allan Pinkus. Strictly positive definite functions on a real inner product space. *Adv. Comput. Math.*, 20:263–271, 2004.

[282] Allan Pinkus. *Totally positive matrices.* Cambridge Tracts in Mathematics, Vol. 181, Cambridge University Press, Cambridge, 2010.

[283] Jim Pitman. Probabilistic bounds on the coefficients of polynomials with only real zeros. *J. Combin. Th. Ser. A*, 77(2):279–303, 1997.

[284] Georg Pólya. Über Annäherung durch Polynome mit lauter reellen Wurzeln. *Rend. Circ. Mat. Palermo*, 36:279–295, 1913.

[285] Georg Pólya. Algebraische Untersuchungen über ganze Funktionen vom Geschlechte Null und Eins. *J. reine angew. Math.*, 145:224–249, 1915.

[286] Georg Pólya. Über Trigonometrische integrale mit nur reellen Nullstellen. *J. reine angew. Math.*, 158:6–18, 1927.

[287] Georg Pólya. Über die algebraisch-funktionentheoretischen Untersuchungen von J.L.W.V. Jensen. *Kgl. Danske Vid. Sel. Math.-fys. Medd.*, 7(17), 34 pp., 1927.

[288] Georg Pólya and Issai Schur. Über zwei Arten von Faktorenfolgen in der Theorie der algebraischen Gleichungen. *J. reine angew. Math.*, 144:89–113, 1914.

[289] Georg Pólya and Gabor Szegő. *Aufgaben und Lehrsätze aus der Analysis. Band II: Funktionentheorie, Nullstellen, Polynome Determinanten, Zahlentheorie.* Springer-Verlag, Berlin, 1925.

[290] Tiberiu Popoviciu. Sur l'approximation des fonctions convexes d'ordre supérieur. *Mathematica (Cluj)*, 8:1–85, 1934.

[291] Emilio Porcu, Moreno Bevilacqua, and Marc G. Genton. Spatio-temporal covariance and cross-covariance functions of the great circle distance on a sphere. *J. Amer. Statist. Assoc.*, 111(514):888–898, 2016.

[292] Alexander E. Postnikov. Total positivity, Grassmannians, and networks. *Preprint*, arXiv:math.CO/0609764, 2006.

[293] Augustine Francois Poulain. Théorèmes généraux sur les équations algébriques. *Nouv. Ann. Math.*, 6:21–33, 1867.

[294] Carl Johan Ragnarsson, Wesley Wai Suen, and David G. Wagner. Correction to a theorem on total positivity. *Canad. Math. Bull.*, 49(2):281–284, 2006.

[295] Qazi Ibadur Rahman and Gerhard Schmeisser. *Analytic theory of polynomials.* London Mathematical Society Monographs, Vol. 26. The Clarendon Press, Oxford University Press, New York, 2002.

[296] Robert Reams. Hadamard inverses, square roots and products of almost semidefinite matrices. *Linear Algebra Appl.*, 288:35–43, 1999.

[297] Paul Ressel. Laplace-transformation nichtnegativer und vektorwertiger maße. *Manuscripta Math.*, 13:143–152, 1974.

[298] Bernhard Riemann. Über die Anzahl der Primzahlen unter einer gegebenen Größe. *Monat. Berliner Akad.*, pp. 671–680, 1859.

[299] Konstanze Christina Rietsch. Quantum cohomology rings of Grassmannians and total positivity. *Duke Math. J.*, 110(3):523–553, 2001.

[300] Konstanze Christina Rietsch. Totally positive Toeplitz matrices and quantum cohomology of partial flag varieties. *J. Amer. Math. Soc.*, 16(2):363–392, 2003.

[301] A. Wayne Roberts and Dale E. Varberg. *Convex functions.* Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York–London, Vol. 57, Pure and Applied Mathematics, 1973.

[302] R. Tyrrell Rockafellar. *Convex analysis.* Princeton Mathematical Series, Vol. 28, Princeton University Press, Princeton, 1970.

[303] Hugo E. Rossi and Michéle Vergne. Analytic continuation of the holomorphic discrete series of a semi-simple Lie group. *Acta Math.*, 136:1–59, 1976.

[304] Adam J. Rothman, Elizaveta Levina, and Ji Zhu. Generalized thresholding of large covariance matrices. *J. Amer. Statist. Assoc.*, 104(485):177–186, 2009.

[305] Edward John Routh. *Stability of a given state of motion.* London, 1877.

[306] Walter Rudin. Positive definite sequences and absolutely monotonic functions. *Duke Math. J*, 26(4):617–622, 1959.

[307] David Ruelle. Characterization of Lee–Yang polynomials. *Ann. of Math. (2)*, 171(1):589–603, 2010.

[308] Siddhartha Sahi. Binomial coefficients and Littlewood–Richardson coefficients for Jack polynomials. *Int. Math. Res. Not.*, 2011(7):1597–1612, 2011.

[309] Konrad Schmüdgen. Around Hilbert's 17th problem. *Documenta Math.*, Extra Volume, 'Optimization Stories', pp. 433–438, 2012.

[310] Konrad Schmüdgen. *The moment problem.* Graduate Texts in Mathematics, Vol. 277, Springer, Cham, 2017.

[311] Isaac Jacob Schoenberg. Über variationsvermindernde lineare Transformationen. *Math. Z.*, 32:321–328, 1930.

[312] Isaac Jacob Schoenberg. Zur Abzählung der reellen Wurzeln algebraischer Gleichungen. *Math. Z.*, 38:546–564, 1934.

[313] Isaac Jacob Schoenberg. Remarks to Maurice Fréchet's article "Sur la définition axiomatique d'une classe d'espace distanciés vectoriellement applicable sur l'espace de Hilbert". *Ann. of Math. (2)*, 36(3):724–732, 1935.

[314] Isaac Jacob Schoenberg. Extensions of theorems of Descartes and Laguerre to the complex domain. *Duke Math. J.*, 2(1):84–94, 1936.

[315] Isaac Jacob Schoenberg. On certain metric spaces arising from Euclidean spaces by a change of metric and their imbedding in Hilbert space. *Ann. of Math. (2)*, 38(4):787–793, 1937.

[316] Isaac Jacob Schoenberg. Metric spaces and completely monotone functions. *Ann. of Math. (2)*, 39(4):811–841, 1938.

[317] Isaac Jacob Schoenberg. Metric spaces and positive definite functions. *Trans. Amer. Math. Soc.*, 44(3):522–536, 1938.

[318] Isaac Jacob Schoenberg. Positive definite functions on spheres. *Duke Math. J.*, 9(1):96–108, 1942.

[319] Isaac Jacob Schoenberg. Contributions to the problem of approximation of equidistant data by analytic functions. Part A. On the problem of smoothing or graduation. A first class of analytic approximation formulae. *Quart. Appl. Math.*, 4(1):45–99, 1946.

[320] Isaac Jacob Schoenberg. On totally positive functions, Laplace integrals and entire functions of the Laguerre–Pólya–Schur type. *Proc. Natl. Acad. Sci. USA*, 33(1):11–17, 1947.

[321] Isaac Jacob Schoenberg. On variation-diminishing integral operators of the convolution type. *Proc. Natl. Acad. Sci. USA*, 34(4):164–169, 1948.

[322] Isaac Jacob Schoenberg. Some analytical aspects of the problem of smoothing. In: *Studies and Essays presented to R. Courant on his 60th birthday* (K.O. Friedrichs, O.E. Neugebauer, J.J. Stoker, Eds.), pp. 351–370, Interscience Publ., New York, 1948.

[323] Isaac Jacob Schoenberg. On Pólya frequency functions. II. Variation-diminishing integral operators of the convolution type. *Acta Sci. Math. (Szeged)*, 12:97–106, 1950.

[324] Isaac Jacob Schoenberg. On Pólya frequency functions. I. The totally positive functions and their Laplace transforms. *J. d'Analyse Math.*, 1:331–374, 1951.

[325] Isaac Jacob Schoenberg. On the zeros of the generating functions of multiply positive sequences and functions. *Ann. of Math. (2)*, 62(3):447–471, 1955.

[326] Isaac Jacob Schoenberg. A note on multiply positive sequences and the Descartes rule of signs. *Rend. Circ. Mat. Palermo*, 4:123–131 1955.

[327] Isaac Jacob Schoenberg and Anne M. Whitney. On Pólya frequency functions. III. The positivity of translation determinants with an application to the interpolation problem by spline curves. *Trans. Amer. Math. Soc.*, 74:246–259, 1953.

[328] Issai Schur. Zur Theorie der linearen homogenen Integralgleichungen. *Math. Ann.*, 67:306–339, 1909.

[329] Issai Schur. Bemerkungen zur Theorie der beschränkten Bilinearformen mit unendlich vielen Veränderlichen. *J. reine angew. Math.*, 140:1–28, 1911.

[330] Issai Schur. Zwei Sätze über algebraische Gleichungen mit lauter reellen Wurzeln. *J. reine angew. Math.*, 144:75–88, 1914.

[331] Issai Schur. Über Potenzreihen, die im Innern des Einheitskreises beschränkt sind. *J. reine angew. Math.*, 147:205–232, 1917.

[332] Binyamin Schwarz. Totally positive differential systems. *Pacific J. Math.*, 32(1):203–229, 1970.

[333] Johann Andreas Segner. *Dissertatio epistolica, qua regulam harriotti...* PhD dissertation, Univ. Jena, 1728.

[334] Ambikeshwar Sharma and Jean Tzimbalario. A class of cardinal trigonometric splines. *SIAM J. Math. Anal.*, 7(6):809–819, 1976.

[335] James Alexander Shohat and Jacob Davidovich Tamarkin. *The problem of moments.* AMS Mathematical Surveys, American Mathematical Society, New York, 1943.
[336] Carl Ludwig Siegel. Über die analytische Theorie der quadratischen Formen. *Ann. of Math. (2)*, 36(3):527–606, 1935.
[337] Wacław Sierpińsky. Sur l'équation fonctionelle $f(x+y) = f(x) + f(y)$. *Fund. Math.*, 1(1):116–122, 1920.
[338] Barry Martin Simon. *Loewner's theorem on monotone matrix functions.* Grundlehren der mathematischen Wissenschaften, Vol. 354, Springer-Verlag, Berlin, 2019.
[339] Mark Skandera. Inequalities in products of minors of totally nonnegative matrices. *J. Algebraic Combin.*, 20(2):195–211, 2004.
[340] Serge Skryabin. On the graded algebras associated with Hecke symmetries, II. The Hilbert series. *J. Algebraic Combin.*, published online, DOI: 10.1007/s10801-021-01103-0.
[341] Alexander J. Smola, Zoltán L. Ovári, and Robert C. Williamson. Regularization with dot-product kernels. *Adv. Neural Inf. Proc. Syst.* (NIPS 2000; T.K. Leen, T.G. Dietterich, and V. Tresp, Eds.), pp. 308–314, MIT Press, 2001.
[342] Alexander J. Smola, Bernhard Schölkopf, and Klaus-Robert Müller. The connection between regularization operators and support vector kernels. *Neural Netw.*, 11(4):637–649, 1998.
[343] Sergei Lvovich Sobolev. Sur un théorème d'analyse fonctionnelle. *Rec. Math. (Mat. Sbornik) N.S.* 4(46):471–497, 1938.
[344] Alan David Sokal. When is a Riesz distribution a complex measure? *Bull. Soc. Math. France*, 139(4):519–534, 2011.
[345] Ingo Steinwart. On the influence of the kernel on the consistency of support vector machines. *J. Mach. Learn. Res.*, 2(1):67–93, 2002.
[346] James Drewry Stewart. Positive definite functions and generalizations, an historical survey. *Rocky Mountain J. Math.*, 6(3):409–434, 1976.
[347] Thomas Joannes Stieltjes. Recherches sur les fractions continues. *Ann. Fac. Sci. Toulouse*, 1–122, 1894–95.
[348] Charles François Sturm. Mémoire sur la résolution des équations numériques. *Bull. Sci. Férussac*, 11:419–425, 1829.
[349] Bernd Sturmfels. Totally positive matrices and cyclic polytopes. *Linear Algebra Appl.*, 107:275–281, 1988.
[350] Gabor Szegő. Bemerkungen zu einem Satz von J.H. Grace über die Wurzeln algebraischer Gleichungen. *Math. Z.*, 13:28–55, 1922.
[351] Edward Burr Van Vleck. On the location of roots of polynomials and entire functions. *Bull. Amer. Math. Soc.*, 35(5):643–683, 1929.
[352] Vladimir Naumovich Vapnik. The nature of statistical learning theory. *Statistics for Engineering and Information Science*, Springer, New York, 2000.
[353] Harkrishan Lal Vasudeva. Positive definite matrices and absolutely monotonic functions. *Indian J. Pure Appl. Math.*, 10(7):854–858, 1979.
[354] Prateek Kumar Vishwakarma. Positivity preservers forbidden to operate on diagonal blocks. *Trans. Amer. Math. Soc.*, in press; arXiv:math.CA/2002.00332.
[355] John von Neumann and Isaac Jacob Schoenberg. Fourier integrals and metric geometry. *Trans. Amer. Math. Soc.*, 50:226–251, 1941.
[356] Jan Vybíral. A variant of Schur's product theorem and its applications. *Adv. Math.*, 368, Article 107140, 9 pp., 2020.
[357] David G. Wagner. Total positivity of Hadamard products. *J. Math. Anal. Appl.*, 163(2):459–483, 1992.
[358] David G. Wagner. Multivariate stable polynomials: theory and applications. *Bull. Amer. Math. Soc.*,48(1):53–84, 2011.
[359] Nolan Russell Wallach. The analytic continuation of the discrete series. *Trans. Amer. Math. Soc.*, 251, 1979. Part I, pp. 1–17; Part II, pp. 19–37.
[360] John Wallis. *Arithmetica infinitorum, Opera mathematica I.* Oxford, 1656.
[361] Hsien-Chung Wang. Two-point homogeneous spaces. *Ann. of Math. (2)*, 55(1):177–191, 1952.
[362] Yi Wang and Yeong-Nan Yeh. Polynomials with real zeros and Pólya frequency sequences. *J. Combin. Theory Ser. A*, 109(1):63–74, 2005.
[363] Yi Wang and Bin Zhang. Hadamard powers of polynomials with only real zeros. *Linear Algebra Appl.*, 439:3173–3176, 2013.

[364] Karl Theodor Wilhelm Weierstrass. Zur theorie der eindeutigen analytischen Funktionen, *Mathematische Abhandlungen der Akademie der Wissenschaftenzu Berlin*, S. 11–60 (French Translation: (1879), Mémoire sur les fonctions analytiques uniformes. Traduit par E. Picard, *Annales Scientifiques de l'École Normale Supérieure*, 2^e Série, 8: 111–150), 1876.

[365] Hans Felix Weinberger. A characterization of the Pólya frequency functions of order 3. *Appl. Anal.*, 15(1–4):53–69, 1983.

[366] Louis Weisner. Polynomials whose roots lie in a sector. *Amer. J. Math.*, 64(1):55–60, 1942.

[367] Philip White and Emilio Porcu. Towards a complete picture of stationary covariance functions on spheres cross time. *Electron. J. Statist.*, 13(2):2566–2594, 2019.

[368] Anne M. Whitney. A reduction theorem for totally positive matrices. *J. d'Analyse Math.*, 2(1):88–92, 1952.

[369] David Vernon Widder. Necessary and sufficient conditions for the representation of a function by a doubly infinite Laplace integral. *Bull. Amer. Math. Soc.*, 40(4):321–326, 1934.

[370] David Vernon Widder. *The Laplace transform.* Princeton Legacy Library, Princeton University Press, Princeton, 1941.

[371] John Wishart. The generalised product moment distribution in samples from a normal multivariate population. *Biometrika*, 20A(1/2):32–52, 1928.

[372] Hans Sylvain Witsenhausen. Minimum dimension embedding of finite metric spaces. *J. Combin. Th. Ser. A*, 42(2):184–199, 1986.

[373] Yuan Xu. Positive definite functions on the unit sphere and integrals of Jacobi polynomials. *Proc. Amer. Math. Soc.*, 146(5):2039–2048, 2018.

[374] Yuan Xu and Elliott Ward Cheney. Strictly positive definite functions on spheres. *Proc. Amer. Math. Soc.*, 116(4):977–981, 1992.

[375] Chen Ning Yang and Tsung-Dao Lee. Statistical theory of equations of state and phase transitions. I. Theory of Condensation. *Phys. Rev.*, 87:404–409, 1952.

[376] Xingzhi Zhan. *Matrix theory.* Graduate Studies in Mathematics, Vol. 147, American Mathematical Society, Providence, 2013, 264 pp.

[377] Bin Zhang and Steve Horvath. A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.*, 4: Article 17, 45 pp. (electronic), 2005.

[378] Fuzhen Zhang. *Matrix theory: Basic results and techniques.* Universitext, Springer-Verlag, New York, 2011, xvii+399 pp.

[379] Johanna Ziegel. Convolution roots and differentiability of isotropic positive definite functions on spheres. *Proc. Amer. Math. Soc.*, 142(6):2063–2077, 2014.

A

B

C

D

E

F

G

H