THE ENTRYWISE CALCULUS AND DIMENSION-FREE POSITIVITY PRESERVERS, WITH AN APPENDIX ON SPHERE PACKINGS

APOORVA KHARE

ABSTRACT. We present an overview of a classical theme in analysis and matrix positivity: the question of which functions preserve positive semidefiniteness when applied entrywise. In addition to drawing the attention of experts such as Schoenberg, Rudin, and Loewner, the subject has attracted renewed attention owing to its connections to various applied fields and techniques. In this survey we will focus mainly on the question of preserving positivity in all dimensions. Connections to distance geometry and metric embeddings, positive definite sequences and functions, Fourier analysis, applications and covariance estimation, Schur polynomials, and finite fields will be discussed.

The Appendix contains a mini-survey of sphere packings, kissing numbers, and their "lattice" versions. This part overlaps with the rest of the article via Schoenberg's classification of the positive definite functions on spheres, aka dimension-free entrywise positivity preservers with a rank constraint – applied via Delsarte's linear programming method.

CONTENTS

Date: November 11, 2025.

2020 *Mathematics Subject Classification.* 15-02, 26-02, 52-02, 52-03; 15B48, 51F99, 15B48, 42A70, 15A15, 15A45, 15A83, 62J10, 47B34, 42A82, 43A35, 51K05, 30L05, 52C17, 33C55, 11H31, 94B27.

Key words and phrases. positive semidefinite matrix, entrywise calculus, Toeplitz matrix, Hankel matrix, absolutely monotonic function, metric geometry, spherical embedding, positive definite function, two-point homogeneous space, Schur polynomials, symmetric function identities, sample covariance, covariance estimation, finite fields; sphere packing, packing density, lattice packing, E_8 lattice, Leech lattice, kissing number, Hermite constant, spherical harmonics, addition theorem, Gegenbauer polynomial, Chebyshev polynomial, spherical code.

1. Introducing positivity preservers

The goal of this article is to survey a foundational result in matrix analysis, whose origins can be traced back to Pólya and Szegő exactly one hundred years ago (following Schur). This result, originally proved by Schoenberg (and then Rudin and many others), continues to yield connections to active areas of mathematics and applied fields.

The result in question combines two evergreen ingredients in mathematics: preserver problems and positive matrices. The notion of positivity is as old as mathematics – starting with counting and measuring. More pertinently, positivity of real symmetric matrices occurs at least as early as the second partial derivative test for local minima (via the Hessian matrix). On the complex side, an early occurrence of positive semidefinite matrices is in Pick and Nevanlinna's solutions of their eponymous interpolation problem (1910s).

Recall that a Hermitian matrix $A \in \mathbb{C}^{n \times n}$ is said to be *positive definite* if the associated quadratic form $Q(x) := x^* A x$, $x \in \mathbb{C}^n$ is positive definite (this notation has appeared as early as 1868 in [135] – in connection with the existence of the E_8 lattice; see Section A.6 in the Appendix). The non-strict relaxation of this condition is that of *positive semidefiniteness* of a Hermitian matrix A:

$$x^* A x = \langle Ax, x \rangle \geqslant 0 \ \forall x \in \mathbb{C}^n. \tag{1.1}$$

Classical results by Sylvester and others provide numerous characterizations of this notion:

Theorem 1.1. *The following are equivalent for a complex (resp. real) Hermitian matrix* $A_{n\times n}$*:*

1. *A is positive semidefinite (henceforth termed* **psd***, or simply* positive*):* $x^*Ax \geqslant 0\ \forall x \in \mathbb{C}^n$ *(resp.* $x \in \mathbb{R}^n$*).*
2. *The eigenvalues of A are all in* $[0,\infty)$*.*
3. $A = B^*B$ *for some* $B \in \mathbb{C}^{n\times n}$ *(resp.* $B \in \mathbb{R}^{n\times n}$*).*
4. *There exist vectors in* $\mathbb{C}^n$ *(resp.* $\mathbb{R}^n$*), say* $\mathbf{x}_1,\dots,\mathbf{x}_n$*, such that A is their* Gram matrix*:* $a_{ij} = \langle \mathbf{x}_i, \mathbf{x}_j\rangle$ *for all* $1 \leqslant i,j \leqslant n$*.*
5. *(Sylvester's criterion.) The principal minor* $\det A_{I\times I}$ *is nonnegative, for all* $I \subseteq [n] := \{1,\dots,n\}$*.*

The following fact is also standard.

Lemma 1.2. *Let A be the Gram matrix of any finite set of vectors drawn from $\mathbb{R}^r$. Then A has rank at most r, with equality if and only if the vectors span $\mathbb{R}^r$.*

In this survey, the **question of interest** is to understand the functions that are applied to matrices and preserve positivity. Understanding preservers of mathematical structures is an age-old question; for example, in matrix theory one of the first such results is by Frobenius, who in 1897 classified all determinant-preserving linear maps on matrix algebras [49]. Just as another example: Marcus [100] and Russo–Dye [123] classified linear preservers of the unitary group in $\mathbb{C}^{n\times n}$ and in general C^*-algebras, respectively. However, even after a century of work, basic questions in preservers of positive matrices remain unanswered. As a first example, the linear preservers of positive semidefinite matrices are not fully determined; for this and other aspects of the area, see the survey articles [62, 98] and the monograph [106]. (The goal of this survey is to discuss nonlinear preservers, and we will mention another open question below.)

Definition 1.3. Returning to the question of interest – and removing the linearity constraint – there are two natural ways in which a function acts on a Hermitian matrix $A_{n\times n} = U^*DU$ (by the spectral theorem), with $D = \operatorname{diag}(\lambda_1,\dots\lambda_n)$:

- On its spectrum, via the *functional calculus*: $f(A) := U^*f(D)U$, where $f(D) := \operatorname{diag}(f(\lambda_1),\dots,f(\lambda_n))$; and
- On its entries, via the *entrywise calculus*: $f[A] := (f(a_{ij}))_{i,j=1}^n$.

There is also a second notion of positivity:

$$A = (a_{ij}) \in \mathbb{C}^{n\times n} \text{ is } \textit{entrywise nonnegative} \text{ if } a_{ij} \in [0,\infty)\ \forall i,j \in [1,n]. \tag{1.2}$$

While the functional calculus is the more well-known mechanism, this work will mainly focus on the entrywise calculus. From above, we find four ways in which a function may act on matrix spaces and preserve positivity:

1. $f(A)$ is psd if A is psd.
2. $f[A]$ is entrywise nonnegative if A is so.
3. $f(A)$ is entrywise nonnegative if A is so.
4. $f[A]$ is psd if A is psd.

Notation. Unless otherwise declared, we will henceforth focus on real symmetric matrices, and real-valued functions acting on them.

Now the first two of the four classifications above are easy:

Proposition 1.4. *The functions satisfying conditions (1) or (2) above, are precisely the functions* $f : [0,\infty) \to [0,\infty)$*.*

The third was worked out by Hansen, for all positive matrices with entries in a symmetric interval (for completeness, we also mention the related work [10] of Bharali–Holtz):

Theorem 1.5 ([67, Theorem 3.3(1)]). *Fix $0 < \rho \leqslant \infty$, and let a function $f : (-\rho, \rho) \to \mathbb{R}$. The following are equivalent.*

(1) If A is any real symmetric matrix (of any dimension) with spectrum in $(-\rho, \rho)$ and nonnegative entries, then $f(A)$ has nonnegative entries.
(2) f is given on $(-\rho, \rho)$ by a convergent power series $\sum_{k \geqslant 0} c_k x^k$ with nonnegative coefficients: $c_k \geqslant 0 \ \forall k \geqslant 0$.

This leaves the fourth and final question, which is the subject of this article:

Question 1.6. *Which entrywise maps preserve positive semidefiniteness (of matrices of all dimensions)?*

2. Schoenberg's theorem and its (classical) variants

We now embark on the study of the question above. Our journey begins exactly one hundred years ago, which is when the question was asked, in the well-known 1925 book [117] of Pólya and Szegő. The authors also supplied a large class of functions that are entrywise positivity preservers – the reason behind this is a celebrated 1911 result of Schur:

Theorem 2.1 (Schur product theorem, [128, Satz VII]). *If $n \geqslant 1$, and two $n \times n$ matrices A, B are positive semidefinite, then so is their* Schur/entrywise product

$$A \circ B := (a_{ij} b_{ij})_{i,j=1}^n. \tag{2.1}$$

Among other things, this theorem is useful in proving that the closed convex cone of positive definite kernels (defined below) on $X \times X$ for any set X is moreover closed under multiplication.

Proof. There are several proofs of this result: for instance, $A \circ B$ is a principal submatrix of the Kronecker product $A \otimes B$, which is positive semidefinite. Alternately, write

$$A = \sum_{i \geqslant 1} \lambda_i v_i v_i^*, \qquad B = \sum_{j \geqslant 1} \mu_j u_j u_j^*, \qquad \text{with all } \lambda_i, \mu_j \geqslant 0$$

by the spectral theorem and Theorem 1.1. As the entrywise/Schur product is bilinear, $A \circ B = \sum_{i,j \geqslant 1} \lambda_i \mu_j (v_i \circ u_j)(v_i \circ u_j)^*$, and this is positive semidefinite from first principles. $\square$

The Schur product theorem helps find entrywise positivity preservers as follows. It is clear from the definition (1.1) that the positive semidefinite (psd) matrices form a closed convex cone: they are stable under sums, positive dilations, and entrywise limits. In addition, by Theorem 2.1 they are also closed under Schur products. Thus, the set of entrywise maps preserving positivity is also closed under these operations. In addition, this set contains the functions $f(x) \equiv 1, x$ – since the latter leaves each psd matrix unchanged, while the former sends it entrywise to the all-ones matrix $\mathbf{1}_{n \times n} := (1)_{i,j=1}^n$, and this rank-one matrix is psd. Thus, the closure of $\{1, x\}$ under sums, positive dilations, products, and limits preserves positivity of matrices of all sizes. This closure is precisely the set of convergent power series with nonnegative coefficients, and this was the 1925 observation of Pólya–Szegő:

Definition 2.2. Given a subset $I \subseteq \mathbb{C}$, define $\mathbb{P}_n(I)$ to be the set of $n \times n$ positive semidefinite matrices with entries in I.

Theorem 2.3 ([117, Problem 37]). *Let $I \subseteq \mathbb{R}$ be an interval, and $f(x) = \sum_{k=0}^{\infty} c_k x^k$ be a power series convergent on I, with all $c_k \in [0, \infty)$. Then the entrywise map $f[-]$ sends $\mathbb{P}_n(I)$ to $\mathbb{P}_n(\mathbb{R})$ for all $n \geqslant 1$.*

2.1. **Matrices with entries in** $[-1, 1]$**.** After stating their result, Pólya–Szegő asked if there are other preservers. This was answered by Schoenberg (a student of Schur and Sanielevici) 17 years later – for continuous functions. The next result – and its later variants stated below – can be considered as collectively forming a deep converse to the Schur product theorem.

Theorem 2.4 ([127, Theorem 2]). *Let $I = [-1, 1]$ and let $f : I \to \mathbb{R}$ be continuous. The following are equivalent.*

(1) The entrywise map $f[-]$ sends $\mathbb{P}_n(I)$ to $\mathbb{P}_n(\mathbb{R})$ for all $n \geqslant 1$.
(2) f is given on I by a convergent power series $\sum_{k=0}^{\infty} c_k x^k$ with all $c_k \geqslant 0$.

Definition 2.5. Note that each such power series $f(x) = \sum_{k \geqslant 0} c_k x^k$ is infinitely differentiable on $I' = (0, 1)$ and satisfies: $f^{(k)} \geqslant 0$ on I'. Such a function is said to be *absolutely monotone/monotonic* on I'.

Remark 2.6. The easy implication is (2) $\implies$ (1), and this is precisely Theorem 2.3. The hard part is (1) $\implies$ (2), which was Schoenberg's contribution. In fact, Schoenberg assumed positivity preservation on an even smaller set – the matrices in $\mathbb{P}_n([-1, 1])$ with all diagonal entries 1, aka *correlation matrices* – and deduced absolute monotonicity. (From this, one can deduce the power series representation – and in particular, real analyticity – using a 1929 theorem of Bernstein [9].) This effort to "reduce the test set" will recur in this section.

Let C_1 denote the multiplicative closed convex cone of entrywise positivity preservers of $\bigcup_{n \geqslant 1} \mathbb{P}_n([-1, 1])$. In a sense, the countably many (rescaled) monomials $\mathbb{R}_{\geqslant 0} x^k$ are the "extreme rays" of C_1 that are continuous.

It is natural to ask what happens if the continuity assumption is removed. In this case, there are two other extreme rays that are discontinuous – but only at the endpoints – and they are generated by the following functions on $[-1, 1]$:

$$f_+(x) := \lim_{n \to \infty} x^{2n} = \mathbb{1}(x = \pm 1), \qquad f_-(x) := \lim_{n \to \infty} x^{2n+1} = \mathbb{1}(x = 1) - \mathbb{1}(x = -1),$$

where $\mathbb{1}(E)$ denotes the indicator of an event/statement E. Thus, nonnegative linear combinations of f_+, f_-, and the functions in Schoenberg's theorem 2.4 are indeed (possibly discontinuous) preservers. A natural question is if there are no others, and this was affirmatively answered in 1978 by Christensen and Ressel:

Theorem 2.7 ([21, Theorems 1,2]). *Let $I = [-1, 1]$ and let $f : I \to \mathbb{R}$. The following are equivalent.*

(1) The entrywise map $f[-]$ sends $\mathbb{P}_n(I)$ to $\mathbb{P}_n(\mathbb{R})$ for all $n \geqslant 1$.
(2) The function f is equal to a convergent power series plus two other terms:

$$f(x) = \sum_{k=0}^{\infty} c_k x^k + c_{-1} \left(\mathbb{1}(x = 1) - \mathbb{1}(x = -1) \right) + c_{-2} \mathbb{1}(x = \pm 1), \qquad x \in I,$$

with $c_k \geqslant 0$ for all $k \geqslant -2$ and $\sum_{k \geqslant -2} c_k < \infty$.

The methodologies employed in showing the above results are different: Schoenberg used spherical integrals and ultraspherical polynomials to prove his result, while Christensen–Ressel's proof was convexity-theoretic, involving Choquet theory and Bauer simplices.

2.2. **Positive definite sequences, Toeplitz matrices on the circle, and Rudin's result.** If one removes the endpoints from the domain, then the assumption of continuity may also be dispensed with, with greater ease. This was first achieved by Rudin in 1959 – here is a reformulation of his result:

Theorem 2.8 ([122, Theorems I,IV])**.** *Suppose $I = (-\rho, \rho)$ where $0 < \rho \leqslant \infty$, and $f : I \to \mathbb{R}$. The following are equivalent.*

(1) The entrywise map $f[-]$ sends $\mathbb{P}_n(I)$ to $\mathbb{P}_n(\mathbb{R})$ for all $n \geqslant 1$.
(2) If $A \in \mathbb{P}_n(I)$ is Toeplitz of rank at most 3, then $f[A] \in \mathbb{P}_n(\mathbb{R})$.
(3) f is given on I by a convergent power series $\sum_{k=0}^\infty c_k x^k$ with all $c_k \geqslant 0$.

Note the significant reduction of the test set in part (2), from all positive matrices of all sizes in part (1) – in the spirit of Remark 2.6.

Rudin was studying preservers of *positive definite sequences*[1], and we now take some time to motivate these, from complex function theory. In 1907, Carathéodory published a solution to the following question:

Characterize all analytic functions f such that $f(0) = 1$ and f maps the unit disk $D(0, 1)$ into the right half-plane $\Re(z) > 0$.

His solution [19] was that if one writes $f(z) = 1 + \sum_{k=0}^\infty (a_k + ib_k) z^k$, then the above condition holds if and only if for each $n \geqslant 1$, the point $(a_1, b_1, \dots, a_n, b_n) \in \mathbb{R}^{2n}$ lies in the convex hull of the curve

$$\{(2\cos\theta, -2\sin\theta, 2\cos 2\theta, \dots, 2\cos n\theta, -2\sin n\theta) \,:\, 0 \leqslant \theta \leqslant 2\pi\}. \tag{2.2}$$

In 1911, Toeplitz observed [140] that the constraints (2.2) can be rephrased algebraically, in terms of the positivity of certain related Hermitian quadratic forms for all $n \geqslant 0$:

$$\sum_{i,j=1}^n c_{i-j} z_i \overline{z_j} \geqslant 0 \ \forall z = (z_1, \dots, z_n) \in \mathbb{C}^n, \qquad \text{where } c_0 = 2,\ c_k = a_k - ib_k,\ c_{-k} = \overline{c_k}\ (k > 0).$$

In other words, the semi-infinite matrix

$$T = (t_{ij}) := (c_{i-j})_{i,j \geqslant 0} \text{ is positive semidefinite.} \tag{2.3}$$

Moreover, Toeplitz's matrix T here has the property that the entries along any "diagonal line" (i.e., parallel to the main diagonal) are all equal – a structure that is now called a *Toeplitz matrix*.

Thus, Carathéodory's solution is equivalent to the notion of a positive definite sequence, and it was the preservers of these that Rudin was classifying in [122]. Rudin's motivations come from Fourier analysis, whose connection to positive definite sequences was established at the same time as Toeplitz's work. Indeed, in 1911, Herglotz also published (independently) the work [69], which showed the equivalence of Toeplitz's conditions to the *trigonometric moment problem*: to characterize all sequences $c_n \in \mathbb{C}$ for $n \in \mathbb{Z}$, for which there exists a nonnegative measure μ on $[-\pi, \pi]$ such that the c_n are its Fourier–Stieltjes coefficients $\int_{-\pi}^{\pi} e^{-in\theta} d\mu(\theta)$ for all n.

Herglotz showed that the answer is precisely the positivity condition (2.3). Thus, positive definite sequences are precisely the Fourier–Stieltjes coefficients of nonnegative measures μ on $S^1 \subset \mathbb{C}$:

[1]This is the second occurrence, after "matrices", of the phrase "positive definite", and we will see one more, which naturally leads to entrywise transforms.

Theorem 2.9 ([69])**.** *A complex sequence $(c_n)_{n\in\mathbb{Z}}$ is the Fourier–Stieltjes coefficient-sequence of a nonnegative measure on S^1 if and only if the Toeplitz matrix $T = (c_{i-j})_{i,j\geqslant 0}$ is positive semidefinite – in other words, $c : \mathbb{Z} \to \mathbb{C}$ is a* positive definite function *(defined below).*

We add for completeness that alongside the solution to the above function theory problem by Carathéodory, and the two equivalent conditions by Toeplitz and Herglotz, comes yet another classical equivalent condition. This is the famous Herglotz–Riesz (integral) representation theorem for the aforementioned analytic maps, proved by both authors independently in 1911. We state a more general version, wherein $f(0)$ need not equal 1:

Theorem 2.10 ([69, 118])**.** *A function $f(z) = u(z)+iv(z)$ is analytic on $D(0,1)$ with image in the closed right half-plane $\Re(z) \geqslant 0$, if and only if there exists a finite positive measure μ on $[0,2\pi]$ such that*

$$f(z) = i \cdot v(0) + \int_0^{2\pi} \frac{e^{i\theta}+z}{e^{i\theta}-z}\, d\mu(\theta).$$

Returning to Rudin: he came to his question about preservers of such sequences in the context of Fourier analysis. He was considering functions operating on spaces of Fourier transforms of L^1 functions on Locally Compact Abelian groups G (such G are abbreviated *LCA* groups), or of measures on G. Rudin studied the torus $G = S^1$, while Kahane and Katznelson were studying similar questions on its dual group $\mathbb{Z}$. The three authors then proved in 1959 with Helson, a "converse Wiener–Levy theorem" in [68]. In the same year, Rudin published his related variant of Schoenberg's theorem [122].

2.3. **Open intervals; Hankel matrices.** We return to the story of entrywise preservers. First note that the rank of the positive Toeplitz matrix in the trigonometric moment problem above, corresponds to the size of the support of the measure μ. Thus, Rudin's Theorem 2.8 shows that working with at most three-point measures suffices to recover real analyticity and absolute monotonicity. This connection between supports of measures and rank constraints of test matrices resurfaced in positivity preservers very recently, and is described below.

Two decades after Rudin's work, a variant of the above results was shown for matrices with strictly positive entries. This was by Vasudeva in 1979:

Theorem 2.11 ([143, Theorem 6])**.** *For $I = (0,\infty)$, the two assertions of Schoenberg's theorem 2.4 are again equivalent.*

Vasudeva's theorem was strengthened twofold very recently: in 2022, Belton–Guillot–Khare–Putinar obtained the same conclusion of absolute monotonicity from hypotheses that were significantly weaker in two ways. First, the domain was changed to $(0,\rho)$ for any $0 < \rho \leqslant \infty$; and the test set in each dimension was once again reduced, this time to *Hankel* matrices of rank at most 2.

Theorem 2.12 ([7, Theorem 9.6 and (proof of) Proposition 8.1])**.** *Suppose $I = (0,\rho)$ or $[0,\rho)$ where $0 < \rho \leqslant \infty$, and $f : I \to \mathbb{R}$. The following are equivalent.*

1. *The entrywise map $f[-]$ sends $\mathbb{P}_n(I)$ to $\mathbb{P}_n(\mathbb{R})$ for all $n \geqslant 1$.*
2. *If $A \in \mathbb{P}_n(I)$ is Hankel of rank at most 2, then $f[A] \in \mathbb{P}_n(\mathbb{R})$.*
3. *f is given on I by a convergent power series $\sum_{k=0}^\infty c_k x^k$ with all $c_k \geqslant 0$.*

Below, we will state the multivariate version of this result. Also note that parallel to Rudin's approach (via Herglotz), in [7] the authors considered transforms of discrete data obtained from a positive measure μ on the real line: *moment sequences.* In other words, a function f sends the kth moment of a positive measure μ on the line, to the kth moment of another positive measure σ_μ. Once again, the rank of the test matrices is governed by the support set of μ, and it turns out that one- and two-point test measures suffice to deduce absolute monotonicity.

The two-sided version of Theorem 2.12, parallel to Rudin's characterization above, was also shown in [7]. Once again, the test set can be greatly reduced in each dimension, from all positive matrices to low rank Hankel matrices.

Theorem 2.13 ([7, Corollary 6.2])**.** *Suppose $0 < \rho \leqslant \infty$ and $I = (-\rho, \rho)$. Given $f : I \to \mathbb{R}$, the following are equivalent.*

(1) The entrywise map $f[-]$ sends $\mathbb{P}_n(I)$ to $\mathbb{P}_n(\mathbb{R})$ for all $n \geqslant 1$.
(2) If $A \in \mathbb{P}_n(I)$ is Hankel of rank at most 3, then $f[A] \in \mathbb{P}_n(\mathbb{R})$.
(3) f is given on I by a convergent power series $\sum_{k=0}^{\infty} c_k x^k$ with all $c_k \geqslant 0$.

Notice that all of the above results are "dimension-free", in that the test set in them consists of matrices of unbounded size. It is interesting that the proof of the last two results above uses a result in *fixed* dimension – see Theorem 4.5. A full account of both of these proofs – which are rather accessible to develop "almost from scratch" – as well as details of the results in multiple sections below, can also be found in the recent monograph [83]. For additional connections and results, see the survey [5].

As a parting note in the real case, understanding positivity preservers immediately leads to characterizing the entrywise maps preserving *monotonicity*:

Definition 2.14. The *Loewner ordering* on real symmetric $n \times n$ matrices is: $A \geqslant B$ if $A - B \in \mathbb{P}_n(\mathbb{R})$. Now given $I \subseteq \mathbb{R}$, a function $f : I \to \mathbb{R}$ is said to be (a) *Loewner positive on* $\mathbb{P}_n(I)$ if $f[A] \geqslant 0$ whenever $A \in \mathbb{P}_n(I)$ (i.e., $A \geqslant 0$); and (b) *Loewner monotone on* $\mathbb{P}_n(I)$ if $f[A] \geqslant f[B]$ whenever $A \geqslant B \geqslant 0$.

In this language, the above "Schoenberg-type theorems" classify the Loewner positive maps. Setting $B = 0$, it is also clear that if $0 \in I$ and f is Loewner monotone, then $f - f(0)$ is Loewner positive. In fact the reverse implication also holds – and is easy to show using the Schur product theorem, once we know the Schoenberg–Rudin theorem above. Thus, we have:

Theorem 2.15 ([83, Theorem 19.2])**.** *Suppose $0 < \rho \leqslant \infty$ and $I = (-\rho, \rho)$. The following are equivalent for an arbitrary map $f : I \to \mathbb{R}$:*

(1) $f[-]$ is Loewner monotone on $\mathbb{P}_n(I)$ for all $n \geqslant 1$.
(2) $f[-]$ is Loewner monotone on the Hankel matrices in $\mathbb{P}_n(I)$ of rank at most 3, for all $n \geqslant 1$.
(3) f is given on I by a power series $\sum_{k=0}^{\infty} c_k x^k$, with $c_k \geqslant 0$ for all $k > 0$ (and any $c_0 \in \mathbb{R}$).

The reader can compare and contrast this to Loewner's celebrated characterization of matrix monotone maps in the functional calculus [99, 133].

2.4. **Complex domains.** We conclude this section by discussing two results proved in the literature for positivity preserving maps acting on complex Hermitian matrices. Following his proof for preservers of real positive matrices, Rudin [122] made an observation parallel to that of Pólya–Szegő, for complex matrices. Namely, Rudin observed that the conjugation map $z \mapsto \overline{z}$ also preserves positive semidefiniteness. Since the preservers form a multiplicatively closed convex cone, it follows that the entrywise functions

$$z \mapsto z^j \overline{z}^k \qquad (j, k \geqslant 0)$$

also preserve positivity on complex matrices of all sizes. Again taking nonnegative linear combinations, followed by limits, yields a large family of preservers; and Rudin conjectured in his 1959 work [122] that there are no others. This was shown by Herz soon after, in 1963:

Theorem 2.16 ([73, Théorème 1])**.** *Denote by $D(0,1)$ the open unit disk in $\mathbb{C}$, and say $f : D(0,1) \to \mathbb{C}$. The following are equivalent.*

(1) The entrywise map $f[-]$ sends $\mathbb{P}_n(D(0,1))$ to $\mathbb{P}_n(\mathbb{C})$ for all $n \geqslant 1$.

(2) The function f is a convergent power series, of the form $f(z) = \sum_{j,k \geqslant 0} c_{jk} z^j \overline{z}^k$ for $z \in D(0,1)$, with $c_{jk} \geqslant 0\ \forall j,k \geqslant 0$. Moreover, such a representation for f is unique.

A final result along these "classical lines" returns to the "Schoenberg" version of the complex setting, of functions on the closed unit disk. The entrywise positivity preservers were classified in this setting by Christensen–Ressel in 1982, under Schoenberg's assumption of continuity:

Theorem 2.17 ([22, Corollary 1])**.** *Let $f : \overline{D}(0,1) \to \mathbb{C}$ be a continuous map on the closed unit disk. Let $\mathcal{H}$ be an infinite-dimensional complex Hilbert space, with unit sphere S. The following are equivalent.*

(1) f is a "positive definite kernel" on S – i.e., for any finite set of points $z_1, \dots, z_n \in S$, the matrix with (i,j) entry $f(\langle z_i, z_j \rangle)$ is positive semidefinite.

(2) As in the previous result: f is a convergent power series of the form $f(z) = \sum_{j,k \geqslant 0} c_{jk} z^j \overline{z}^k$ for $z \in \overline{D}(0,1)$, with $c_{jk} \geqslant 0\ \forall j,k \geqslant 0$. Such a representation for f is unique.

As a consequence, these are also equivalent to:

(3) $f[-] : \mathbb{P}_n(\overline{D}(0,1)) \to \mathbb{P}_n(\mathbb{C})$ for all $n \geqslant 1$.

3. Schoenberg's motivations: distance geometry

We now discuss the classical motivations of Schoenberg in arriving at Theorem 2.4, which is a seminal result that has spawned much subsequent activity (not only in generalizing and refining it, but in other, modern domains as well, as is mentioned below). Schoenberg was motivated by the study of *metric/distance geometry.* Indeed, since the advent of Descartes in the 1600s, it has been the norm to think of vectors in Euclidean space $\mathbb{R}^n$ as n-tuples of real numbers, and of distances between them via the Pythagorean metric. However, for almost two millennia before Descartes, geometry meant working with Euclid's postulates and using points, lines, angles, distances, etc.

In the early 20th century, there was a revival of this "distance geometry" perspective. We single out the Vienna Circle of mathematicians and philosophers, in which Karl Menger (a student of Hahn), Tarski, Hahn, von Neumann, Gödel, Taussky, and others met regularly to discuss mathematics. A prevalent theme of their Kolloquium (1928–1936, see [103]) was the study of metric spaces and of properties intrinsic to them, such as curvature, homogeneity, and metric convexity. Indeed, the foundations of metric space theory had been then-recently established, due to works of Birkhoff, Fréchet, and Hausdorff among others, and the Vienna Kolloquium's investigations led them to advances not only in mathematics, but also in mathematical economics (von Neumann's fixed point theorem – the precursor to the one by Kakutani that was later used by Nash – is a case in point, as is the work of Abraham Wald).

3.1. Metric embeddings.

Let $X = \{x_0, \dots, x_n\}$ be a finite set endowed with a metric $d = d_X$. A seminal 1910 result by Fréchet states:

Theorem 3.1 ([46])**.** *Every metric space of size $n+1$ isometrically embeds into $(\mathbb{R}^n, \ell_\infty)$ – where $\ell_\infty(x,y) := \sup_i |x_i - y_i|$ denotes the sup-norm.*

Remark 3.2. In fact the embedding Fréchet provides is reminiscent of – and the precursor to – the Kuratowski embedding [93]. Fréchet's result was subsequently improved by Witsenhausen [148] to $(\mathbb{R}^{n-1}, \ell_\infty)$ if $n \geqslant 2$.

It is now natural to ask which finite, or separable, metric spaces embed into "usual" Euclidean space $\mathbb{R}^n$, or into their limit $\mathbb{R}^\infty = \bigcup_{n \geqslant 1} \mathbb{R}^n$, or into its closure $\ell^2(\mathbb{R})$. Following characterizations by Menger [102] and Fréchet [47], Schoenberg provided in 1935 a characterization that related metric geometry and matrix positivity:

Theorem 3.3 ([125, Theorem 1])**.** *Let a finite metric space $X = \{x_0, \dots, x_n\}$, and write $d_{ij} := d_X(x_i, x_j)$ for $i, j = 0, \dots, n$. Then X embeds isometrically into some Euclidean space (equivalently, into ℓ^2) if and only if its* modified Cayley–Menger matrix

$$CM'(X) := (d_{i0}^2 + d_{j0}^2 - d_{ij}^2)_{i,j=1}^n$$

is positive semidefinite. Moreover, if this happens then the smallest dimensional space $\mathbb{R}^r$ into which X embeds is precisely the rank of $CM'(X)$.

Proof. We outline only one direction, which is the illuminating calculation relating distances and inner products: if X is Euclidean, so that we identify each x_i with an isometrically embedded copy in $\mathbb{R}^k$, then

$$\begin{aligned} CM'(X)_{ij} = d_{i0}^2 + d_{j0}^2 - d_{ij}^2 &= \|x_i - x_0\|^2 + \|x_j - x_0\|^2 - \|(x_i - x_0) - (x_j - x_0)\|^2 \\ &= 2\langle x_i - x_0, x_j - x_0 \rangle. \end{aligned} \tag{3.1}$$

But these form a Gram matrix, which is positive semidefinite (see Theorem 1.1). □

Parallel to flat space embeddings, Schoenberg also characterized when a finite metric space embeds into a Euclidean sphere of unit radius, $S^{r-1} \subset \mathbb{R}^r$. Recall that on any such sphere (for any r), any two antipodal points have distance π, while two non-antipodal points x, y and the origin lie on a unique plane, which slices the sphere along a great circle. Now the intrinsic spherical distance $\sphericalangle x, y$ equals the length of the shorter arc joining them:

$$\sphericalangle x, y := \arccos\langle x, y \rangle, \qquad \text{i.e.,} \qquad \langle x, y \rangle = x \cdot y = \cos \sphericalangle x, y, \qquad \forall x, y \in S^{r-1}. \tag{3.2}$$

This metric exists on each S^{r-1}, hence on their union over $r \geqslant 2$, and hence on its closure – which is the unit sphere $S^\infty \subset \ell^2$. This means that applying $\cos(\cdot)$ entrywise to a spherical distance matrix $(\sphericalangle x_i, x_j)_{i,j=0}^n$ in S^∞ yields a Gram matrix. The converse is also not hard to show, leading to another 1935 result of Schoenberg:

Theorem 3.4 ([125, Theorem 2])**.** *A finite metric space $X = \{x_0, \dots, x_n\}$ embeds isometrically into a Euclidean unit sphere (with its intrinsic angle metric) if and only if X has diameter $\leqslant \pi$ and the entrywise map* $\cos[-]$ *sends its distance matrix D_X to a positive semidefinite matrix* $\cos[D_X]$*. Moreover, the smallest Euclidean dimension r for which X embeds spherically in S^{r-1} is the rank of* $\cos[D_X]$.

3.2. **Positive definite functions.** We make two observations about Theorem 3.4: (a) matrix positivity again plays an indispensable role in it; and (b) the theorem features an early occurrence of the entrywise calculus: maps that take distance matrices to positive ones; or via composing with the metric, two-variable symmetric maps that take $X \times X$ to $\mathbb{P}_{|X|}(\mathbb{R})$. This notion was abstracted into

Definition 3.5.

(1) (Schoenberg, 1938, [126].) A *positive definite function* on a metric space (X, d_X) is a map $f : [0, \infty) \to \mathbb{R}$, such that for any points $x_1, \dots, x_n \in X$, the matrix $(f(d_X(x_i, x_j)))_{i,j=1}^n$ is positive semidefinite.

(2) (Mercer, 1909, [104]; Mathias, 1923, [101].) This is the "traditional" – and different – definition: a *positive definite function* on a group G is a map $f : G \to \mathbb{C}$ such that for any points $g_1, \dots, g_n \in G$, the matrix $(f(g_i^{-1} g_j))_{i,j=1}^n$ is positive semidefinite.

Indeed, it is the latter notion that appears in the Carathéodory–Toeplitz–Herglotz results discussed after Theorem 2.8 above. We again digress here with a discussion of the early history of such functions, following the comprehensive survey [137] by James Stewart.[2] Stewart was a student of J.L.B. Cooper (who descended from Titchmarsh and hence Hardy), and the survey is an offspring of his doctoral dissertation on "Positive definite functions and generalizations."

As the name suggests, positive definite functions send "squares of domain sets" to psd matrices. Indeed, following the use of Smith [135] and others of "positive definite" for matrices, it was Mercer [104] who extended in 1909 the notion to *kernels*: these are maps $f : X \times X \to \mathbb{C}$ for an arbitrary set X, such that $f(x, x') = \overline{f(x', x)}$ and the quadratic form induced by f is positive semidefinite. Examples of such kernels had also appeared in Hilbert's 1904–1910 articles on integral equations [76].

Later, Mathias [101] independently rediscovered in 1923 such kernels – on the group $(\mathbb{R}, +)$. He called them "positive definite"; observed using the Schur product theorem that they form a multiplicatively closed convex cone; and showed that if f is such a kernel on $(\mathbb{R}, +)$ then its Fourier transform $\widehat{f}(t) := \int_{\mathbb{R}} e^{-itx} f(x) dx$ is nonnegative (if it exists). In a sense this is Bochner's theorem over $\mathbb{R}$, but the connection to a density function on $\mathbb{R}$ was shown by Bochner a decade later:

Theorem 3.6 ([14]). *A continuous map $f : \mathbb{R} \to \mathbb{R}$ is positive definite if and only if there exists a (unique) probability measure μ on $\mathbb{R}$ whose Fourier–Stieltjes transform is f:*

$$f(x) = \int_{\mathbb{R}} e^{itx} d\mu(t).$$

We end this part with two historical remarks.

Remark 3.7. Bochner also proved his eponymous theorem for $G = \mathbb{R}^r$ in the following year [15]; the general version for locally compact abelian (LCA) groups came a few years later, due to Povzner, Raikov, and Weil. Other uses of positive definite kernels included Moore and Aronszajn's works on reproducing kernels (see e.g. [114]), the theory of harmonics on homogeneous spaces (Cartan, Weyl, Itô), and its comprehensive generalization by Krein [91] – which subsumes the preceding three authors' work as well as that of Gelfand and Raikov. See [137, Section 8] for more details. For a more recent survey taking the reader from positive definiteness to harmonic analysis and operator theory, see Shapiro's article [132].

Remark 3.8. Bochner-type theorems characterize positive definite functions on a LCA group G, as Fourier transforms of probability measures on the dual group $\widehat{G}$. From this perspective, one should note that the "general" LCA-form of Bochner's theorem was first proved by Herglotz two decades prior to Bochner – see Theorem 2.9 over the dual groups $(S^1, \mathbb{Z})$.

We will mention Herglotz next in Section A.7 in the Appendix on sphere packings – given that he is credited by Müller with the Addition Theorem for spherical harmonics.

3.3. Reconciling the two notions of positive definiteness.

Having discussed the "group" notion of positive definiteness in Definition 3.5(2), we now come to Schoenberg's work on positive definite functions. He introduced the "metric" variant in Definition 3.5(1) in [126], and showed that a metric space X is Euclidean (or embeds isometrically in ℓ^2) if and only if it

[2] The reader who has taught freshman calculus, may recognize this author for a different reason.

is separable and the Gaussian family $\{\exp(-cx^2) : c > 0\}$ is "metric" positive-definite on X. Schoenberg then turned his attention to such positive-definite functions on (Euclidean and Hilbertian) spheres, in the 1942 paper [127] bearing this title.

We explain briefly here, why in the Euclidean sphere context, Schoenberg's setting reconciles with the "other", group-setting that was introduced by Mercer/Mathias and taken forward by Bochner and others. More precisely, we explain how each positive definite function on the sphere, in Schoenberg's metric setting, is a G-invariant positive definite kernel (which incorporates both the "metric" and "group" settings), and on a two-point homogeneous space G/H. We first introduce the relevant notions.

Definition 3.9.

(1) A compact metric space (X, d_X) is said to be *two-point homogeneous* if there is a compact group G acting on X, such that given points $p, q, r, s \in X$ with $d_X(p, q) = d_X(r, s)$, there exists $g \in G$ with $gp = r, gq = s$.
(2) We also isolate the common structure underlying both notions in Definition 3.5: given a topological space X, a continuous kernel : $X \times X \to \mathbb{C}$ is *positive definite* if given any $k \geqslant 1$ and any points $x_1, \dots, x_k \in X$, the matrix $(K(x_i, x_j))_{i,j=1}^k \in \mathbb{P}_k(\mathbb{C})$.

Remark 3.10. Schoenberg's (continuous) positive definite functions are real-valued kernels which factor through the distance map: $K(\cdot, \cdot) = f \circ d_X(\cdot, \cdot)$ – whereas the Mathias–Mercer definition factors through "fractions in G": $K(\cdot, \cdot) = f \circ \eta(\cdot, \cdot)$, where $\eta(g, h) = g^{-1}h$.

Two-point homogeneity is a strong condition; for instance, setting $q = p, s = r$ implies that G acts transitively on X. Let us fix a basepoint $x_0 \in X$ and denote $H := \mathrm{Stab}_G(x_0)$ (a compact subgroup of G), so that $X = G/H$.

Remark 3.11. If G is infinite and connected, then X turns out to be a rank-one Riemannian symmetric space of compact type; these have been classified by Wang [147]. In this case, (G, H) is also an example of a Gelfand pair [53].

Let us now turn our attention to a distinguished **special case**:

Proposition 3.12. *Let $r \geqslant 2$ and let $X = (S^{r-1}, \sphericalangle)$, the unit sphere in $\mathbb{R}^r$. This is a two-point homogeneous space as above, with $G = SO(r)$, and $H = SO(r-1) \oplus \{1\}$ for $x_0 = \mathbf{e}_r$, the north pole.*

Proof. Let $u \in S^{r-1}$ be any unit vector; then u can be completed to an ordered orthonormal basis $(u_1, \dots, u_{r-1}, u)$ such that the orthogonal matrix $U = [u_1 | \cdots | u_{r-1} | u]$ has determinant one. Thus $SO(r)$ acts transitively on S^{r-1}, and the stabilizer of $x_0 = \mathbf{e}_r$ is $SO(r-1) \oplus 1$. Note that G acts on $X = S^{r-1}$ by isometries, since $\langle \cdot, \cdot \rangle$ is $O(r)$-invariant.

It remains to show two-point homogeneity. Given $a, b, c, d \in S^{r-1}$ with $\sphericalangle a, b = \sphericalangle c, d$, we may first apply the G-transitive action to assume that $a = c = \mathbf{e}_r$. We will use here – and below – the basic fact (3.2) that the spherical distance on S^{r-1} is connected to the inner product. Thus, writing $b = (b', b_r), d = (d', d_r)$ as tuples in $\mathbb{R}^r$, it follows that $b_r = d_r$, and hence $\|b'\| = \|d'\|$. From above, there exists $g \in SO(r-1) \oplus \{1\}$ sending $b' \mapsto d'$ and hence $b \mapsto d$. Therefore $g \in H$ sends $b, \mathbf{e}_r$ to $d, \mathbf{e}_r$, respectively. ☐

We next show:

Lemma 3.13. *Let $X = S^{r-1}, G, H, x_0$ be as above, and let K be a continuous kernel on $X \times X$ into any codomain C. Then K is invariant under the diagonal action of G if and only if $K(x, y) \equiv \psi(x \cdot y)\ \forall x, y \in X$ for some continuous map $\psi : [-1, 1] \to C$.*

Proof. The sufficiency is clear. Conversely, two pairs (x, y) and (x', y') in $(S^{r-1})^2$ are G-conjugate if and only if $\sphericalangle x, y = \sphericalangle x', y'$, if and only if $x \cdot y = x' \cdot y'$. ☐

We can now conclude the analysis. A continuous map $f : [0, \pi] \to \mathbb{R}$ is a positive definite function on $(S^{r-1}, \sphericalangle)$ in the Schoenberg/metric sense, if and only if

$$\psi := f \circ \cos^{-1} : [-1, 1] \to \mathbb{R} \tag{3.3}$$

sends Gram matrices from S^{r-1} to positive semidefinite matrices. By Lemma 3.13, this is equivalent to a continuous G-invariant positive definite kernel $K : S^{r-1} \times S^{r-1} \to \mathbb{R}$. □

3.4. **Schoenberg: from positive definite functions on spheres to positivity preservers.** Having reconciled the two notions of positive definiteness (and subsumed them by kernels in Definition 3.9), we conclude this section on Schoenberg's motivations. As we saw above, Schoenberg began his studies in metric geometry by understanding metric embeddings into Euclidean spaces and spheres. The latter case naturally led him to introduce "metric" positive-definite functions, which he then studied (in the cited and other papers).

Schoenberg then classified in 1942 all continuous positive definite functions on S^{r-1} for each $r \geqslant 2$ (which he wrote composing with $\cos^{-1}$, so with domain $x \cdot y \in [-1, 1]$).

Theorem 3.14 ([127, Theorem 1])**.** *Fix an integer $r \geqslant 2$. Given a continuous map $f : [-1, 1] \to \mathbb{R}$, the following are equivalent:*

(1) $f \circ \cos$ is positive definite on $(S^{r-1}, \sphericalangle)$.
(2) The map $f(x) = \sum_{k=0}^{\infty} c_k C_k^{(r)}(x)$, where $c_k \geqslant 0$ $\forall k$, and for varying r, the functions $\{\frac{C_k^{(r)}(x)}{C_k^{(r)}(1)} : k \geqslant 0\}$ are precisely the first Chebyshev, Legendre, or Gegenbauer orthogonal polynomials.

Below, we reinterpret Schoenberg's result from a modern perspective; and further below in the Appendix – see Section A.7 – we will derive the "easier" half (2) $\implies$ (1) from the theory of spherical harmonics. Here we continue to the *Hilbert sphere* S^∞, which is the closure in ℓ^2 of $\bigcup_{r \geqslant 2} S^{r-1}$. Continuous positive definite functions on S^∞ are in a sense the "intersection" of the maps in Theorem 3.14 over all r, and Schoenberg showed in the same work:

Theorem 3.15 ([127, Theorem 2])**.** *A continuous map $f : [-1, 1] \to \mathbb{R}$ is positive definite on S^∞ if and only if*

$$f(\cos\theta) = \sum_{k \geqslant 0} c_k \cos(\theta)^k, \qquad \theta \in [0, \pi]$$

where all $c_k \geqslant 0$ and $\sum_k c_k < \infty$.

One can free this result from the sphere context by setting $x = \cos\theta$. The result then translates into: $f[-]$ sends spherical Gram matrices to positive semidefinite matrices (see (3.3) and the next lines) if and only if $f(x) = \sum_{k \geqslant 0} c_k x^k$ on $[-1, 1]$. (See Section 4.4 for a longer explanation.) This is precisely Schoenberg's theorem 2.4 on entrywise positivity preservers.

3.5. **From Schoenberg's positive definite functions to sphere packings.** Before we discuss the connections of Schoenberg and Rudin's results to applied fields, and a modern perspective on positivity preservers, we mention that Schoenberg's theorem 3.14 has another interesting and important application: to the old problem of sphere packings in any dimension. As this is not directly related to the question of preserving positivity, we defer this discussion to the Appendix; the connection to Schoenberg's result is given in Section A.7.

4. Modern motivations; fixed dimension

Interest in positivity preservers has also been renewed in recent times, because of their relevance in modern statistical methodology to analyze high-dimensional data. In classical statistics, one typically considers a few random variables (p) and has a large number of observations/data points of them. That is, the sample size n is much larger than p, which renders robustness to traditional statistical estimators. However, in recent times a new paradigm has emerged, wherein the number of random variables is very large ($\sim$100,000) and the number of data points is consequently very small. This motivates strongly, and leads to, many novel results on entrywise positivity preservers in a fixed dimension, as well as their connections to hitherto disparate areas in mathematics: symmetric function theory, finite fields, and graph theory, to name a few. We refer the reader to the survey [6] as well as the monograph [83, Chapter 7] for more on this; here we provide only a short exposition.

4.1. **An application to machine learning.** Before discussing the fixed dimension setting in earnest, we begin with a quick digression into an application of entrywise positivity preservers in the "dimension-free" case, i.e., for matrices of all sizes. In this context, we continue to discuss positive definite kernels on real Hilbert spaces $\mathcal{H}$: as in Theorem 2.17, these are maps $K : \mathbb{R} \to \mathbb{R}$ such that $(K(\langle x_i, x_j \rangle))_{i,j=1}^k$ is positive semidefinite for all choices of points $x_1, \dots, x_k \in \mathcal{H}$.

It turns out that every such map K yields a reproducing-kernel Hilbert space, and this is an important tool in Machine Learning (see, e.g., [116, 136, 142]). By Lemma 1.2, the Gram matrices $(\langle x_i, x_j \rangle)$ are precisely the positive semidefinite matrices (of all sizes) whose rank is bounded above by $\dim \mathcal{H}$. Thus, Rudin's theorem 2.8 yields the complete classification of all such kernels – and in 1959, predating their application in machine learning by decades.

The next subsection describes how Schoenberg's theorem 2.4 also contains in it the seeds of – and similarly predates by decades – the concept of *regularization* in modern high-dimensional data analysis. For now, we remark for completeness that Schoenberg's results on positive definite functions on spheres also have applications to Gaussian random fields, pseudo-differential equations with radial basis functions, and approximating functions and interpolating data on spheres – such as the earth in geospatial modeling. Moreover, Schoenberg's theorem 3.3 on Euclidean embeddings of metric spaces is a crucial ingredient in multidimensional scaling [32].

4.2. **Rudin vs. Schoenberg.** We take a moment to compare and contrast the two results: Rudin's Theorem 2.8 vs. Schoenberg's Theorem 3.14. In both theorems, the test matrices consist of Gram matrices of arbitrarily large finite sets of vectors drawn from a subset $X \subset \mathbb{R}^r$:

- For Rudin, X was the open ball of any fixed radius $0 < \rho \leqslant \infty$, and his classification was independent of the precise value of $r \geqslant 3$. In this case, the positivity preservers on this test set are positive sums of monomials.
- For Schoenberg, X was the unit sphere S^{r-1}, which led to the test matrices being correlation matrices – so all diagonal entries are 1 – and the rank bounded above by r. In this case, the continuous preservers of positivity on this test set are positive sums of Gegenbauer polynomials.

4.3. **Applied fields and regularization.** We now come to the study of preservers in fixed dimension, strongly motivated by applied fields. Let us consider a concrete example: analyzing temperatures on the earth's surface. This has been a subject of intensive study in recent years in climate science. With the advent of technology, there are thousands of weather stations and other locations where one has temperature markers.

Another concrete example involves understanding gene-gene interactions, in trying to detect the early onset/early warning for cardiovascular diseases or cancer, say. There are thousands of genes that are studied today. Alternately, one can consider the behavior over time of financial instruments (e.g. in the stock market) and their interdependencies.

The common theme in all of these examples is of "complex multivariate structures" and trying to understanding their interactions. This is a key challenge in modern applied fields – and one of the simplest measures of dependency between any two such variables is the linear dependency, captured in their covariance or correlation:

$$Cov(X,Y) := \mathbb{E}[(X-\mathbb{E}X)(Y-\mathbb{E}Y)].$$

Covariance analysis has been a leading and robust mechanism for data analysis. The difference now is that the sample covariance matrices

$$\widehat{\Sigma} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})(x_i-\overline{x})^T \tag{4.1}$$

are enormous in dimension, since there are $p \gg 0$ random variables being measured. Moreover, as the sample size n is very small, the $p \times p$ matrix $\widehat{\Sigma}$ is highly singular, which is unfavorable to subsequent statistical analysis.

Thus, various workarounds have been suggested to "improve the properties" of such matrices and render them more amenable to statistical techniques. A popular approach has been to apply iterative methods ("compressed sensing") – some names in this regard are Donoho, Daubechies, Candes, and Tao. While these methods work well for a few thousand random variables, they are too expensive for matrices of order ~100,000. As a result, new methodologies are called for.

One alternative in the field of (ultra-)high dimensional covariance estimation, which has emerged in recent times, is to *regularize* sample covariance matrices. In other words, one applies a regularizing function *entrywise* to each covariance or correlation.

Example 4.1. A popular regularizer is *hard-thresholding*. Suppose the true covariance matrix of the population (or of a probability distribution) underlying the sampled data, is: $\Sigma = \begin{pmatrix} 1 & 0.2 & 0 \\ 0.2 & 1 & 0.5 \\ 0 & 0.5 & 1 \end{pmatrix}$. Whereas the sample covariance matrix is $\widehat{\Sigma} = \begin{pmatrix} 0.95 & 0.18 & 0.02 \\ 0.18 & 0.96 & 0.47 \\ 0.02 & 0.47 & 0.98 \end{pmatrix}$.

It is then natural to threshold small entries (i.e., change them to zero if their absolute value is below a "threshold"), with the idea that the variables are actually independent but the observed value is noise: $\widetilde{\widehat{\Sigma}} = \begin{pmatrix} 0.95 & 0.18 & \mathbf{0} \\ 0.18 & 0.96 & 0.47 \\ \mathbf{0} & 0.47 & 0.98 \end{pmatrix}$.

Such an operation applies directly on the cone, induces sparsity (number of zero entries), and is scalable because it is entrywise – no iterative or algorithmic procedures required. It also displays good consistency and other properties. (See e.g. [11, 71, 72, 97, 150] for some papers that study regularization.) However, one also needs to ensure that the thresholded (or more generally, regularized) matrix is itself a proxy for the true covariance – and in particular, is positive semidefinite. This is where one has a theoretical gap, in that it is not clear *for which regularizing entrywise maps is the transformed matrix guaranteed to be positive semidefinite.* Thus, while the problem emerges directly and naturally from applied fields, it brings us back squarely to the question of classifying entrywise positivity preservers.

Moreover, there is motivation from this perspective to study the question of preservers in a fixed dimension – because in a given applied problem, the number of random variables (aka the dimension of the problem) is known. So there is no need to study dimension-free preservers; indeed, only studying these restricts one to using power series with nonnegative coefficients to regularize covariance matrices, which is unnecessarily restrictive *and* which does not induce or preserve sparsity. This strongly motivates trying to classify entrywise positivity preservers – aka regularizing maps – in fixed dimension.

4.4. **Positivity preservers in fixed dimension: rank and sparsity constraints.** Resuming the narrative from the previous section, Schoenberg's 1942 classification of positive definite functions can be interpreted today in terms of regularization. Namely, a continuous function is positive definite on S^{r-1} for some r, if and only if – via (3.3) – $f \circ \cos$ sends Gram matrices, of any size / number of vectors but with all vectors in S^{r-1}, to positive semidefinite matrices. By Theorem 1.1 and Lemma 1.2, this is equivalent to the entrywise maps preserving positivity, on matrices of any dimension but with rank at most r.

Similarly, if one lets the dimension grow unbounded, $f : [-1, 1] \to \mathbb{R}$ is positive definite on S^∞ if and only if $f[-]$ is a positivity preserver on correlation matrices (i.e., psd matrices with diagonal entries 1). Thus, Schoenberg's theorems 3.14 and 3.15 classified the regularizers of correlation matrices of any size, but constraining (or not) the *rank*.

The question we are discussing here is different: now the dimension itself is constrained (and hence, so is the rank). This is not only a natural theoretical "next step" after Schoenberg's and Rudin's results, it is also motivated from the modern perspective of big data (as discussed above).

Unfortunately, few results are known in this setting. We mention results along a few of these fronts. First, the problem in fixed dimension $n = 1$ is trivial: the preservers are clearly all functions : $[0, \infty) \to [0, \infty)$. For $n = 2$, the situation is more involved, and was resolved in 1979 by Vasudeva. We write here his result in a slightly more general setting; the proof is virtually unchanged.

Theorem 4.2 ([143]; taken from [83, Theorem 6.7])**.** *Suppose $0 < \rho \leqslant \infty$, and the domain I is either $[0, \rho)$ or $(0, \rho)$. Now an entrywise map $f : I \to \mathbb{R}$ preserves positivity on $\mathbb{P}_2(I)$ if and only if f is nonnegative, nondecreasing, and multiplicatively midconvex on I (this last means that $f(\sqrt{xy})^2 \leqslant f(x)f(y)$ for all $x, y \in I$).*

In particular, if we set $I^+ := I \setminus \{0\}$, then: any such function is continuous on I^+, and is either never zero or always zero on I^+.

The next case is $n = 3$. Remarkably, in this case the problem remains **open**. Thus, we do not have a classification of the dimension-n positivity preservers for any $n \geqslant 3$.

In the absence of a classification for arbitrary functions acting on all matrices in $\mathbb{P}_n$, several refinements have been proposed and studied, yielding "restricted" preserver results. We list a few of these in which the *test matrices* are additionally constrained.

(1) One can impose *rank-constraints*: $f[-]$ sends matrices in $\mathbb{P}_n$ with rank at most l to matrices in $\mathbb{P}_n$ with rank at most k. Rank-constraints are natural in both theory and applications: on the theoretical side, Gram matrices arising from Euclidean spheres S^{r-1} have rank at most r (in Schoenberg's work [127]), and Rudin's theorem 2.8 as well as Theorems 2.12 and 2.13 also had rank constraints, arising from the support sets of measures on the circle and line, respectively. On the applied side, rank constraints arise naturally from the sample size, which is typically small in applications. For results along these lines, see the 2017 paper [60].
(2) Alternately, one can impose *sparsity constraints*: impose zero entries in pre-fixed positions. These positions may be determined from a combinatorial perspective (one forms graphs associated to a zero pattern) or from domain-specific knowledge in

applications, where various random variables are known to be independent or at least uncorrelated, e.g., Gaussian graphical models. For such results, see the work [58].

(3) One can instead also study preservers of structured matrices such as Toeplitz or Hankel matrices, as was done in Theorems 2.8, 2.12, and 2.13.

Independently, one can constrain the set of functions that act on test matrices. We discuss two families; the first consists of *power functions.* These are a natural family of entrywise maps to consider, and they are also used in practice to induce sparsity on matrices by sending very small / spurious observed correlations to (very close to) zero.

By the Schur product theorem, all integer powers $x^k, k \geqslant 0$ entrywise preserve positivity in every dimension. We now turn to non-integer powers – and hence, only consider matrices with positive entries. The following result was shown in 1977 by FitzGerald and Horn:

Theorem 4.3 ([44, Theorem 2.2]). *Given an integer $n \geqslant 2$, the entrywise power x^α, $\alpha \in \mathbb{R}$ preserves positivity on $\mathbb{P}_n((0,\infty))$ if and only if $\alpha \in \mathbb{Z}_{\geqslant 0} \cup [n-2,\infty)$.*

This interesting result has seen variants for preservers of positive matrices with negative entries and variants of power maps; rank constraints; and sparsity constraints. See e.g. [57, 59, 74]. In all of these papers, entrywise powers preserving certain matrix properties are classified, and in all cases the solution set equals the non-negative integers up to some integer C, followed by all real numbers in $[C,\infty)$. This point of phase transition is called the *critical exponent* (for that particular matrix property). In particular, in [59] a new graph invariant is defined and computed for all chordal graphs.

We conclude this part with a significant strengthening of Theorem 4.3. It turns out that there is a multiparameter family of rank two positive matrices, *each of which* encodes the entire set of power preservers of Loewner positivity. This was shown by Jain in 2019:

Theorem 4.4 ([80, Theorem 1.1]). *Given $n \geqslant 2$, choose distinct positive reals $x_1,\dots,x_n$. Also let $\alpha \in \mathbb{R}$. Then x^α entrywise preserves positivity on the matrix $(1+x_ix_j)_{i,j=1}^n$ if and only if $\alpha \in \mathbb{Z}_{\geqslant 0} \cup [n-2,\infty)$.*

4.5. **Connection to symmetric polynomials.** In addition to these modern results, there is essentially only one classical result for entrywise preservers of positivity in a fixed dimension (without any restrictions on the test matrices or the functions). This is due to Loewner, who wrote it in a letter to Josephine Mitchell in 1967 (courtesy: Stanford Library Archives). Later, the result appeared in the 1969 PhD thesis of his student, Roger A. Horn [79], and was subsequently refined in [60, 84]. We state an alternate version from [7, Theorem 4.2]:

Theorem 4.5 ("Stronger" Loewner–Horn, [79, Theorem 1.2]). *Fix a dimension $n \geqslant 3$ and a scalar $0 < \rho \leqslant \infty$, and let $I = (0,\rho)$. Let $f : I \to \mathbb{R}$ be such that $f[-]$ preserves positivity on $\mathbb{P}_2(I)$ and on all Hankel matrices in $\mathbb{P}_n(I)$ of rank at most two. Then $f \in C^{(n-3)}(I)$, with $f, f', \dots, f^{(n-3)}$ nonnegative on I. Moreover, $f^{(n-3)}$ is convex and nondecreasing on I.*

Interestingly, this result, combined with Bernstein's theorem on absolutely monotone functions [9], provides a pathway to proving the dimension-free Schoenberg–Rudin theorem for preservers of positive matrices with *positive* entries. See e.g. [7] for this treatment in one and several variables.

The proof of Theorem 4.5, originally due to Loewner, contains the seeds of a surprising connection to symmetric function theory. We now describe this connection, which was discovered only about ten years ago, and is now better understood.

Recall the century-old observation by Pólya–Szegő (or a special case thereof) from 1925: if $f(x)$ is a polynomial with nonnegative coefficients, then $f[-]$ entrywise preserves $\mathbb{P}_n(\mathbb{R})$ for all $n \geqslant 1$. Thus, if one seeks preservers of $\mathbb{P}_n(\mathbb{R})$ or of $\mathbb{P}_n((-\rho,\rho))$ for fixed n, then one should expect more polynomial preservers: ones which have negative coefficients. However, apart from Vasudeva's 2×2 characterization (see above), *no* example of such a polynomial was known for almost a century. Certainly, work of Fischer–Stegeman [43] had shown that if such a preserver on $\mathbb{P}_n((-\rho, \rho))$ is to exist, the n nonzero coefficients of lowest degrees should be positive, as should those of highest degrees if $\rho = \infty$. This is "morally" along the lines of the Loewner–Horn theorem 4.5. The question remains:

Can at least one other coefficient be negative? More generally, which coefficients in a polynomial preserver of $\mathbb{P}_n((-\rho, \rho))$ can be negative, for fixed $n \geqslant 1$ and $\rho \in (0, \infty)$?

This was first answered in 2016 for a special class of polynomials by Belton–Guillot–Khare–Putinar [4], and then in 2021 for all polynomials by Khare–Tao:

Theorem 4.6 ([86])**.** *Fix an integer $n \geqslant 1$ and a scalar $0 < \rho < \infty$. If a polynomial $f(x)$ with real coefficients is an entrywise positivity preserver of $\mathbb{P}_n((-\rho, \rho))$, then its first n coefficients (of lowest degree) are positive. Moreover, it is possible for every other coefficient to be negative.*

If f has precisely $n + 1$ nonzero coefficients, there exists a closed-form expression for the negative threshold for the unique negative coefficient possible (which is the leading term).

As the focus of this article is on studying the dimension-free preservers (originally classified by Schoenberg and then Rudin), we do not provide more details here, referring the reader to the aforementioned papers, as well as the survey [6] and the monograph [83]. We mention, however, that the determinantal calculation that is at the heart of obtaining the closed-form threshold in Theorem 4.6, as well as at the heart of Loewner's theorem 4.5, is as follows:

Given a smooth function $f(t)$, and real scalars u_i, v_i for $1 \leqslant i \leqslant n$, compute the Taylor coefficients of the $n \times n$ determinant $\Delta(t) := \det f[(tu_i v_j)_{i,j=1}^n]$.

The first $\binom{n}{2} + 1$ such derivatives/Taylor coefficients were worked out by Loewner in the 1960s. However, *all* Taylor coefficients had been worked out in the special case $f(t) = 1/(1-t)$ by Cauchy [20], more than a century ago! This is the famous Cauchy Determinantal Formula, which is an important result in symmetric function theory, and which turns out to involve *Schur polynomials* in the (x_i) and in the (y_j) separately. In the 1880s, Frobenius generalized this to f a sum of two geometric series [48].

Eventually, this question was settled by Khare in full generality – in both the analytic and algebraic settings. Here is the latter result (and it subsumes the calculations by Cauchy, Frobenius, and Loewner):

Theorem 4.7 ([84, Theorem 2.1])**.** *Fix a commutative unital ring R, and let t be an indeterminate. Let $f(t) := \sum_{M \geqslant 0} f_M t^M \in R[[t]]$ be an arbitrary formal power series. Given vectors $\mathbf{u}, \mathbf{v} \in R^N$ for some $N \geqslant 1$, we have:*

$$\Delta(t) := \det f[t\mathbf{u}\mathbf{v}^T] = V(\mathbf{u})V(\mathbf{v}) \sum_{M \geqslant \binom{N}{2}} t^M \sum_{\mathbf{n} = (n_N, \dots, n_1) \,\vdash M} s_{\mathbf{n}}(\mathbf{u}) s_{\mathbf{n}}(\mathbf{v}) \cdot \prod_{j=1}^N f_{n_j},$$

where $V(\mathbf{u}) := \prod_{i<j}(u_i - u_j)$ is the Vandermonde determinant for $\mathbf{u}$, and similarly for $V(\mathbf{v})$.

This result (or its analysis counterpart, wherein the inner sum is precisely the Mth Maclaurin coefficient of $\Delta(t)$) yields the interesting conclusion that *every smooth function / power series "gives rise to" all Schur polynomials.* This bridges analysis and symmetric function theory, and also helps explain the appearance of Schur polynomials in the preserver problem in fixed dimension in [4, 86].

Later, Khare and Sahi [85] worked out the analogue of Theorem 4.7 for the matrix permanent $\mathrm{perm} f[(tu_i v_j)_{i,j=1}^n]$, and in fact for all irreducible characters – even more generally, all complex class functions – of the symmetric group and of every subgroup. Thus, the entrywise calculus also connects (surprisingly) to group representations and symmetric functions.

As a concluding trivia, we sketch in the next figure the closely knit academic lineage of several of the experts in this area, having mentioned some of their contributions above.[3]

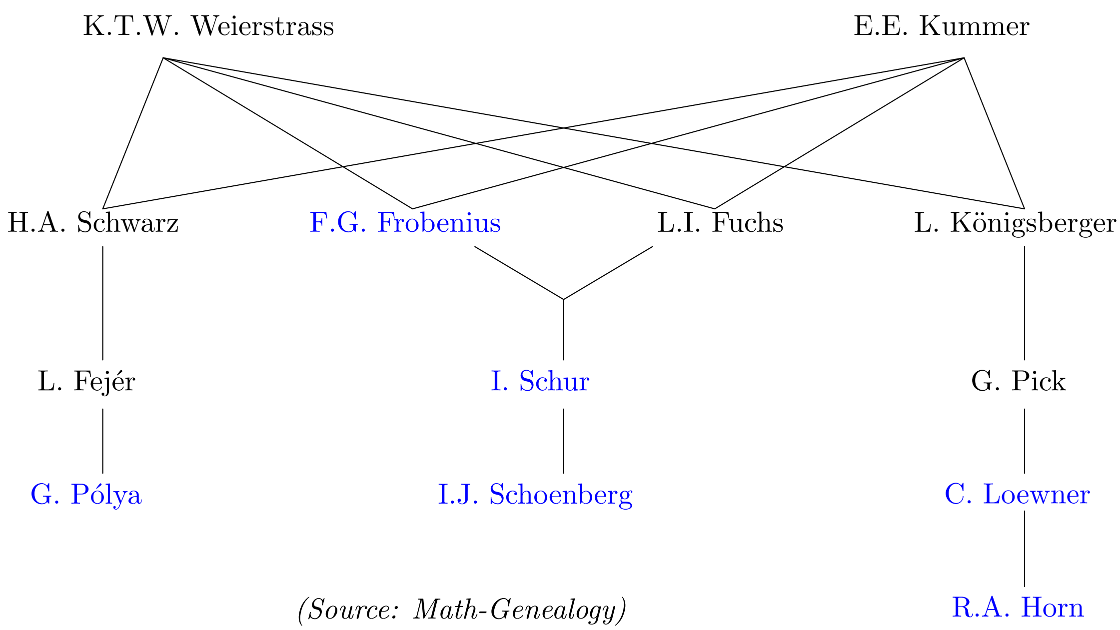

FIGURE 4.1. Math-Genealogy of some of the experts in positivity, its preservers, and connections

4.6. **Acting only on off-diagonal entries.** A variant of Schoenberg's theorem 2.4, with a modern twist, is as follows. Recall that Schoenberg was classifying the entrywise maps sending Gram matrices to themselves – equivalently, sending covariance matrices to themselves. Now if the test matrices are correlation matrices (i.e., Gram matrices of vectors on S^∞), then one may want to preserve the self-correlations 1 along the diagonals, while regularizing the other correlations. Thus, a natural variant of the entrywise action is as follows:

Definition 4.8. Given a domain $I \subseteq \mathbb{R}$, a function $f : I \to \mathbb{R}$, and a square matrix $A = (a_{ij})$, define $f^*[A]$ to have diagonal entries a_{ii} and other entries $f(a_{ij})$.

In 2015, Guillot and Rajaratnam showed that, perhaps surprisingly, the dimension-free off-diagonal entrywise preservers are once again solutions to Schoenberg's theorem – but with an additional constraint:

Theorem 4.9 ([61, Theorem 4.21])**.** *Fix a scalar $0 < \rho \leqslant \infty$ and let $I := (-\rho, \rho)$. The following are equivalent for a function $f : I \to \mathbb{R}$:*

(1) $f^*[-]$ preserves positive semidefiniteness on $\mathbb{P}_n(I)$ for all sizes $n \geqslant 1$.
(2) The function $f(x) = \sum_{k \geqslant 0} c_k x^k$ on I for some scalars $c_k \geqslant 0$, and such that $|f(x)| \leqslant |x|$ on I. (So if $\rho = \infty$, then $f(x) \equiv cx$ on $\mathbb{R}$ with $c \in [0, 1]$.)

[3]Fejér had a remarkable list of PhD students – here we name some others who feature in this article and the Appendix: Paul Erdös, László Fejes Tóth, Pál Turán, and John von Neumann.

This line of inquiry was taken forward by Vishwakarma [146], who generalized the problem in two ways. First, the function f now avoids prescribed principal submatrices / diagonal *blocks* in each dimension, not just 1×1 blocks / diagonal entries; and second, on these diagonal blocks a different function g acts. Vishwakarma classified most of these cases, for $g(x) = \alpha x^k$ with $\alpha \in (0, \infty)$ and $k \in \mathbb{Z}_{\geqslant 0}$. Here is a special case of his main result.

Theorem 4.10 ([146]; taken from [83, Theorem 20.3])**.** *Fix $0 < \rho \leqslant \infty$, and let $I = (-\rho, \rho)$ and $f, g : I \to \mathbb{R}$. Now fix for each $n \geqslant 1$ a collection T_n of subsets of $[n] := \{1, \dots, n\}$ – i.e., $T_n \subseteq 2^{[n]}$. Given $A \in I^{n \times n}$, define $(g, f)_{T_n}[A] \in \mathbb{R}^{n \times n}$ to be the matrix with (i, j) entry $g(a_{ij})$ if there exists $E \in T_n$ with $i, j \in E$. If no such E exists then set $(g, f)_{T_n}[A]_{ij} := f(a_{ij})$.*

Now assume $T_n \not\subseteq \{\{1\}, \dots, \{n\}\}$ for some $n \geqslant 3$; and T_n partitions a subset of $[n]$ for all $n \geqslant 1$. If $g(x) = \alpha x^k$ with $\alpha \in (0, \infty)$ and $k \in \mathbb{Z}_{\geqslant 0}$, and $(g, f)_{T_n}[A] \in \mathbb{P}_n(\mathbb{R})$ for all $A \in \mathbb{P}_n(I)$, then there are three cases:

1. *If for all $n \geqslant 3$ we have $T_n = \{[n]\}$ or $\{\{1\}, \dots, \{n\}\}$, then f is a convergent power series with nonnegative coefficients, and $0 \leqslant f \leqslant g$ on $[0, \rho)$.*
2. *If T_n is not a partition of $[n]$ for some $n \geqslant 3$, then $f(x) = cg(x)$ for some $c \in [0, 1]$.*
3. *If neither (a) nor (b) holds, then $f(x) = cg(x)$ for some $c \in [-1/(K-1), 1]$, where*

$$K := \max_{n \geqslant 1} |T_n| \in [2, +\infty].$$

The first case is akin to Schoenberg's theorem, in the form of the final solution set. The next case is much more restrictive; but it is the third part of the result that is striking. Recall the profusion of Schoenberg-type results above, which unanimously reveal absolutely monotonic functions (power series with nonnegative coefficients) as the dimension-free preservers. Nevertheless, part (3) – in the special case $g(x) \equiv x$ – reveals the possibility of $f(x) \equiv cx$ with $c < 0$. This is the first – and to date, the only – instance of a dimension-free setting, in which the positivity preserver is not absolutely monotonic.

5. Allowing negative eigenvalues; multivariate versions

5.1. **Preservers of matrices with negative inertia.** We now discuss some results that are mostly taken from very recent preprints, beginning with negative inertia preservers. Having discussed entrywise maps preserving matrices with all nonnegative eigenvalues, it is natural to ask what happens if one allows a few negative eigenvalues. This was worked out by Belton–Guillot–Khare–Putinar in [8], and we present a few of the findings.

Definition 5.1. Given integers $n \geqslant 1$ and $0 \leqslant k \leqslant n$, and a domain $I \subseteq \mathbb{C}$, let $\mathcal{S}_n^{(k)}(I)$ denote the Hermitian $n \times n$ matrices with all entries in I and exactly k negative eigenvalues.

Also denote by $\overline{\mathcal{S}_n^{(k)}}(I)$ the "closure", wherein the entries still stay in I but the *negative inertia* (i.e., number of eigenvalues < 0) is at most k: $\overline{\mathcal{S}_n^{(k)}}(I) = \bigsqcup_{j=0}^{k} \mathcal{S}_n^{(j)}(I)$.

Finally, the *inertia* of an $n \times n$ Hermitian matrix is the triple (n_-, n_0, n_+), where the coordinates denote the numbers of negative, zero, and positive eigenvalues, respectively.

Our goal is to understand the entrywise preservers of negative inertia on $\bigcup_{n \geqslant k} \mathcal{S}_n^{(k)}(I)$ for all integers $k \geqslant 0$. Note that the case of $k = 0$ is precisely the (dimension-free) Schoenberg–Rudin theorem above. We state the next result only for $I = (-\rho, \rho)$ (and in it, compute the inertia preservers as well). The cases of $I = (0, \rho)$ and $[0, \rho)$ will be presented through their multivariate versions below.

Theorem 5.2 ([8, Theorems 1.2 and 1.3]). *Fix an integer $k \geqslant 0$ and a scalar $0 < \rho \leqslant \infty$. Let $I = (-\rho, \rho)$ and $f : I \to \mathbb{R}$.*

(1) *Then $f[-]$ preserves the inertia of all matrices in $\mathcal{S}_n^{(k)}(I)$ for all $n \geqslant k$ if and only if $f(x) \equiv cx$ for some $c > 0$.*
(2) *The map $f[-]$ preserves the negative inertia of all matrices in $\mathcal{S}_n^{(k)}(I)$ for all $n \geqslant k$ if and only if:*
   (a) *$k = 0$: This is the Schoenberg–Rudin theorem, and f must be a convergent power series on I with nonnegative Maclaurin coefficients.*
   (b) *$k = 1$: $f(x) \equiv cx$ or $f(x) \equiv -c$, for some $c > 0$.*
   (c) *$k \geqslant 2$: $f(x) \equiv cx$ for some $c > 0$.*

Thus, the class of dimension-free (negative) inertia preservers is very rigid for $k > 0$. It turns out that a more interesting question is to classify the entrywise maps sending $\mathcal{S}_n^{(k)}(I)$ to $\overline{\mathcal{S}_n^{(l)}}$. This too admits a complete solution, whose multivariate version is perhaps more clarifying to state. For now, we provide the meat of the assertion in the form of a summary:

Theorem 5.3 ([8, Theorem A]). *Fix nonnegative integers k, l and a scalar $0 < \rho \leqslant \infty$. Let $I = (-\rho, \rho)$ and $f : I \to \mathbb{R}$ be such that $f[-] : \mathcal{S}_n^{(k)}(I) \to \overline{\mathcal{S}_n^{(l)}}(\mathbb{R})$ for all $n \geqslant k, l$.*

(1) *If $k = 0$, then $f(x)$ equals a(ny) real number $f(0)$ plus a convergent power series with nonnegative Maclaurin coefficients and vanishing at $x = 0$.*
(2) *If $k > 0$, then f is linear or constant.*

5.2. **Preservers of positivity and of inertia, in several variables.** Schoenberg's theorem has a natural multivariable generalization, for any number of variables $m \geqslant 1$. Note by the Schur product theorem that if $A_1, \dots, A_m$ are any matrices in $\mathbb{P}_n(\mathbb{C})$ for some $n \geqslant 1$, then their entrywise product $A_1 \circ \cdots \circ A_m \in \mathbb{P}_n(\mathbb{C})$. Reformulated, this says that the function $f(\mathbf{x}) = x_1 \cdots x_m$ entrywise sends $\mathbb{P}_n(\mathbb{C})^m$ to $\mathbb{P}_n(\mathbb{C})$ for all $n \geqslant 1$. In general, we define

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

Restricting to real matrices, the easy implication of the following 1995 generalization of Schoenberg's theorem – by FitzGerald–Micchelli–Pinkus – is clear:

Theorem 5.4 ([45, Theorem 2.1]). *Let $I = \mathbb{R}$. An entrywise map $f : I^m \to \mathbb{R}$ sends $\mathbb{P}_n(I)^m$ to $\mathbb{P}_n(\mathbb{R})$ for all $n \geqslant 1$, if and only if f equals a convergent power series with nonnegative coefficients on I^m:*

$$f(\mathbf{x}) = \sum_{\alpha \in \mathbb{Z}_{\geqslant 0}^m} c_\alpha \mathbf{x}^\alpha \qquad \text{for all } \mathbf{x} \in I^m, \quad \text{with all } c_\alpha \geqslant 0. \tag{5.1}$$

This result was recently strengthened by Belton–Guillot–Khare–Putinar in two ways. First, the domain was reduced and also varied, to one- and two- sided domains. Second, the test set of matrices was severely reduced, to low rank Hankel ones.

Theorem 5.5 ([7]). *Let $0 < \rho \leqslant \infty$, and let $I = (0, \rho), [0, \rho)$, or $(-\rho, \rho)$. Fix a positive integer m and suppose $f : I^m \to \mathbb{R}$. The following are equivalent.*

(1) *The entrywise map $f[-]$ sends m-tuples in $\mathbb{P}_n(I)^m$ to $\mathbb{P}_n(\mathbb{R})$.*
(2) *Let $H_n(I)$ denote the Hankel matrices in $\mathbb{P}_n(I)$ – of rank at most 2 for $I = (0, \rho), [0, \rho)$, and rank at most 3 for $I = (-\rho, \rho)$. Then $f[-] : H_n(I)^m \to \mathbb{P}_n(\mathbb{R})$ for all $n \geqslant 1$.*
(3) *The map f is as in (5.1), on all of I^m.*

Skeleton of proof. That (3) $\implies$ (1) is via the Schur product theorem, and that (1) $\implies$ (2) is immediate. That (2) $\implies$ (3) was shown in [7] in the following locations:

- For $I = (0, \rho)$: see Theorem 9.6 in *loc. cit.*
- For $I = [0, \rho)$: see Theorem 9.6 and the proof of Proposition 9.8.
- For $I = (-\rho, \rho)$: see Theorem 9.11 and the subsequent remarks. $\square$

Having understood positivity preservers, we turn to the multivariate analogue of Theorem 5.3. The following is [8, Theorem B]:

Theorem 5.6 (Schoenberg-type theorem with negativity constraints). *Let $I := (-\rho, \rho)$, $(0, \rho)$, or $[0, \rho)$, where $0 < \rho \leqslant \infty$. Also fix integers $m \geqslant 1$ and $k_1, \dots, k_m, l \geqslant 0$. Rearrange the negative inertias k_p such that any zero values are at the start; thus there exists $0 \leqslant m_0 \leqslant m$ such that $k_1 = \cdots = k_{m_0} = 0 < k_{m_0+1}, \dots, k_m$.*

Now given any function $f : I^m \to \mathbb{R}$, the following are equivalent.

1. *The entrywise map $f[-]$ sends $\times_{p=1}^m \overline{\mathcal{S}_n^{(k_p)}}(I)$ to $\overline{\mathcal{S}_n^{(l)}}(\mathbb{R})$ for all $n \geqslant \max_p k_p$.*
2. *The entrywise map $f[-]$ sends $\times_{p=1}^m \mathcal{S}_n^{(k_p)}(I)$ to $\overline{\mathcal{S}_n^{(l)}}(\mathbb{R})$ for all $n \geqslant \max_p k_p$.*
3. *There exists a function $F : (-\rho, \rho)^{m_0} \to \mathbb{R}$ and a non-negative constant c_p for each $p = m_0 + 1, \dots, m$ such that*
   (a) *we have the representation*
$$f(\mathbf{x}) = F(x_1, \dots, x_{m_0}) + \sum_{p=m_0+1}^{m} c_p x_p \qquad \text{for all } \mathbf{x} \in I^m, \tag{5.2}$$
   (b) *the function $\mathbf{x}' := (x_1, \dots, x_{m_0}) \mapsto F(\mathbf{x}') - F(\mathbf{0}_{m_0})$ is absolutely monotone, that is, it is represented on I^{m_0} by a convergent power series with all Maclaurin coefficients non-negative, and*
   (c) *the inequality $\mathbf{1}_{F(\mathbf{0})<0} + \sum_{p : c_p > 0} k_p \leqslant l$ holds.*

In addition to judiciously chosen test matrices and analysis techniques, an interesting additional ingredient in the proof involves the use of Sidon sets (also termed B-sets) from number theory and additive combinatorics, whose use was pioneered by Singer [134] and Erdös–Turán [39]; see also Bose–Chowla [16].

Having discussed the real case, we end by mentioning the multivariable complex case too. In this case – following Herz's theorem 2.16, its multivariate counterpart was again shown in 1995 by FitzGerald–Micchelli–Pinkus:

Theorem 5.7 ([45, Theorem 3.1]). *Given an integer $m \geqslant 1$ and a function $f : \mathbb{C}^m \to \mathbb{C}$, the following are equivalent.*

1. *The entrywise map $f[-]$ sends $\mathbb{P}_n(\mathbb{C})^m$ to $\mathbb{P}_n(\mathbb{C})$ for all $n \geqslant 1$.*
2. *The function $f(\mathbf{z}) = \sum_{\alpha, \beta \in \mathbb{Z}_{\geqslant 0}^m} c_{\alpha,\beta} \mathbf{z}^\alpha \overline{\mathbf{z}}^\beta$ for all $\mathbf{z} \in \mathbb{C}^m$, with all $c_{\alpha,\beta} \geqslant 0$.*

This result was extended to classify the preservers of negative inertia. Here we present the complex analogue of Theorem 5.6; this is [8, Theorem C]:

Theorem 5.8. *Let the integers $m, k_1, \dots, k_m, l$ be as in Theorem 5.6. Given any function $f : \mathbb{C}^m \to \mathbb{C}$, the following are equivalent.*

1. *The entrywise transform $f[-]$ sends $\times_{p=1}^m \overline{\mathcal{S}_n^{(k_p)}}(\mathbb{C})$ to $\overline{\mathcal{S}_n^{(l)}}(\mathbb{C})$ for all $n \geqslant \max_p k_p$.*
2. *The entrywise transform $f[-]$ sends $\times_{p=1}^m \mathcal{S}_n^{(k_p)}(\mathbb{C})$ to $\overline{\mathcal{S}_n^{(l)}}(\mathbb{C})$ for all $n \geqslant \max_p k_p$.*
3. *There exists a function $F : \mathbb{C}^{m_0} \to \mathbb{C}$ and non-negative constants c_p and d_p for each $p = m_0 + 1, \dots, m$ such that*

(a) we have the representation

$$f(\mathbf{z}) = F(z_1, \dots, z_{m_0}) + \sum_{p=m_0+1}^{m} \left(c_p z_p + d_p \overline{z_p}\right) \qquad \text{for all } \mathbf{z} \in \mathbb{C}^m, \tag{5.3}$$

(b) the function $\mathbf{z}' := (z_1, \dots, z_{m_0}) \mapsto F(\mathbf{z}') - F(\mathbf{0}_{m_0})$ *is represented on* $\mathbb{C}^{m_0}$ *by a convergent power series in* $\mathbf{z}'$ *and* $\overline{\mathbf{z}'}$ *with non-negative coefficients, as in Theorem 5.7, and*

(c) $f(\mathbf{0}_m) = F(\mathbf{0}_{m_0})$ *is real, and we have* $\mathbf{1}_{F(\mathbf{0})<0} + \sum_{p:c_p>0} k_p + \sum_{p:d_p>0} k_p \leqslant l$*.*

Thus, the rich class of preservers in this result and Theorem 5.6 mix the absolutely monotone class of Schoenberg, with the rigid class of nonnegative homotheties. Notice that if all $k_p = l = 0$, we recover Schoenberg's (multivariate) real and complex theorems.

6. Recent developments on positivity preservers

We are nearing the end of this survey. This rather short section has three relatively disconnected parts.

6.1. **Preservers of positivity and of non-positivity.** In all of the results mentioned above, we have focused on classifying the functions f such that $f[A]$ is positive if A is so. (We omit the inertia and multivariate considerations of Section 5.) Here we consider the natural parallel question: what are the functions such that $f[A]$ is positive semidefinite *if and only if* A is so? A close variant is to replace "positive semidefinite" by "positive definite".

Here are the answers, from recent work by Guillot–Gupta–Vishwakarma–Yip. Remarkably, the answers are the same for all dimensions and for any fixed dimension greater than two:

Theorem 6.1 ([56, Theorem 1.8])**.** *Fix a dimension* $n \geqslant 3$ *and let* $\mathbb{F} = \mathbb{R}$ *or* $\mathbb{C}$*. The following are equivalent for an arbitrary function* $f : \mathbb{F} \to \mathbb{F}$*:*

(1) $A \in \mathbb{F}^{n\times n}$ *is a positive definite matrix if and only if* $f[A]$ *is.*
(2) $A \in \mathbb{F}^{n\times n}$ *is a positive semidefinite matrix if and only if* $f[A]$ *is.*
(3) f *is a positive multiple of a continuous field automorphism of* $\mathbb{F}$*. That is,* $f(x) \equiv cx$ *if* $\mathbb{F} = \mathbb{R}$*, and* $f(z) \equiv cz$ *or* $c\overline{z}$ *if* $\mathbb{F} = \mathbb{C}$*, for some* $c > 0$*.*

In fact the authors prove significantly stronger results in [56], in that they study the problem for matrices with entries in a wide class of sub-domains of $\mathbb{R}$ and $\mathbb{C}$.

6.2. **Preservers over finite fields.** We next study Schoenberg's theorem over a nonstandard setting: finite fields $\mathbb{F} = \mathbb{F}_q$. In analogy to the real case, here one defines a scalar to be *positive* if it is the square of a nonzero element – i.e., a nonzero quadratic residue. Notice that these elements still form half of the units $\mathbb{F}_q^\times$ when the prime power q is odd.

Defining positive matrices is more challenging. It turns out that over finite fields, even the basic characterizations of positive semidefinite matrices – in Theorem 1.1 – are not all available. Instead, we use a different characterization of positive *definite* matrices, one that is unavailable for semidefinite matrices:

A real symmetric matrix is positive definite if and only if its leading principal minors are all positive.

It turns out that this notion can be adapted usefully to finite fields. This was studied in detail by Cooper–Hanna–Whitlatch in [31], and they showed that when q is even or $q \equiv 3$ mod 4, such matrices $A_{n\times n}$ admit a Cholesky decomposition: $A = LL^T$, where L is lower triangular with entries in $\mathbb{F}_q$ and positive diagonal entries. Thus, we work with:

Definition 6.2. A symmetric $n \times n$ matrix over a finite field $\mathbb{F}_q$ is *positive definite* if its leading principal $1 \times 1, \dots, n \times n$ minors are squares of nonzero elements in $\mathbb{F}_q$.

We now present a natural class of entrywise preservers of positive *definiteness* over $\mathbb{F}_q$, in the spirit of Pólya and Szegő's century-old result. Namely, if $\operatorname{char}(\mathbb{F}_q) =: p$ and $x \mapsto x^p$ is the Frobenius automorphism, then

$$\det A^{\circ p} = \det(a_{ij}^p) = \sum_{\sigma \in S_n} \det(\sigma) \prod_{i=1}^n a_{i\sigma(i)}^p = \left(\sum_{\sigma \in S_n} \det(\sigma) \prod_{i=1}^n a_{i\sigma(i)} \right)^p = (\det A)^p.$$

From this it follows that the same functions as in Theorem 6.1(3) are preservers: positive multiples of field automorphisms $x \mapsto cx^{p^k}$, with $c \in \mathbb{F}_q$ positive and $k \geqslant 0$ in $\mathbb{Z}$. Remarkably, Guillot–Gupta–Vishwakarma–Yip showed that – akin to Theorem 6.1 – for every fixed $n \geqslant 3$, there are no other preservers:

Theorem 6.3 ([55])**.** *Let $q = p^\ell$ for a prime $p \geqslant 2$ and an integer $\ell \geqslant 1$, and $f : \mathbb{F}_q \to \mathbb{F}_q$. The following are equivalent.*

(1) The map $f[-]$ preserves $\mathbb{P}_n(\mathbb{F}_q)$ for some $n \geqslant 3$.
(2) The map $f[-]$ preserves $\mathbb{P}_n(\mathbb{F}_q)$ for all $n \geqslant 3$.
(3) $f(x) = cx^{p^k}$ for some $c \in \mathbb{F}_q$ positive and $0 \leqslant k \leqslant \ell - 1$.

If moreover p is odd, these are also equivalent to:

(4) $f(0) = 0$ and f is an automorphism of the Paley graph over $\mathbb{F}_q$, i.e.,

$$(f(a) - f(b))^{(q-1)/2} = (a - b)^{(q-1)/2} \qquad \forall a, b \in \mathbb{F}_q.$$

Thus, unlike the "classical" Schoenberg–Rudin theorem, both the "if-and-only-if" version in Theorem 6.1, as well as Theorem 6.3 over finite fields, admit solutions in each fixed dimension (and hence in the dimension-free setting). Moreover, the proof over finite fields is completely different than over the reals or complex numbers, and involves an interesting mix of tools: the quadratic character $x^{(q-1)/2}$, the celebrated Weil character bounds, results of Carlitz on Paley graphs, and the Erdös–Ko–Rado theorem, to name a few.

Remark 6.4. We add that the papers [56] and [55] prove many more results on "if-and-only-if" positivity preservers as well as preservers over finite fields; what is provided above and here are merely a sample. We refer the reader to these works for more details.

6.3. **Schoenberg's theorem for integer matrices.** We end this section by returning full circle to Schoenberg's theorem in characteristic zero – but this time over matrices with *integer* entries. In this case, one works with functions $f : \mathbb{Z} \to \mathbb{R}$ – which have a discrete domain, so one cannot even take limits of function-values!

Remarkably, Schoenberg's theorem still holds here, with somewhat stronger hypotheses:

Definition 6.5. A real matrix is said to be *partially defined* if a subset of its entries are unspecified. Such a matrix is said to be *positive (semi)definite* if there exists a choice of each unspecified entry such that the resulting *matrix completion* is positive (semi)definite.

We now come to the main new tool that is required to state the result.

Definition 6.6. Given a subset $I \subseteq \mathbb{R}$, a function $f : I \to \mathbb{R}$ is a *partially defined entrywise positivity preserver* if for every partially defined psd matrix A with specified entries in I, the partially defined matrix $f[A]$ is also psd.

In other words, for every partially defined psd matrix A with specified entries in I, both A and $f[A]$ can be completed (via some choices of real scalars) to psd matrices.

With these notions at hand, we state "Schoenberg's theorem over $\mathbb{Z}$":

Theorem 6.7 (Damase–Pascoe [34], forthcoming). *A function $f : \mathbb{Z} \to \mathbb{R}$ is a partially defined entrywise positivity preserver, if and only if $f(n) = \sum_{k\geqslant 0} c_k n^k$ for all integers n, with all $c_k \in [0,\infty)$.*

In other words, as "usual" f is the restriction to $\mathbb{Z}$ of an entire function that is absolutely monotonic on $(0,\infty)$. Note that one implication is precisely the Pólya–Szegő 1925 observation, via the Schur product theorem.

Remark 6.8. Theorem 6.7 is valid even if one replaces $\mathbb{Z}$ by any $X \subseteq \mathbb{R}$ with $\sup X = \infty$.

While Theorem 6.7 is indeed a pleasing addition, it is natural to ask if one can get rid of the "partially defined" test matrices:

Question 6.9. Suppose a (continuous) entrywise map $f : \mathbb{Z} \to \mathbb{R}$ preserves positivity on $\bigcup_{n\geqslant 1} \mathbb{P}_n(\mathbb{Z})$. Is f the restriction to $\mathbb{Z}$ of an entire function with nonnegative Maclaurin coefficients?

Appendix A. Sphere packings and kissing numbers in Euclidean space

This Appendix contains a parallel mini-survey of two famous problems in discrete geometry: understanding sphere packings and kissing numbers for spheres in Euclidean space. This also connects to the main body of this article via Schoenberg's theorem 3.14 (which classifies the dimension-free entrywise positivity preservers with a rank constraint). See Section A.7.

Much of this material can be found in the well known monograph [30] by Conway and Sloane. See also [151], as well as the classic 1964 text by Rogers [121].

The question of sphere packings is informally stated as follows: given a dimension $n \geqslant 1$, how does one stack the maximum amount of congruent balls – or *spheres* – $B_{\mathbb{R}^n}(a_k, r)$ for a fixed $r > 0$ and centers $a_1, a_2, \dots \in \mathbb{R}^n$, which can intersect only along their boundaries.

A.1. Problem statement and early history.

We now write the question more formally in order to set notation. We will work with unit spheres – i.e., $r = 1$ – and denote each sphere by its center, so that the condition of non-overlaps is equivalent to all centers being at least distance 2 apart.

Definition A.1. Fix a dimension $n \geqslant 1$.

1. A *(sphere) packing* in $\mathbb{R}^n$ (or in general, in any metric space) is a countably infinite subset (of "centers") $\mathcal{P} \subset (\mathbb{R}^n, \|\cdot\|_2)$ such that $\|x - y\|_2 \geqslant 2$ for all $x \neq y$ in $\mathcal{P}$.
2. A packing is a *lattice packing* if $\mathcal{P}$ is a lattice in $\mathbb{R}^n$.
3. The *density* of a packing $\mathcal{P}$ is (informally) the maximum proportion of space filled by the spheres, and is defined via the formula

$$\Delta_{\mathcal{P}} := \limsup_{M\to\infty} \frac{1}{(2M)^n} \mathrm{Vol}_{\mathbb{R}^n}\left([-M,M]^n \cap \bigcup_{x\in\mathcal{P}} B_{\mathbb{R}^n}(x,1) \right). \tag{A.1}$$

In other words, take a large cube $[-M,M]^n$ in $\mathbb{R}^n$ and intersect it with $\mathcal{P}$; now compute the percentage of this intersection inside the cube; and take $M \to \infty$. Finally, we have:

Definition A.2. The *sphere packing density* of $\mathbb{R}^n$ is $\Delta_{\mathbb{R}^n} := \sup\limits_{\text{packings } \mathcal{P}} \Delta_{\mathcal{P}}$.

On a related note, define the *lattice packing density* of $\mathbb{R}^n$ to be

$$\Delta^{(L)}_{\mathbb{R}^n} := \sup_{\text{lattice packings } \mathcal{P}} \Delta_{\mathcal{P}}. \tag{A.2}$$

The **question** of interest is to compute $\Delta_{\mathbb{R}^n}, \Delta^{(L)}_{\mathbb{R}^n} \in [0,1]$ for $n > 1$ (for $n = 1$, they are clearly 1). This question has a storied history. Perhaps the earliest version in modern times involves understanding close-packing of spheres, which came up in the 1580s when Sir Walter Raleigh asked the English astronomer and mathematician Thomas Harriot[4] about efficiently stacking cannonballs (on ships). Harriot published in the 1590s an analysis of stacking patterns – which also led towards atomic theory – and later corresponded with his German contemporary Johannes Kepler, who stated a conjectural form in 1611 in the work [82]: the maximum packing density in three dimensions is achieved by a pyramidal piling, akin to oranges in grocery stores.

It took almost four centuries to fully (and positively) settle Kepler's conjecture.[5] During this time, many researchers studied the problem in three *and two* dimensions. For $n = 2$: in 1773, Lagrange [94] studied extremal quadratic forms and deduced that the lattice density in $\mathbb{R}^2$ is achieved by the hexagonal/honeycomb packing: $\Delta^{(L)}_{\mathbb{R}^2} = \dfrac{\pi}{2\sqrt{3}}$. However, the bound for all packings was realized much later. In 1910, Thue showed [139] that this is indeed the unconstrained packing density: $\Delta_{\mathbb{R}^2} = \dfrac{\pi}{2\sqrt{3}}$; however, it is generally believed that his proof has a gap. A complete proof was provided in the 1940s by Fejes Tóth [40, 41].

The $n = 3$ case was harder, and an early result is by Gauss, who showed in his book review [51] of Seeber's book on quadratic forms [130][6] that Lagrange's 1773 methods could be modified to yield the 3-dimensional analogue: $\Delta^{(L)}_{\mathbb{R}^3} = \dfrac{\pi}{3\sqrt{2}}$. In other words, Kepler's cannonball packing achieves the lattice packing density in $\mathbb{R}^3$. (See Section A.3 for the connection of Lagrange's and Gauss' study of extremal quadratic forms to sphere packings.)

Note that there are not one, but at least two different ways to achieve the optimal cannonball packing density. Namely, on top of each "sheet" of spheres, one arranges the next sheet according to either the *FCC (face-centered cubic)* arrangement – also denoted *CCP* – or the *HCP (hexagonal close-packed)* arrangement. As one gets two such choices each time, both with the same density, there are uncountably many arrangements of "sheets" in 3-space which yield the Kepler packing density. This uncountable family is collectively termed as *Barlow packings*, in honor of the crystallographer William Barlow's 1883 work [3].

Subsequently, the Kepler conjecture featured in Hilbert's 18^{th} problem [75], at the turn of the 20th century.

A.2. **Applications.** A quick digression: the problem of sphere packings was not only of intrinsic mathematical interest (or for efficiently stacking cannonballs or oranges, in real life), it features in other fields too – starting with Harriot and Barlow, whose aforementioned works relate sphere packings to crystallography. Sphere packings are a reasonable starting point for modeling the structure of gases, liquids, crystals, and granular media, and can yield density estimates for "idealized materials".

[4]On a historical note, Thomas Harriot made remarkable contributions to mathematics and physics: he was the first to observe sunspots using a telescope, preceded Galileo in drawing a map of the moon using a telescope, and is said to have studied refraction and discovered Snell's law before Snell. In mathematics, he pioneered the modern way of computing with algebraic unknowns, and proved Girard's theorem on the area of a triangle on the unit sphere (predating Girard).

[5]On a light note: this means in particular that it isn't only Fermat's Last Theorem that had to wait for centuries for a resolution.

[6]Seeber was a German mathematician and physicist who is especially known for his mathematical work focusing on crystallography. Thus, between Lagrange, Gauss, Seeber, and the French physicist Auguste Bravais – who is known for working on the lattice theory of crystals – one can regard this as the time when lattices became formalized and mainstream in mathematics.

In the mathematical sciences, we now know that sphere packings are related to number theory (via modular forms) and to optimization. One can also connect the question to physics, including statistical mechanics and the Thomson problem. E.g. one imagines the centers of the spheres to be electrons that repel one another, and the densest packing is a "minimum energy configuration" of electrons.

However, arguably the most important application of (higher-dimensional) sphere packings is to communication theory: they are the continuous analogues, in a sense, of length n error-correcting codes. This emerges from work of Shannon, Hamming, and others. Roughly speaking, one is transmitting a set of signals – represented here by points $x \in \mathbb{R}^n$, with n usually large. (E.g., the coordinates may be amplitudes at different frequencies, typically in the hundreds or more.) The transmission channel may have a noise level $\varepsilon > 0$, so one expects the transmitted signal x to be received at the other end as some $y \in B_{\mathbb{R}^n}(x, \varepsilon)$. Thus, if one builds the signal set / "vocabulary" such that any two signals have distance at least 2ε, then this allows the mechanism to "error-correct". Whereas if this is not the case, certain received signals could not be "decoded" to recover uniquely the transmitted signal.

Moreover, we would like to have as large a vocabulary as possible – which translates precisely into maximally packing ε-spheres in a fixed space. This is an important real-world application: error-correcting codes are used by cell phones, the internet, and even space probes to send signals reliably.

Example A.3. We provide an early example of such a code, and it is discrete in nature. Given a set F and integers $n, k \geqslant 1$, we first set some notation.

(1) A *code* (or *F-code*) *of length* n is a subset $S \subset F^n$.
(2) Such a subset is a *k-error correcting code* if the *Hamming distance* between any $x \neq y \in S$ (i.e., the number of coordinates where $x_i \neq y_i$) is at least $2k+1$.
(3) A k-error correcting F-code of length n is *perfect* if every word in F^n is "detectable" by S up to error at most k (in particular, the balls in the next equality are disjoint):
$$F^n = \bigsqcup_{s \in S} B_{Hamming}(s, k).$$

For instance, here is a binary 1-error correcting code of length 7:
$$(1,1,0,1,0,0,0), \quad (0,1,1,0,1,0,0), \quad (0,0,1,1,0,1,0), \quad (0,0,0,1,1,0,1).$$
It is easy to check that the Hamming distance between any two unequal points is 4. Thus, the code can detect (and correct) errors in data transmission of a single bit. $\square$

The first perfect 1-error correcting codes were discovered in 1947 by Hamming (but published in 1950) [66], and in 1949 by Golay [54], in two landmark papers in the field. Since these works and another concurrent seminal work by Shannon [131], a problem of considerable interest in mathematics, information theory, and applications has been to try and construct binary codes with large minimum distance. (For sphere packings, in the sequel we will consider subsets/codes on the surface of a sphere, termed *spherical codes*.)

A.3. **Optimizing quadratic forms: from Lagrange to Hermite, to lattice packings.** Before we return to Kepler's conjecture and sphere packings in other dimensions, we explain how Lagrange's study of quadratic forms leads to lattice packings, via work of Hermite (and via a constant that is named after him). Some of this account is taken from [112, Section 1].

Given integers a, b, c, Lagrange was studying which integers can be represented by quadratic forms $ax^2 + bxy + cy^2$ for $x, y \in \mathbb{Z}$ – inspired by previous work of Euler and of Fermat (who had studied the "sum of two squares" case $a = c = 1, b = 0$). Thus, in his 1770s work [94], Lagrange generalized Euclid's algorithm to such binary quadratic forms. This was further generalized in 1850 by Hermite [70] along a common theme to this survey and to Theorem 1.1: for $x = (x_1, \dots, x_n) \in \mathbb{R}^n$, define the quadratic form $Q(x) := \sum_{i,j=1}^n q_{ij} x_i x_j$ for some real symmetric *positive definite* matrix/quadratic form Q. Hermite showed:

Theorem A.4 ([70]). *Given a positive definite real matrix $Q_{n\times n}$, there exist integers $x_1, \dots, x_n$ such that*

$$0 < Q((x_1, \dots, x_n)) \leqslant (4/3)^{(n-1)/2} \det(Q)^{1/n}.$$

If we define $\|Q\| := \inf_{x\in\mathbb{Z}^n\setminus\{0\}} Q(x)$, it turns out that this infimum is positive and is attained. Thus the ratio $\|Q\|/\det(Q)^{1/n}$ can be bounded above for all positive definite Q. This yields the constant named after Hermite:

Definition A.5. The *nth Hermite constant* $\gamma_n := \sup_{Q=Q^T \text{ positive definite}} \dfrac{\|Q\|}{\det(Q)^{1/n}}$. (This supremum is in fact attained.)

Moreover, the bivariate case of these facts had been shown by Lagrange:

Theorem A.6 ([94]). *The supremum γ_2 exists/is attained, and equals $\sqrt{4/3}$.*

It turns out that determining the Hermite constant is equivalent to a restricted version of computing the packing density: doing so for *lattices*. Recall that a lattice in $\mathbb{R}^n$ is the $\mathbb{Z}$-span of an $\mathbb{R}$-basis.

Definition A.7. Given a lattice $L \subset \mathbb{R}^n$, its *covolume* is the n-dimensional volume of $\mathbb{R}^n/L$; and its *least length* $\lambda_1(L)$ is the length of any shortest nonzero element of L.

The Hermite constant can be described in lattice-theoretic terms:

Proposition A.8. *γ_n equals the supremum of $\lambda_1(L)^2$ over all unit covolume lattices:*

$$\gamma_n := \sup\{\lambda_1(L)^2 \,:\, \mathrm{Vol}_{\mathbb{R}^n}(\mathbb{R}^n/L) = 1\}. \tag{A.3}$$

Proof. At the outset, note that if L is any lattice, then rescaling L by $c > 0$ rescales the least length by c and the covolume by c^n. Thus,

$$\frac{\lambda_1(cL)^2}{\mathrm{Vol}_{\mathbb{R}^n}(\mathbb{R}^n/cL)^{2/n}} = \frac{\lambda_1(L)^2}{\mathrm{Vol}_{\mathbb{R}^n}(\mathbb{R}^n/L)^{2/n}}, \tag{A.4}$$

and so we need to compare the supremum of this right-hand side to γ_n.

This is done using a standard correspondence between quadratic forms $Q > 0$ and lattices L. Given Q, Theorem 1.1 yields an invertible matrix $B_Q = [\mathbf{v}_1|\cdots|\mathbf{v}_n]$ whose columns have Gram matrix Q. This leads to the lattice $L(B_Q) := \oplus_{i=1}^n \mathbb{Z}\mathbf{v}_i$. Moreover, the choice of basis is unique up to an orthogonal change: $B_Q^T B_Q = Q = C_Q^T C_Q$ if and only if $B_Q C_Q^{-1} \in O(n)$.

Now if $B_Q = UC_Q$ for orthogonal U, then we compute:

$$\begin{aligned} \|Q\| = &\inf_{x\in\mathbb{Z}^n\setminus\{0\}} x^T Q x = \inf_{x\in\mathbb{Z}^n\setminus\{0\}} x^T B_Q^T B_Q x = \inf_{x\in\mathbb{Z}^n\setminus\{0\}} \left\| \sum_{i=1}^n x_i \mathbf{v}_i \right\|^2 = \lambda_1(L)^2, \\ \det(Q)^{1/n} = &\det(B_Q^T B_Q)^{1/n} = |\det(B_Q)|^{2/n} = \mathrm{Vol}_{\mathbb{R}^n}(\mathbb{R}^n/L)^{2/n}, \end{aligned} \tag{A.5}$$

and both calculations are unchanged under $B_Q \to UB_Q = C_Q$.

Conversely, given a lattice L, choose an ordered basis that generates it, say $(\mathbf{v}_1, \dots, \mathbf{v}_n)$. Now define $B_L := [\mathbf{v}_1|\cdots|\mathbf{v}_n]$, and $Q_L := B_L^T B_L$. Also note that if C_L is any other generating basis then $C_L = PB_L$ for some matrix $P \in GL_n(\mathbb{Z})$ (so $\det P = \pm 1$), and hence $Q \rightsquigarrow P^T Q P$. Now $\|P^T Q P\| = \lambda_1(L)^2 = \|Q\|$ by (A.5), since the lattice L is unchanged; and $\det(P^T Q P) = \det(Q)$. Moreover, one can reverse both calculations in (A.5), with B_Q replaced by B_L.

It follows that the set of ratios in the definition of γ_n (across all $Q > 0$) equals the set of ratios on the right side in (A.4) (across all L). Taking suprema, the result follows. ☐

Having gone from quadratic forms to lattices, we now mention the *lattice packing problem*: given a dimension $n \geqslant 1$, compute $\Delta_{\mathbb{R}^n}^{(L)}$. This turns out to be equivalent to the above:

Proposition A.9. *Given $n \geqslant 1$, computing the lattice packing density of $\mathbb{R}^n$ is equivalent to determining the Hermite constant γ_n.*

Proof. The result holds because of a simple equation governing the dependence between three quantities: (a) the lattice packing density $\Delta_{\mathbb{R}^n}^{(L)}$, (b) the Hermite constant γ_n, and (c) the n-dimensional volume ν_n of the unit ball in $\mathbb{R}^n$ – see [12, Equation (1)]:

$$\gamma_n = 4(\Delta_{\mathbb{R}^n}^{(L)}/\nu_n)^{2/n} \qquad \iff \qquad \Delta_{\mathbb{R}^n}^{(L)} = (\gamma_n/4)^{n/2}\nu_n = (\gamma_n/4)^{n/2}\frac{\pi^{n/2}}{\Gamma(n/2+1)}. \tag{A.6}$$

In fact this relationship holds for each lattice, and is not mysterious. Define the *packing radius* $r(L)$ to be the largest scalar such that placing an n-sphere of this radius at each point of L yields a packing. Thus, $r(L) = \lambda_1(L)/2$. Now a "fundamental domain" for $\mathbb{R}^n$ is the n-dimensional parallelopiped $\mathbb{R}^n/L$ (also called "parallelotope"), and the space covered in it by the packing is 2^n-many spherical "sectors/caps" – which make up exactly one sphere of radius $\lambda_1(L)/2$, so of n-volume $\nu_n \cdot (\lambda_1(L)/2)^n$. Thus, the density of the packing for L is:

$$\Delta_L := \frac{(\lambda_1(L)/2)^n}{\mathrm{Vol}_{\mathbb{R}^n}(\mathbb{R}^n/L)}\nu_n = \frac{(\lambda_1(L)^2/4)^{n/2}}{\mathrm{Vol}_{\mathbb{R}^n}(\mathbb{R}^n/L)} \cdot \frac{\pi^{n/2}}{\Gamma(n/2+1)}.$$

Clearly, rescaling the lattice by c changes both the numerator and denominator by a factor of c^n, so we may normalize to assume that L has covolume 1. This is precisely (A.6) before taking the supremum, because of the proof of Proposition A.8 (and it holds for every L of covolume 1), so now take the supremum. ☐

A.4. **Lattice packing densities in low dimensions; Hermite constants.** We now return to the story of the Kepler conjecture, following Lagrange, Gauss, Barlow, and Hilbert (see above). In 1953, László Fejes Tóth suggested a recipe for ascertaining Kepler's conjecture (see [42]), via checking a finite – but very large – number of cases to solve a finite-variable optimization problem. This would require advanced computing tools, which have since become available. In the 1990s Hales, together with his student Ferguson, applied linear programming techniques to try and minimize a function with 100+ variables, which they had shown would suffice to compute the sphere packing density of $\mathbb{R}^3$. Their research program took up 2+ years, 100,000 linear programming problems, and 3 gigabytes of computer programs. The findings appeared in the long articles [63, 65] in 2005–06. A decade later, Hales and many coauthors completed and published a formal proof of the computer-assisted calculations [64]. (There was also a proof of Kepler's conjecture by Wu-Yi Hsiang; but following criticism by Gábor Fejes Tóth and others, the proof is currently regarded as incomplete.)

This ends the story of sphere packings in three dimensions; note that in all cases discussed so far, the centers of the spheres in at least one configuration in each dimension (which is the only one in $\mathbb{R}$ and $\mathbb{R}^2$) lie on a lattice. These are precisely the lattices of Lie type A_n for $n = 1, 2, 3$ – i.e., the lattices generated in the hyperplane $(1, \dots, 1)^\perp \subset \mathbb{R}^{n+1}$ by the simple roots $\alpha_i = \mathbf{e}_i - \mathbf{e}_{i+1}$ for $1 \leqslant i \leqslant n$.

From above, the search for the Hermite constant is the same as that for the lattice packing density of Euclidean space. We have discussed the $n = 1, 2, 3$ cases above, and $\gamma_n, \Delta_{\mathbb{R}^n}^{(L)}$ are precisely known for just six other values of n to date. In the following table we summarize these results – via noting the values of $\gamma_n, \nu_n, \Delta^{(L)}_{\mathbb{R}^n}$, the associated lattices (and all but one are of simple Lie type), and their discoverers – or the discoverers of $\Delta^{(L)}_{\mathbb{R}^n}$, which is equivalent.

n	1	2	3	4	5	6	7	8	24
γ_n^n	1	$\frac{4}{3}$	2	4	8	$\frac{64}{3}$	64	2^8	4^{24}
ν_n	2	π	$\frac{4}{3}\pi$	$\frac{1}{2}\pi^2$	$\frac{8}{15}\pi^2$	$\frac{1}{6}\pi^3$	$\frac{16}{105}\pi^3$	$\frac{1}{24}\pi^4$	$\frac{1}{12!}\pi^{12}$
$\Delta^{(L)}_{\mathbb{R}^n}$	1	$\frac{\pi}{2\sqrt{3}}$	$\frac{\pi}{3\sqrt{2}}$	$\frac{\pi^2}{16}$	$\frac{\pi^2}{15\sqrt{2}}$	$\frac{\pi^3}{48\sqrt{3}}$	$\frac{\pi^3}{105}$	$\frac{\pi^4}{384}$	$\frac{\pi^{12}}{12!}$
Lattice (Lie) type	A_1	A_2	A_3	D_4	D_5	E_6	E_7	E_8	Leech
By:	–	Lagrange	Gauss	Korkine, Zolotareff		Hofreiter; Blichfeldt	Blichfeldt		Cohn, Kumar
Reference:	–	[94]	[51]	[88]	[90]	[78]; [13]	[13]		[26]
Year:	–	1773	1831	1872	1877	1933; 1935	1935		2009

TABLE 1. The known Hermite constants and lattice packing densities

A.5. **Kissing numbers: exact answers and bounds.** Having discussed exact answers in a few dimensions for the lattice packing density, before moving to all packings we take a detour into a related problem: determining the *kissing number* in $\mathbb{R}^n$. This is the largest number of non-overlapping unit spheres that a unit sphere can simultaneously touch, or "kiss" tangentially, and we will denote it by $k(n)$. It is not hard to show that $k(1) = 2$ and $k(2) = 6$. The $n = 3$ case was the subject of a famous 1694 debate between Newton (who thought it was 12) and Gregory (who thought it was 13). Thus, this question is also called the *thirteen spheres problem*, and the integer of interest is also called the *Newton number* or *contact number*. It was solved more than 250 years later – in 1952 – by Schütte and van der Waerden [129], who showed that $k(3)$ is indeed 12.

In four dimensions, the kissing number was computed by Musin in 2008 to be 24 [111]. The only other kissing numbers that are known are in dimensions $n = 8, 24$ – and they were both computed independently in 1979 by Levenshtein [96] and by Odlyzko–Sloane [113]: $k(8) = 240$ and $k(24) = 196560$. Remarkably, these two numbers also arise from spheres with centers in the same two lattices as for sphere packings! We will sketch their proofs below.

n	1	2	3	4	8	24
$k(n) = A(n, \pi/3)$	2	6	12	24	240	196560
Lattice (Coxeter) type	A_1	A_2	H_3	D_4	E_8	Leech
By:	–	–	Schütte, van der Waerden	Musin	Levenshtein; Odlyzko–Sloane	
Reference:	–	–	[129]	[111]	[96]; [113]	[96]; [113]
Year:	–	–	1952	2003	1979	1979

TABLE 2. The known kissing numbers

In addition to these exact results, and decades before Musin's 2008 article, upper and lower bounds were sought – for general n, not specific values. (This will also be the theme when we consider the (lattice) packing density of $\mathbb{R}^n$ for general n.) E.g., Coxeter [33] proposed some upper bounds in 1963. Wyner provided in 1965 an asymptotic lower bound of $2^{0.2075n(1+o(1))}$ [149]. But the object of our focus here and below is a ~50-year old upper bound due to Kabatiansky and Levenshtein [81].

Definition A.10. Given an angle $\psi \in [0, \pi]$ and a dimension $n \geqslant 2$, a finite subset $X \subset S^{n-1}$ is a *spherical ψ-code* if the spherical distance between any two vectors in X is at least ψ:

$$\sphericalangle x, y \geqslant \psi \qquad \Longleftrightarrow \qquad \langle x, y \rangle \leqslant \cos \psi.$$

Let $A(n, \psi)$ denote the size of any maximum-cardinality spherical ψ-code in S^{n-1}.

Remark A.11. The name "code" is akin to that in "error-correcting code" – see Example A.3 – in that it too stands for a set of points in a metric space, with all nonzero distances uniformly bounded below.

Additionally, spherical codes have their origins in a well-known question in mathematical biology: the *Tammes problem*, formulated in 1930 by Tammes [138]. The problem asks: given integers $n, N \geqslant 2$, pack N points on S^{n-1} such that the minimum distance between distinct points gets maximized.

As far as sphere packings and kissing numbers go, note that if two non-overlapping unit spheres kiss a common unit sphere, the closest that their centers can get is precisely when the three centers form an equilateral triangle. Thus, every packing or kissing problem involves $\psi \geqslant \pi/3$, i.e., $\cos \psi \leqslant 1/2$. In particular, the kissing number is

$$k(n) = A(n, \pi/3), \qquad \forall n \geqslant 2. \tag{A.7}$$

Now the celebrated 1978 upper bound of Kabatiansky–Levenshtein is:

Theorem A.12 ([81, Corollary 1])**.** *For all $\psi \in (0, 63°)$, we have*

$$n^{-1} \log_2 A(n, \psi) \leqslant \frac{-1}{2} \log_2(1 - \cos \psi) - 0.099 + o(1).$$

In particular, $k(n) = A(n, \pi/3) \leqslant 2^{n(0.401+o(1))}$.

A.6. **The E_8 and Leech lattices.** Notice in the aforementioned results – on both the lattice packing and kissing number problems – that beyond the first three or four dimensions, two dimensions stood out: $n = 8, 24$. This is because in both of these dimensions, both of the above problems get solved when the centers lie in the same, remarkable lattice of rank 8 or 24. Thus, for completeness we write down characterizations for both lattices.

The E_8 lattice is widely studied in the context of Lie theory, mathematical and particle physics, and Hamming codes (among other areas). It is a rank 8 lattice with any of the following properties:

(1) It is at once integral, even, and unimodular. Namely, $x \cdot y$ is an integer for all $x, y \in E_8$; $x \cdot x$ is moreover even for all x; and E_8 has a covolume 1.
(2) An alternate description is the set of points $x = (x_1, \dots, x_8) \in \mathbb{R}^8$ such that $\sum_i x_i$ is an even integer, and the x_i are either all integers or all half-integers.

E_8 is also the root lattice for the largest exceptional complex simple Lie algebra; we provide its Dynkin diagram.

It is known that the norm-square $Q : E_8 \to \mathbb{R}$, $Q(x) := \|x\|^2$ is a positive definite quadratic form. (Thus, up to isometry, E_8 is also the unique even, unimodular, positive definite lattice of rank 8.) This connects E_8 to both themes in this article: positivity and lattice packings. Indeed, the existence of such a form was proved in the very first article cited in this article, in 1868 by Smith [135]. It was then explicitly constructed first by Korkine–Zolotareff [89]

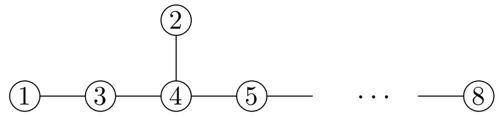

FIGURE A.1. The nodes (or simple roots) of the E_8 Dynkin diagram

in 1873, in the same series of works in which they computed the Hermite constant in dimensions 4 and 5 (see above). In 1938, Mordell [107] showed the uniqueness of a lattice with the above properties (he wrote a related work on lattice packings and γ_n, mentioned below).

Remark A.13. Moreover, one checks that the shortest norm-square of a nonzero vector in E_8 is $\lambda_1(E_8)^2 = 2$, and there are precisely 240 such vectors "nearest" to the origin. Thus, if one places spheres of packing radius $r(E_8) = \lambda_1(E_8)/2$ at these lattice points, they kiss the congruent sphere centered at the origin. Thus, $k(8) \geqslant 240$.

We next mention the *Leech lattice* Λ_{24}, which also features outside sphere packings in geometry (including higher-dimensional versions of the Tammes problem in mathematical biology), modular forms, group theory (via Conway groups), coding theory, and moonshine theory and vertex operator algebras. This lattice was introduced in 1967 by Leech [95], following lifting from $(\mathbb{Z}/2\mathbb{Z})^{24}$ to $\mathbb{Z}^{24}$ the extended Golay code [54] – this has minimal Hamming distance 8 (see Example A.3). The Leech lattice is the unique (up to isometry) unimodular even lattice in $\mathbb{R}^{24}$, and it has least length 2. Once again, one checks that there are 196560 points in Λ_{24} of least length (i.e., closest to the origin). Thus, $k(24) \geqslant 196560$.

A.7. **Kissing numbers and spherical codes, via Delsarte – and Schoenberg.** Here we elaborate on the *linear programming method* that led to the Kabatiansky–Levenshtein asymptotic upper bound for the kissing number $k(n)$. This method was pioneered by Delsarte in the 1970s – originally in [36] for cardinalities of binary codes; then in [37] for more general "association schemes"; and finally in 1977 with Goethals and Seidel [38], where Gegenbauer polynomials enter into the picture. This is the promised connection to Schoenberg's work (see Theorem 3.14) and to the above survey on positivity preservers.

A.7.1. *Refresher on Gegenbauer polynomials.* We recall here some basics on Gegenbauer polynomials $G_k^{(n)}(t)$, which formed the basis (literally!) of Schoenberg's classification of positive definite functions over spheres S^{n-1}. These can be defined in multiple ways: for instance, via their generating function

$$(1 - 2rt + r^2)^{(2-n)/2} = \sum_{k=0}^{\infty} r^k C_k^{(n)}(t); \qquad G_k^{(n)}(t) := \frac{C_k^{(n)}(t)}{C_k^{(n)}(1)}$$

if $n \geqslant 3$ – and for $n = 2$, we have:

$$\frac{1 - rt}{1 - 2rt + r^2} = \sum_{k=0}^{\infty} r^k G_k^{(2)}(t).$$

A second recipe to define these polynomials is via a three term recurrence, for any $n \geqslant 2$:

$$G_0^{(n)}(t) = 1, \quad G_1^{(n)}(t) = t, \qquad G_k^{(n)}(t) = \frac{(2k + n - 4)tG_{k-1}^{(n)}(t) - (k - 1)G_{k-2}^{(n)}(t)}{k + n - 3} \ \forall k \geqslant 2.$$

A third description is that the $G_k^{(n)}(t)$ are degree k polynomials that are normalized – $G_k^{(n)}(1) = 1$ – and form an orthogonal system with respect to integrating on $[-1, 1]$ against the measure $(1-t^2)^{(n-3)/2}\,dt$. (This is the projection to $[-1,1]$ of the surface measure $d\omega_{n-1}$ of the sphere.)

These orthogonal polynomials subsume various special cases. Setting $n = 2, 3, 4$, we recover, respectively: Chebyshev polynomials of the first kind, Legendre polynomials, and Chebyshev polynomials of the second kind. Moreover, while the above are perfectly adequate, self-contained definitions/characterizations of Gegenbauer polynomials, we now provide a fourth, beautiful description, wherein they arise naturally through spherical harmonics.

A.7.2. *Spherical harmonics and the Addition Theorem.* The following account is taken from the survey [115] and the early part of Müller's notes [110].

We work over the sphere $S^{n-1} \subset \mathbb{R}^n$. Thus, for $0 \neq x \in \mathbb{R}^n$, write $x = \|x\|\xi$, with $\xi \in S^{n-1}$. The area element on the sphere S^{n-1} will be denoted by $d\omega_{n-1} = d\omega_{n-1}(\xi)$, and the $(n-1)$-dimensional "surface area" of the sphere is

$$\omega_{n-1} := \int_{S^{n-1}} d\omega_{n-1} = \frac{2\pi^{n/2}}{\Gamma(n/2)}.$$

Thus $\omega_0 = |S^0| = 2$, $\omega_1 = 2\pi$, $\omega_2 = 4\pi$, and so on.

Next, we have the *Laplace operator* $\Delta_n := \sum_{j=1}^n \frac{\partial^2}{\partial x_j^2}$, and its null vectors among polynomials:

Definition A.14. Given integers $n \geqslant 2$ and $k \geqslant 0$, a *spherical harmonic* in n dimensions of degree/order k is a harmonic (i.e. $\Delta_n f \equiv 0$) polynomial $H_k(x) = H_k((x_1, \dots, x_n))$ that is homogeneous of degree k, and is now restricted to S^{n-1}. We will denote these by $H_k(\xi)$. Let $\mathcal{SH}_{n,k}$ denote the space of such polynomials.

The above integral defines an inner product on the span of all spherical harmonics (of all degrees) $\bigoplus_{k\geqslant 0} \mathcal{SH}_{n,k}$:

$$\langle f, g\rangle := \int_{S^{n-1}} f(\xi) g(\xi)\, d\omega_{n-1}. \tag{A.8}$$

By Green's theorem, one checks that for nonnegative integers $k \neq k'$ and corresponding degree spherical harmonics $H_k, H_{k'}$,

$$\begin{aligned} 0 = \int_{\|x\|\leqslant 1} (H_k \Delta_n H_{k'} - H_{k'} \Delta_n H_k)\, dx &= \int_{S^{n-1}} \left(H_k \frac{\partial H_{k'}}{\partial r} - H_{k'} \frac{\partial H_k}{\partial r} \right) d\omega_{n-1} \\ &= \int_{S^{n-1}} (k'-k) H_k(\xi) H_l(\xi)\, d\omega_{n-1}, \end{aligned}$$

since H_k has normal derivative on S^{n-1} (in the $r = \|x\|$ direction) equal to

$$\frac{\partial}{\partial r} H_k(r\xi)\Big|_{r=1} = k r^{k-1} H_k(\xi)\Big|_{r=1} = k\, H_k(\xi), \quad \forall k \geqslant 0.$$

Thus, spherical harmonics of differing degrees are orthogonal – and for a given degree $k \geqslant 0$, one uses e.g. Gram–Schmidt to obtain an orthonormal basis of $\mathcal{SH}_{n,k}$. One also has:

$$N(n,k) := \dim \mathcal{SH}_{n,k} = \binom{k+n-2}{k} + \binom{k+n-3}{k-1}, \qquad \forall n \geqslant 2,\ k \geqslant 0. \tag{A.9}$$

Example A.15 ($n = 2$)**.** We consider a well known special case, where $n = 2$ and the sphere is the unit circle. In this case $\dim \mathcal{SH}_{2,k} = 2$ for $k > 0$, and $\mathcal{SH}_{2,0} = \mathbb{R} \cdot 1$. For $k > 0$, two linearly independent degree k spherical harmonics are $\Re\,(x_2 + ix_1)^k$ and $\Im\,(x_2 + ix_1)^k$. Now introduce polar coordinates: $x_1 = r\cos\theta, x_2 = r\sin\theta$. Using this, we obtain an orthonormal set of spherical harmonics for each $k > 0$:

$$S_{k,1} := \frac{1}{\sqrt{\pi}\, r^k}\Re(x_2 + ix_1)^k = \frac{1}{\sqrt{\pi}}\cos k(\tfrac{\pi}{2} - \theta),$$
$$S_{k,2} := \frac{1}{\sqrt{\pi}\, r^k}\Im(x_2 + ix_1)^k = \frac{1}{\sqrt{\pi}}\sin k(\tfrac{\pi}{2} - \theta).$$

This example shows that when one thinks of S^1 not as a torus but as a one-dimensional sphere, the way to generalize Fourier series to higher dimensions is via spherical harmonics.

Now we bring in the orthogonal group.

Lemma A.16. *If $u : \mathbb{R}^n \to \mathbb{R}$ is smooth and $A \in O(n)$, then $\Delta_n(u \circ A) \equiv (\Delta_n u) \circ A$ on $\mathbb{R}^n$. Thus, if $H_k(\xi)$ is a spherical harmonic then so is $H_k(A\xi)$.*

Proof. The first assertion is an explicit computation, and immediately yields the second. □

Proposition A.17. *Fix integers $n \geqslant 2$ and $k \geqslant 0$ and an orthogonal matrix $A \in O(n)$. Suppose $\{S_{n,l} : 1 \leqslant l \leqslant N(n,k)\}$ is an orthonormal basis of $(\mathcal{SH}_{n,k}, \langle\cdot,\cdot\rangle)$.*

1. *Then so is $\{\xi \mapsto S_{n,l}(A\xi) : 1 \leqslant l \leqslant N(n,k)\}$.*
2. *Define the matrix $C^{(A)}$ via expanding $S_{n,l}(A\cdot -)$ in the orthonormal basis above:*

$$S_{n,l}(A\xi) = \sum_{r=1}^{N(n,k)} c_{lr}^{(A)} S_{n,r}(\xi), \qquad C^{(A)} := (c_{lr}^{(A)})_{l,r=1}^{N(n,k)}.$$

Then $C^{(A)}$ is also orthogonal, and $A \mapsto C^{(A)}$ is a group homomorphism $: O(n) \to O(N(n,k))$. In other words, $\mathcal{SH}_{n,k}$ is a finite-dimensional unitary representation of $O(n)$, under $A \cdot S_{n,l}(\xi) := S_{n,l}(A^T\xi)$.

3. *The kernel*

$$F : S^{n-1} \times S^{n-1} \to \mathbb{R}; \qquad F(\xi,\eta) := \sum_{l=1}^{N(n,k)} S_{n,l}(\xi)S_{n,l}(\eta) \tag{A.10}$$

is invariant under the diagonal action of $O(n)$ on its arguments, and hence depends only on their cosine $t = \langle\xi,\eta\rangle$.

Note that the group map $A \mapsto C^{(A)}$ need not be injective. For instance, let n be even and $A = -\mathrm{Id}_n$. Then $S_{n,l}(A\xi) = S_{n,l}(\xi)$ for all $k \geqslant 0$, $1 \leqslant l \leqslant N(n,k)$, and $\xi \in S^{n-1}$.

Proof.

(1) Fix $A \in O(n)$ and define $T_{n,l}(\xi) := S_{n,l}(A\xi)$. This is also in $\mathcal{SH}_{n,k}$ by Lemma A.16, so we have structure constants:

$$S_{n,l}(A\xi) = T_{n,l}(\xi) = \sum_{r=1}^{N(n,k)} c_{lr}^{(A)} S_{n,r}(\xi).$$

Now compute $\langle T_{n,l}, T_{n,m}\rangle$ in two ways: as an inner product and as an integral. First, since the $S_{n,l}$ are orthonormal, $\langle T_{n,l}, T_{n,m}\rangle = \sum_{r=1}^{N(n,k)} c_{lr}^{(A)} c_{mr}^{(A)}$. Second, since S^{n-1} and its surface measure $d\omega_{n-1}$ are invariant under the orthogonal group,

$$\langle T_{n,l}, T_{n,m}\rangle = \int_{S^{n-1}} S_{n,l}(A\xi)S_{n,m}(A\xi)\, d\omega_{n-1} = \int_{S^{n-1}} S_{n,l}(\xi)S_{n,m}(\xi)\, d\omega_{n-1} = \delta_{l,m}.$$

Equating the two expressions proves the assertion – and also shows that the matrix $C^{(A)} = (c_{lr}^{(A)})$ is orthogonal: $C^{(A)}(C^{(A)})^T = \mathrm{Id}_{N(n,k)}$.

(2) It suffices to show $A \mapsto C^{(A)}$ is multiplicative. This is just a formal exercise – given $A, B \in O(n)$, we have:

$$\begin{aligned} S_{n,l}(AB\xi) = \sum_{r=1}^{N(n,k)} c_{lr}^{(A)} S_{n,r}(B\xi) &= \sum_{r=1}^{N(n,k)} c_{lr}^{(A)} \sum_{q=1}^{N(n,k)} c_{rq}^{(B)} S_{n,q}(\xi) \\ &= \sum_{q=1}^{N(n,k)} \left(\sum_{r=1}^{N(n,k)} c_{lr}^{(A)} c_{rq}^{(B)} \right) S_{n,q}(\xi). \end{aligned}$$

On the other hand, $S_{n,l}(AB\xi) = \sum_{q=1}^{N(n,k)} c_{lq}^{(AB)} S_{n,q}(\xi)$. Hence $C^{(AB)} = C^{(A)} C^{(B)}$. The translation into $\mathcal{SH}_{n,k}$ being a unitary representation is now another formal exercise.

(3) To show (A.10), first note that F is invariant under the diagonal action of $O(n)$:

$$F(A\xi, A\eta) = \sum_{l=1}^{N(n,k)} \sum_{m,q=1}^{N(n,k)} c_{lm}^{(A)} c_{lq}^{(A)} S_{n,m}(\xi) S_{n,q}(\eta) = \sum_{m,q=1}^{N(n,k)} S_{n,m}(\xi) S_{n,q}(\eta) \sum_{l=1}^{N(n,k)} c_{lm}^{(A)} c_{lq}^{(A)}.$$

But the inner sum is the (m, q)th entry of $(C^{(A)})^T C^{(A)} = \mathrm{Id}_{N(n,k)}$, so we get

$$F(A\xi, A\eta) = F(\xi, \eta) \qquad \forall \xi, \eta \in S^{n-1},\ A \in O(n).$$

In particular, this invariance holds for $A \in SO(n)$. We are now done by Lemma 3.13, since $K = F$ is clearly continuous. □

With these preliminaries on spherical harmonics and Gegenbauer polynomials, one can identify exactly what is the function of the cosine in (A.10):

Theorem A.18 (Addition Theorem, [109, Equation (3.18)])**.** *Let n, k, and $\{S_{n,l} : 1 \leqslant l \leqslant N(n,k)\}$ be as in Proposition A.17. Then the function in* (A.10) *is the rescaled kth Gegenbauer polynomial:*

$$\sum_{l=1}^{N(n,k)} S_{n,l}(\xi) S_{n,l}(\eta) = \frac{N(n,k)}{\omega_{n-1}} G_k^{(n)}(\langle \xi, \eta \rangle), \tag{A.11}$$

where $\omega_{n-1} = \dfrac{2\pi^{n/2}}{\Gamma(n/2)}$ is the surface area of S^{n-1}.

Remark A.19. This result likely first appeared in work of Müller [109]; but at the start of this paper he attributes the ideas in the entire paper to a lecture given by Herglotz to the Göttingen Mathematical Society on November 1, 1945. Müller again credits Herglotz for the Addition Theorem in his book [110].

Example A.20. Before proceeding further, we write down the Addition Theorem in the special case discussed above: $n = 2$. Write $\xi = e^{i\theta},\ \eta = e^{i\phi}$, so that

$$\langle \xi, \eta \rangle = \cos\theta \cos\phi + \sin\theta \sin\phi = \cos(\theta - \phi),$$

which is the addition formula for $\cos(\cdot)$. Now

$$S_{2,1}(\xi) S_{2,1}(\eta) + S_{2,2}(\xi) S_{2,2}(\eta) = \frac{1}{\pi} \cos(n(\theta - \phi)).$$

Thus, as a function of $t = \langle \xi, \eta \rangle$, the Addition Theorem specializes to yield:

$$t = \cos(\theta - \phi) \mapsto \cos(n(\theta - \phi)),$$

which is precisely the Chebyshev polynomial (in t) of the first kind. $\square$

A.7.3. *The easier half of Schoenberg's theorem on Gegenbauer polynomials.* We now come to Schoenberg's theorem 3.14. For spherical code bounds like the one by Kabatiansky–Levenshtein, and to compute the kissing numbers in dimensions 8 and 24, we only need the "easier" implication, and this follows quickly from the Addition Theorem:

Proof of Theorem 3.14, easier half. The claim to be proved is: *any $\mathbb{R}_+$-linear combination of Gegenbauer polynomials $G_k^{(n)} \circ \cos$ is positive definite on distance matrices of S^{n-1}.*

To show this, we can remove the $\cos(\cdot)$ and replace distance matrices by their entrywise cosines, aka Gram matrices drawn from S^{n-1}. Now it suffices to work with a single Gegenbauer polynomial $G_k^{(n)}$. But by the Addition Theorem A.18, given any vectors $\xi_1, \dots, \xi_N \in S^{n-1}$ and scalars $x_1, \dots, x_N \in \mathbb{R}$, we have:

$$\begin{aligned} x^T \cdot (G_k^{(n)}(\langle \xi_i, \xi_j \rangle))_{i,j=1}^N \cdot x = & \sum_{i,j=1}^N x_i x_j G_k^{(n)}(\langle \xi_i, \xi_j \rangle) \\ = & \sum_{i,j=1}^N \sum_{l=1}^{N(n,k)} \frac{\omega_{n-1}}{N(n,k)} x_i x_j S_{n,l}(\xi_i) S_{n,l}(\xi_j) \\ = & \frac{\omega_{n-1}}{N(n,k)} \sum_{l=1}^{N(n,k)} \sum_{i,j=1}^N x_i x_j S_{n,l}(\xi_i) S_{n,l}(\xi_j) \\ = & \frac{\omega_{n-1}}{N(n,k)} \sum_{l=1}^{N(n,k)} \left(\sum_{i=1}^N x_i S_{n,l}(\xi_i) \right)^2 \geqslant 0, \end{aligned} \tag{A.12}$$

and so $G_k^{(n)}[-]$ indeed sends Gram matrices from S^{n-1} to positive semidefinite matrices. $\square$

Remark A.21. In fact, we do not even need the full power of the "easier half" of Schoenberg's theorem. Below, we will only need that the sum of the matrix entries $\sum_{i,j=1}^N G_k^{(n)}(\langle \xi_i, \xi_j \rangle)$ is nonnegative. But this follows by setting all $x_i = 1$ in (A.12).

We conclude with some remarks on Schoenberg's 1942 paper [127] and a footnote in it. First, to show the above easier half of Theorem 3.14, Schoenberg did not use the Addition Theorem as it was unavailable at the time; instead, he used an inductive and intrinsic Addition Formula for Gegenbauer polynomials $\{G_k^{(n)} : k \geqslant 0\}$ in terms of $\{G_k^{(n-1)} : k \geqslant 0\}$.

Second, there is a path from Schoenberg's results in [127] to not just regularization of covariance matrices, but also to a celebrated conjecture in complex analysis. Note that $G_k^{(n)} \circ \cos$ is positive definite on S^{n-1} for all k, hence on S^{n-2}. Hence by the "harder half" of Schoenberg's theorem 3.14, $G_k^{(n)}$ must be a nonnegative real combination of lower order Gegenbauer polynomials $\{G_k^{(n-1)} : k \geqslant 0\}$. Schoenberg noted this in [127, Footnote 2], and then remarked that this should hold when one replaces the integer parameter $n \geqslant 2$ by $\rho + 2$, for a nonnegative real parameter ρ. It later emerged that this had already been worked out in 1884 by Gegenbauer himself [52]. Almost a century later, in 1976 Askey and Gasper used these results to show in [1] the nonnegativity of a sequence of generalized hypergeometric functions ${}_3F_2$. These ideas were subsequently used in 1985 by de Branges, in his famous resolution of the Bieberbach conjecture [17].

A.7.4. *Application: kissing numbers are attained on E_8 and Λ_{24}, via the Delsarte–Goethals–Seidel bound.* To take stock of the last few pages: we have come from spherical harmonics, to the Addition Theorem, to the easier half of Schoenberg's theorem 3.14 on positive definite functions on spheres, aka positivity preservers on correlation matrices with rank bounded above. We now use this result to compute the kissing numbers $k(8)$ and $k(24)$, via Delsarte's method. In fact we show a stronger statement, which involves one more notion.

Definition A.22. Given an integer $n \geqslant 1$, let $k^{(L)}(n)$ denote the *lattice kissing number of* $\mathbb{R}^n$, i.e., the largest number of unit spheres that touch S^{n-1}, and whose centers lie on a lattice L with $\lambda_1(L) = 2$.

The following observations are clear. First, we have $2n \leqslant k^{(L)}(n) \leqslant k(n)$ for all n, since one can place sphere-centers at $\{\pm 2\mathbf{e}_j : 1 \leqslant j \leqslant n\}$ to kiss the unit sphere around the origin. (Thus, $L = \oplus_{j=1}^n \mathbb{Z}(2\mathbf{e}_j)$.) Second, if $k(n)$ is attained at a lattice, then $k(n) = k^{(L)}(n)$. This is indeed the case for $n = 1, 2, 3, 4$. Now one can show:

Theorem A.23 ([96, 113])**.** *We have $k(8) = k^{(L)}(8) = 240$ (attained on E_8) and $k(24) = k^{(L)}(24) = 196560$ (attained on Λ_{24}).*

The proof uses a famous upper bound on spherical codes, by Delsarte–Goethals–Seidel:

Theorem A.24 ([38])**.** *Fix an integer $n \geqslant 2$ and an angle $\psi \in [0, \pi]$. Let $f(t) = \sum_{k=0}^{d} c_k G_k^{(n)}(t)$, where $n \geqslant 2$ and all $c_k \in [0, \infty)$. Further assume that $c_0 > 0$ and $f(t) \leqslant 0$ for all $t \in [-1, \cos\psi]$. Then we have the upper bound*

$$A(n, \psi) \leqslant \frac{f(1)}{c_0}. \tag{A.13}$$

As promised, this will use the entrywise calculus and Schoenberg's positivity preserver f!

Proof. Choose $N := A(n, \psi)$-many points $\xi_1, \dots, \xi_N$ on the sphere with pairwise angles at least ψ. Apply the entrywise map $f[-]$ to their Gram matrix. Then,

$$\sum_{i,j=1}^{N} f(\langle \xi_i, \xi_j \rangle) = \sum_{k=0}^{d} c_k \sum_{i,j=1}^{N} G_k^{(n)}(\langle \xi_i, \xi_j \rangle) \geqslant c_0 \sum_{i,j=1}^{N} G_0^{(n)}(\langle \xi_i, \xi_j \rangle) = c_0 N^2,$$

using Remark A.21 and that $G_0 \equiv 1$.

Also note that by hypothesis, f applied to any off-diagonal entry is non-positive, since distinct points have inner product at most $\cos(\psi)$. Therefore,

$$\sum_{i,j=1}^{N} f(\langle \xi_i, \xi_j \rangle) = N f(1) + \sum_{i \neq j} f(\langle \xi_i, \xi_j \rangle) \leqslant N f(1).$$

Combining the two bounds gives $A(n, \psi) = N \leqslant f(1)/c_0$. □

Finally, we employ Theorem A.24 to show:

Proof of Theorem A.23, taken from [113]. We first prove the $n = 8$ case. We in fact showed (but did not state) in Section A.6 that $k(8) \geqslant k^{(L)}(8) \geqslant 240$. Now consider the carefully chosen degree 6 polynomial

$$\begin{aligned} f(t) &= \frac{320}{3}(t+1)(t+\tfrac{1}{2})^2 t^2 (t-\tfrac{1}{2}) \\ &= G_0^{(8)} + \frac{16}{7} G_1^{(8)} + \frac{200}{63} G_2^{(8)} + \frac{832}{231} G_3^{(8)} + \frac{1216}{429} G_4^{(8)} + \frac{5120}{3003} G_5^{(8)} + \frac{2560}{4641} G_6^{(8)}, \end{aligned} \tag{A.14}$$

where we suppress the "(t)" from the Gegenbauer polynomials, and one solves for the coefficients of the polynomials $G_k^{(8)}$ by a triangular change of basis from the monomial basis.

Then $f \leqslant 0$ on $[-1, 1/2]$, so one can apply Theorem A.24 with $\psi = \pi/3$ to upper bound the kissing number:

$$k(8) = A(8, \pi/3) \leqslant f(1)/c_0 = 240.$$

Combined with the above lower bounds, we are done by sandwiching.

The proof is similar for $n = 24$: we had seen that $k(24) \geqslant k^{(L)}(24) \geqslant 196560$. Now consider

$$\begin{aligned} f(t) &= \frac{1490944}{15}(t+1)(t+\tfrac{1}{2})^2(t+\tfrac{1}{4})^2 t^2 (t-\tfrac{1}{4})^2(t-\tfrac{1}{2}) \\ &= G_0^{(24)} + \frac{48}{23}G_1^{(24)} + \frac{1144}{425}G_2^{(24)} + \frac{12992}{3825}G_3^{(24)} + \frac{73888}{22185}G_4^{(24)} \\ &\quad + \frac{2169856}{687735}G_5^{(24)} + \frac{59062016}{25365285}G_6^{(24)} + \frac{4472832}{2753575}G_7^{(24)} \\ &\quad + \frac{23855104}{28956015}G_8^{(24)} + \frac{7340032}{20376455}G_9^{(24)} + \frac{7340032}{80848515}G_{10}^{(24)}. \end{aligned} \tag{A.15}$$

Once again $f \leqslant 0$ on $[-1, 1/2]$, so we apply Theorem A.24 with $\psi = \pi/3$ to obtain $k(24) \leqslant 196560$. Combined with the above lower bounds, we are again done. $\square$

A.8. **From spherical codes to sphere packing upper bounds.** We now return from Delsarte, Schoenberg, and kissing numbers, back to packing densities for Euclidean spaces. Recall Theorem A.12 by Kabatiansky–Levenshtein, which involved using linear programming to obtain an upper bound on the Kissing number/spherical code $A(n, \pi/3)$. From this, the authors deduced an upper bound on the packing density itself – which remained the state-of-the-art for many years:

Theorem A.25 ([81])**.** *For all angles $\theta \in [\pi/3, \pi]$, we have the relation*

$$\Delta_{\mathbb{R}^n} \leqslant (1 - \cos\theta)^{n/2} 2^{-n/2} A(n+1, \theta) = \sin(\theta/2)^n A(n+1, \theta). \tag{A.16}$$

In particular, for $\theta = \pi/3$, we have using Theorem A.12:

$$\Delta_{\mathbb{R}^n} \leqslant 2^{-n(0.599+o(1))}. \tag{A.17}$$

Sketch of proof, taken from [29]. Let $\mathcal{P}$ be any packing of $\mathbb{R}^n$ with unit spheres, with density Δ. Consider a solid sphere $\overline{B}_{\mathbb{R}^{n+1}}(0, R) \subset \mathbb{R}^{n+1}$, whose radius R we choose later. Also fix a hyperplane P passing through the origin, and consider the n-dimensional closed disk $D := P \cap \overline{B}_{\mathbb{R}^{n+1}}(0, R)$. Given that D has n-volume equal to that of R^n unit spheres, one expects that on average, a translation of the above sphere packing of $\mathbb{R}^n \cong P$ should intersect D in ΔR^n many *centers* of unit spheres. Choose such a translation, and project these centers onto the upper hemisphere of the boundary $R \cdot S^n$. The spherical distance between any two of these points is bigger than their Euclidean distance, which in turn exceeds the distance between their projections in D.

Now we bring in θ. Ensuring that any two projected points in the hemisphere are at least θ angle apart, is equivalent to their arc having length $\geqslant R\theta$, which holds if the chord joining them has length $\geqslant 2R\sin(\theta/2)$. But we know that all chords have length at least 2, so we equate these bounds and set $R = 1/\sin(\theta/2) > 1/(\theta/2)$. Then

$$\Delta \cdot R^n \leqslant A(n+1, \theta),$$

and this is the desired bound. $\square$

In 2014, Cohn and Zhao improved this bound, using a similarly short argument:

Theorem A.26 ([29])**.** *For all angles $\theta \in [\pi/3, \pi]$, we have the relation*

$$\Delta_{\mathbb{R}^n} \leqslant \sin(\theta/2)^n A(n,\theta). \tag{A.18}$$

This is at least as good, because all $A(n,\theta)$-many points can be arranged along the equatorial sub-sphere S^{n-1} in $S^n \subset \mathbb{R}^{n+1}$.

We now list a few previously shown upper bounds. The first is by Rogers in 1958:

Theorem A.27 ([120])**.** *Let X be a regular $(n+1)$-point simplex in $\mathbb{R}^n$ of side length 2 and a vertex at the origin. Then*

$$\Delta_{\mathbb{R}^n} \leqslant \sigma_n, \qquad \text{where } \sigma_n := \frac{(n+1)\mathrm{Vol}_{\mathbb{R}^n}(B_{\mathbb{R}^n}(0,1)\cap X)}{\mathrm{Vol}_{\mathbb{R}^n}(X)}. \tag{A.19}$$

In other words, the packing density of $\mathbb{R}^n$ cannot exceed the packing density of a regular simplex of edgelength 2 with unit spheres at the vertices.

Our final two bounds here are older: in 1929, Blichfeldt showed in [12] that $\Delta_{\mathbb{R}^n} \leqslant \frac{n+2}{2} \cdot 2^{-n/2}$, and in 1944, Mordell [108] provided a bound in the language of the Hermite constant:

$$\gamma_n \leqslant \gamma_{n-1}^{(n-1)/(n-2)}. \tag{A.20}$$

Remark A.28. Repeated use of Mordell's bound (A.20) shows that

$$\gamma_n \leqslant \gamma_{n-1}^{\frac{n-1}{n-2}} \leqslant \gamma_{n-2}^{\frac{n-1}{n-3}} \leqslant \cdots \leqslant \gamma_2^{n-1}.$$

But Lagrange's theorem A.6 shows $\gamma_2 = \sqrt{4/3}$; now combining these two facts yields Hermite's original theorem A.4, reformulated as: $\gamma_n \leqslant (4/3)^{(n-1)/2} = \gamma_2^{n-1}$.

We end this part by mentioning that work on upper bounds continues – see the very recent work [124], where new upper bounds are shown both for spherical codes (for angles $\theta < 62.997°$), and then for sphere packings in dimensions $n \geqslant 2000$. (This work improves by a constant factor the Kabatiansky–Levenshtein upper bound [81], as did Cohn–Zhao [29].) We add that Cohn maintains a webpage with the latest numerical upper bounds in low dimensions, and references for these.

A.9. **Lower bounds on sphere packings.** In comparison, there are many results in the 20th and 21st centuries that address lower bounds for the packing density $\Delta_{\mathbb{R}^n}$. The first is a simple "folklore" estimate.

Lemma A.29. $\Delta_{\mathbb{R}^n} \geqslant 2^{-n}$.

Proof. In fact, we claim that this estimate is achieved by any "saturated" packing $\mathcal{P}$ – one in which no additional sphere can be inserted. To see why, note as above that all spheres in a packing have centers separated by a distance of 2. Now we claim that $\bigcup_{x\in\mathcal{P}} B_{\mathbb{R}^n}(x,2) = \mathbb{R}^n$ – for if not, there would be a point in the complement, which is 2 apart from all other centers. But then one can add another unit sphere here, which contradicts saturation. Now since the density of the "doubled spheres" is 1, the result follows. □

This bound has seen several improvements throughout the past century and this one – many of the following are lower bounds even for lattice sphere packing densities.

Theorem A.30. *The packing density $\Delta_{\mathbb{R}^n} \cdot 2^n$ is at least as large as:*

1. *(Minkowski, 1905 [105] and Hlawka, 1943 [77].)* $2\zeta(n) = 2\sum_{j=1}^{\infty} j^{-n} = 2 + O(2^{-n})$.
2. *(Rogers, 1947 [119].)* $\dfrac{2n\zeta(n)}{e(1-e^{-n})}$. *In particular, this exceeds $0.73n$ for $n \gg 0$. (See also Davenport and Rogers [35].)*

(3) *(Ball, 1992* [2]*.)* $2(n-1)\zeta(n)$.
(4) *(Krivelevich–Litsyn–Vardy, 2004* [92]*.)* $n/100$.
(5) *(Vance, 2011* [141]*.)* $\dfrac{6n\zeta(n)}{e(1-e^{-n/4})}$, *if* $4|n$.
(6) *(Venkatesh, 2012* [144]*.)* $65963n$ *for all sufficiently large* n*; and (the first super-linear growth:)* $\frac{1}{2}n\log\log n$ *for infinitely many* n.
(7) *(Campos–Jenssen–Michelen–Sahasrabudhe* [18]*.)* $\frac{1-o(1)}{2}n\log n$.
(8) *(Gargava–Viazovska* [50]*.)* $n\log\log n - O(e^{-cn(\log n+O(1))}$ *for infinitely many* n*, where* $c>0$ *is a universal constant.*
(9) *(Klartag* [87]*.)* cn^2*, where* $c>0$ *is a universal constant. (In a sense, this is an adaptation/follow-up of the work of Rogers.)*

Thus, the search for better – maybe even sharp – lower and upper bounds for sphere packings in *general* (and large) dimension n is by no means over, and should see more exciting developments in the years ahead.

A.10. **Conclusion: Cohn–Elkies and Viazovska.** Having discussed (asymptotic) upper and lower bounds for the packing density in all/large dimensions, as well as low-dimensional special cases ($n=1,2,3$), we now conclude this section by mentioning the recent success in determining $\Delta_{\mathbb{R}^n}$ for $n=8,24$ – once again using the lattices E_8 and Λ_{24}, respectively. In these specific dimensions, one uses linear programming bounds once again – not on spherical codes via Delsarte's methods, but directly on $\mathbb{R}^n$ via a 2003 result of Cohn and Elkies. To state it, we first recall that the *Fourier transform* of an L^1 map $f:\mathbb{R}^n\to\mathbb{R}$ is:

$$\widehat{f}(y) := \int_{\mathbb{R}^n} f(x) e^{-2\pi i\langle x,y\rangle}\ dx.$$

We also need the following notion.

Definition A.31. An L^1 map $f:\mathbb{R}^n\to\mathbb{R}$ is said to be *admissible* if there exists $\delta\in(0,\infty)$ such that $|f(x)|$ and $|\widehat{f}(x)|$ are bounded above by a constant times $(1+\|x\|)^{-n-\delta}$.

Now Cohn and Elkies show:

Theorem A.32 ([25])**.** *Suppose* $f:\mathbb{R}^n\to\mathbb{R}$ *is admissible, and* $r>0$ *is such that*

(1) $f(0), \widehat{f}(0) > 0$,
(2) $f(x)\leqslant 0$ *whenever* $\|x\|\geqslant r$*, and*
(3) $\widehat{f}(y)\geqslant 0$ *for all* $y\in\mathbb{R}^n$.

Then one can upper bound the packing density of $\mathbb{R}^n$*:*

$$\Delta_{\mathbb{R}^n}\cdot\frac{\widehat{f}(0)}{f(0)} \quad\leqslant\quad \mathrm{Vol}_{\mathbb{R}^n}(B_{\mathbb{R}^n}(0,r/2)) = (r/2)^n\nu_n = (r/2)^n\frac{\pi^{n/2}}{\Gamma(n/2+1)}. \tag{A.21}$$

If moreover $\lambda_1(L)=r$ *for some lattice* L*, then it attains the packing bound,*

$$\Delta_L = \Delta_{\mathbb{R}^n} = \mathrm{Vol}_{\mathbb{R}^n}(B_{\mathbb{R}^n}(0,r/2))\frac{f(0)}{\widehat{f}(0)} \tag{A.22}$$

if and only if $f\equiv 0$ *on* $L\setminus\{0\}$ *and* $\widehat{f}\equiv 0$ *on* $L^*\setminus\{0\}$.

Recall here that the *dual lattice* L^* is defined to be the lattice either generated by the dual basis to a given basis of L, or equivalently,

$$L^* = \{t\in\mathbb{R}^n\ :\ \langle x,t\rangle\in\mathbb{Z}\ \forall x\in L\}.$$

Remark A.33. Note that the volume on the right of (A.21) may be $\geqslant 1$, in which case the bound is trivial. However, for fixed r the n-volume decreases to 0^+ as $n \to \infty$, and so the bound is certainly relevant for large n.

The proof of Theorem A.32 and the subsequent analysis use the Poisson summation formula. Note that if f is admissible, then $f, \widehat{f}$ are both in L^1 and continuous. Now we have:

Theorem A.34 (Poisson summation)**.** *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is admissible, and $L \subset \mathbb{R}^n$ is a lattice. Then for every $v \in \mathbb{R}^n$,*

$$\sum_{x \in L} f(x+v) = \frac{1}{\mathrm{Vol}_{\mathbb{R}^n}(\mathbb{R}^n/L)} \sum_{t \in L^*} e^{-2\pi i \langle v,t \rangle} \widehat{f}(t), \tag{A.23}$$

with both sides converging absolutely.

Sketch of proof of Theorem A.32 for lattice packings, taken from [23]. Let L be any lattice in $\mathbb{R}^n$. Since any lattice packing of L by spheres of packing radius $r(L) = \lambda_1(L)/2$ has the same packing density under rescaling the lattice and the spheres, let us rescale L such that $r(L) = r/2$. Then the lattice packing density is

$$\Delta_L = \frac{\mathrm{Vol}_{\mathbb{R}^n}(B_{\mathbb{R}^n}(0, r/2))}{\mathrm{Vol}_{\mathbb{R}^n}(\mathbb{R}^n/L)},$$

so it suffices to show the **claim** that if $r(L) = r/2$ then L has covolume at least $\widehat{f}(0)/f(0)$. To see this, apply Poisson summation (A.23) with $v = 0$. As $\lambda_1(L) = r$, the hypotheses give us that (a) the left side of (A.23) is bounded above by $f(0)$, while (b) the right side is bounded below by the ratio of $\widehat{f}(0)$ and the covolume of L. This proves the claim.

For the final assertion, the preceding paragraph says that

$$f(0) \geqslant \sum_{x \in L} f(x) = \frac{1}{\mathrm{Vol}_{\mathbb{R}^n}(\mathbb{R}^n/L)} \sum_{t \in L^*} \widehat{f}(t) \geqslant \frac{\widehat{f}(0)}{\mathrm{Vol}_{\mathbb{R}^n}(\mathbb{R}^n/L)}. \tag{A.24}$$

Hence (by the previous working,) (A.22) holds if and only if the extremal terms in (A.24) are equal, and this happens if and only if both inequalities in (A.24) are equalities. This completes the proof. ☐

This proof is remarkable, in that one ends up throwing out all but one terms on both sides of the Poisson summation formula! So if this approach is to yield a lattice that attains the packing density $\Delta_{\mathbb{R}^n}$ (and hence $\Delta_{\mathbb{R}^n}^{(L)}$), then a remarkably constrained function f would need to exist with all of the above properties. In [23], Cohn calls such an f a *magic function.*

Let us now get even *more* restrictive. If the lattice achieving this result is to be one of the two special lattices – $L = E_8, \Lambda_{24}$ – then we also note that $L^* \cong L$! Moreover, $r = \lambda_1(L)$, which is $\sqrt{2}$ for E_8 and 2 for the Leech lattice Λ_{24}, as mentioned in Section A.6.

Thus, we would then need a magic function f which satisfies the hypotheses of Theorem A.32 and all other constraints above; and moreover, f and $\widehat{f}$ vanish on all nonzero points in E_8 or Λ_{24}. (Finding such an f seems truly "magical"!) Otherwise, one has to try a completely different approach to attaining the packing density.

Initial investigations did seem to be promising. For instance, Cohn–Kumar showed that in 24 dimensions, the packing density is very close to the lattice packing density:

$$\Delta_{\mathbb{R}^{24}}^{(L)} = \Delta_{\Lambda_{24}} \leqslant \Delta_{\mathbb{R}^{24}} \leqslant \Delta_{\Lambda_{24}} \cdot (1 + 1.65 \cdot 10^{-30}).$$

So, does such a magic function exist? The first answer came for $n = 8$ – in which case, it does exist! This was the main result of Viazovska [145], which she announced in an arXiv preprint on "Pi Day 2016".[7] The proof and the underlying magic function quickly got understood, and within a week, Viazovska along with Cohn–Kumar–Miller–Radchenko posted another paper [27] where they found the analogous magic function for the Leech lattice as well. These magic functions show:

Theorem A.35 ([145, 27]). *For $n = 8$, the packing density and lattice packing density agree, and equal $\pi^4/384$ at the lattice E_8. For $n = 24$, the same statement holds, except that the lattice is Λ_{24} and the density is $\pi^{12}/12!$.*

In these works, Viazovska (et al) used the theory of modular forms to come up with the relevant magic functions. This was followed by the 2022 work of Cohn–Kumar–Miller–Radchenko–Viazovska [28], where they also showed that the lattices E_8 and Λ_{24} in fact minimized energy for every potential function that is a completely monotonic function of $\|x\|^2$ – such a phenomenon was previously known to hold only for $n = 1$. In particular, this also generalizes the optimality of the sphere packing densities on these lattices. For additional background and details, we refer the reader to two articles by Cohn. The first is his beautiful account [23] of Viazovska's work and its recent predecessors. The second consists of his lecture notes [24], which address not only sphere packing and kissing numbers, but also spherical harmonics, energy minimization, and more.

Acknowledgments. I firstly thank Gadadhar Misra for asking me to write a survey on Schoenberg's theorem, and also for his patience during my writing the survey and then expanding it to include the Appendix. I thank Mihai Putinar for many enriching additions and references, E.K. Narayanan for useful conversations about spherical harmonics, and Dominique Guillot for going through an early draft in detail and offering numerous helpful suggestions that improved the exposition. I thank Sujit Sakharam Damase for doing so as well, and for valuable discussions about Sections 3.3, 6.3, and Appendix A. In particular, the Appendix would not have been written without his enthusiasm for sphere packings and Gegenbauer (and other orthogonal families of) polynomials, perhaps owing to his upcoming work with James Pascoe [34].

It is also worth pointing out that both positivity and its preservers, as well as sphere packings, are areas of mathematics with numerous connections to other fields and extensive research over more than a century, and to do justice to either in a few pages is not possible or even reasonable. In particular, any omissions and errors are mine.

References

[1] Richard Askey and George Gasper. Positive Jacobi polynomial sums, II. *Amer. J. Math.*, 98(3):709–737, 1976.

[2] Keith Martin Ball. A lower bound for the optimal density of lattice packings. *Int. Math. Res. Not.*, 1992(10):217–221, 1992.

[3] William Barlow. Probable nature of the internal symmetry of crystals. *Nature*, 29:186–188, 1883.

[4] Alexander Belton, Dominique Guillot, Apoorva Khare, and Mihai Putinar. Matrix positivity preservers in fixed dimension. I. *Adv. Math.*, 298:325–368, 2016.

[5] Alexander Belton, Dominique Guillot, Apoorva Khare, and Mihai Putinar. *A panorama of positivity. 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.

[6] Alexander Belton, Dominique Guillot, Apoorva Khare, and Mihai Putinar. *A panorama of positivity. 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

[7] Coincidentally, this date abbreviates to 3/14/16 – perhaps the best date approximating π this century, as it gives π rounded to four decimal places!

– AMS Contemporary Mathematics 743, pp. 109–150, American Mathematical Society, 2020. Parts 1 and 2 (unified) available at arXiv:math.CA/1812.05482.

[7] Alexander Belton, Dominique Guillot, Apoorva Khare, and Mihai Putinar. Moment-sequence transforms. *J. Eur. Math. Soc.*, 24(9):3109–3160, 2022.

[8] Alexander Belton, Dominique Guillot, Apoorva Khare, and Mihai Putinar. Negativity-preserving transforms of tuples of symmetric matrices. *Preprint*, arXiv:math.CA/2310.18041, 2023.

[9] Serge Bernstein. Sur les fonctions absolument monotones. *Acta Math.*, 52(1):1–66, 1929.

[10] Gautam Bharali and Olga Holtz. Functions preserving nonnegativity of matrices. *SIAM J. Matrix Anal. Appl.*, 30(1):84–101, 2008.

[11] Peter John Bickel and Elizaveta Levina. Covariance regularization by thresholding. *Ann. Statist.*, 36(6):2577–2604, 2008.

[12] Hans Frederick Blichfeldt. The minimum value of quadratic forms, and the closest packing of spheres. *Math. Ann.*, 101:605–608, 1929.

[13] Hans Frederick Blichfeldt. The minimum values of positive quadratic forms in six, seven and eight variables. *Math. Z.*, 39:1–15, 1935.

[14] Salomon Bochner. *Vorlesungen über Fouriersche Integrale*. Akad. Verlag., Leipzig, 1932.

[15] Salomon Bochner. Monotone funktionen, Stieltjessche integrale und harmonische analyse. *Math. Ann.*, 108(1):378–410, 1933.

[16] Raj Chandra Bose and Sarvadaman Chowla. Theorems in the additive theory of numbers. *Comment. Math. Helv.*, 37:141–147, 1962.

[17] Louis de Branges. A proof of the Bieberbach conjecture. *Acta Math.*, 154(1-2):137–152, 1985.

[18] Marcelo Campos, Matthew Jenssen, Marcus Michelen, and Julian Sahasrabudhe. A new lower bound for sphere packing. *Preprint*, arXiv:math.MG/2312.10026, 2023.

[19] Constantin Carathéodory. Über den Variabilitätsbereich der Koeffizienten von Potenzreihen, die gegebene Werte nicht annehmen. *Math. Ann.*, 64:95–115, 1907.

[20] Augustin-Louis Cauchy. Mémoire sur les fonctions alternées et sur les sommes alternées. *Exercices Anal. et Phys. Math.*, 2:151–159, 1841.

[21] 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.

[22] Jens Peter Reus Christensen and Paul Ressel. Positive definite kernels on the complex Hilbert sphere. *Math. Z.*, 180(2):193–201, 1982.

[23] Henry Cohn. A conceptual breakthrough in sphere packing. *Notices Amer. Math. Soc.*, 64(2):102–115, 2017.

[24] Henry Cohn. Packing, coding, and ground states. In: *Mathematics and Materials* (M.J. Bowick, D. Kinderlehrer, G. Menon, and C. Radin, Eds.), Vol. 23 of IAS/Park City Mathematics Series, pp. 45–102, American Mathematical Society, Providence, 2017.

[25] Henry Cohn and Noam D. Elkies. New upper bounds on sphere packings I. *Ann. of Math. (2)*, 157(2):689–714, 2003.

[26] Henry Cohn and Abhinav Kumar. Optimality and uniqueness of the Leech lattice among lattices. *Ann. of Math. (2)*, 170(3):1003–1050, 2009.

[27] Henry Cohn, Abhinav Kumar, Stephen D. Miller, Danylo Radchenko, and Maryna S. Viazovska. The sphere packing problem in dimension 24. *Ann. of Math. (2)*, 185(3):1017–1033, 2017.

[28] Henry Cohn, Abhinav Kumar, Stephen D. Miller, Danylo Radchenko, and Maryna S. Viazovska. Universal optimality of the E_8 and Leech lattices and interpolation formulas. *Ann. of Math. (2)*, 196(3):983–1082, 2022.

[29] Henry Cohn and Yufei Zhao. Sphere packing bounds via spherical codes. *Duke Math. J.*, 163(10):1965–2002, 2014.

[30] John Horton Conway and Neil James Alexander Sloane. *Sphere packings, lattices and groups* (3rd ed.). Vol. 290 in: Grundlehren der mathematischen Wissenschaften, lxxiv+706 pp., Springer, New York, 1999.

[31] Joshua Cooper, Erin Hanna, and Hays Whitlatch. Positive-definite matrices over finite fields. *Rocky Mountain J. Math.*, 54(2):423–438, 2024.

[32] 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-Verlag, Berlin, Heidelberg, 2007.

[33] Harold Scott MacDonald Coxeter. An upper bound for the number of equal nonoverlapping spheres that can touch another of the same size. *Proc. Sympos. Pure Math.*, 7:53–71, American Mathematical Society, Providence, 1963.

[34] Sujit Sakharam Damase and James Eldred Pascoe. Work in progress.

[35] Harold Davenport and Claude Ambrose Rogers. Hlawka's theorem in the geometry of numbers. *Duke Math. J.*, 14(2):367–375, 1947.

[36] Philippe Delsarte. Bounds for unrestricted codes, by linear programming. *Philips Res. Rep.*, 27:272–289, 1972.

[37] Philippe Delsarte. An algebraic approach to the association schemes of coding theory. *Philips Res. Rep. Suppl.*, 10:vi+97 pp., 1973.

[38] Philippe Delsarte, Jean-Marie Goethals, and Johan Jacob Seidel. Spherical codes and designs. *Geom. Dedicata*, 6:363–388, 1977.

[39] Paul Erdös and Pál Turán. On a problem of Sidon in additive number theory, and on some related problems. *J. London Math. Soc.*, s1-16(4):212–215, 1941.

[40] László Fejes Tóth. Über einen geometrischen Satz. *Math. Z.*, 46:83–85, 1940.

[41] László Fejes Tóth. Über die dichteste Kugellagerung. *Math. Z.*, 48:676–684, 1942.

[42] László Fejes Tóth. *Lagerungen in der Ebene auf der Kugel und im Raum.* Vol. 65 in: Grundlehren der mathematischen Wissenschaften, xii+240 pp., Springer-Verlag, Berlin, Heidelberg, 1972.

[43] Pal Fischer and Jan D. Stegeman. Functions operating on positive-definite symmetric matrices. *J. Math. Anal. Appl.*, 171(2):461–470, 1992.

[44] Carl Hanson FitzGerald and Roger Alan Horn. On fractional Hadamard powers of positive definite matrices. *J. Math. Anal. Appl.*, 61(3):633–642, 1977.

[45] Carl Hanson FitzGerald, Charles Anthony Micchelli, and Allan Pinkus. Functions that preserve families of positive semidefinite matrices. *Linear Algebra Appl.*, 221:83–102, 1995.

[46] Maurice René Fréchet. Les dimensions d'un ensemble abstrait. *Math. Ann.*, 68:145–168, 1910.

[47] 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.

[48] Ferdinand Georg Frobenius. Über die elliptischen Funktionen zweiter Art. *J. reine angew. Math.*, 93:53–68, 1882.

[49] Ferdinand Georg Frobenius. Über die Darstellung der endlichen Gruppen durch lineare Substitutionen. *Sitzungsber. Preuss. Akad. Wiss.*, Berlin, 994–1015, 1897.

[50] Nihar Gargava and Maryna Viazovska. Mean value for random ideal lattices. *Preprint*, arXiv:math.NT/2411.14973, 2024.

[51] Carl Friedrich Gauss. Untersuchungen über die Eigenschaften der positiven ternären quadratischen Formen von Ludwig August Seber. *Göttingische gelehrte Anzeigen*, 1831.

[52] Leopold Gegenbauer. Zur Theorie der Functionen $C_n^\nu(x)$. *Denks. Akad. Wissen. Wien, Math. Klasse*, 48:293–316, 1884.

[53] Israel Moiseevich Gelfand. Spherical functions in symmetric Riemann spaces. *Doklady Akad. Nauk SSSR (N.S.)*, 70:5–8, 1950.

[54] Marcel Jules Edouard Golay. Notes on digital coding. *Proc. Instt. Radio Eng.*, 37:657, 1949.

[55] Dominique Guillot, Himanshu Gupta, Prateek Kumar Vishwakarma, and Chi Hoi Yip. Positivity preservers over finite fields. *J. Algebra*, 684:479–523, 2025.

[56] Dominique Guillot, Himanshu Gupta, Prateek Kumar Vishwakarma, and Chi Hoi Yip. Entrywise transforms preserving matrix positivity and non-positivity. *Preprint*, arXiv:math.CA/2507.06076, 2025.

[57] 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.

[58] Dominique Guillot, Apoorva Khare, and Bala Rajaratnam. Preserving positivity for matrices with sparsity constraints. *Trans. Amer. Math. Soc.*, 368(12):8929–8953, 2016.

[59] Dominique Guillot, Apoorva Khare, and Bala Rajaratnam. Critical exponents of graphs. *J. Combin. Theory Ser. A*, 139:30–58, 2016.

[60] Dominique Guillot, Apoorva Khare, and Bala Rajaratnam. Preserving positivity for rank-constrained matrices. *Trans. Amer. Math. Soc.*, 369(9):6105–6145, 2017; erratum, *Trans. AMS*, 378(11):8263–8264, *2025.*

[61] Dominique Guillot and Bala Rajaratnam. Functions preserving positive definiteness for sparse matrices. *Trans. Amer. Math. Soc.*, 367:627–649, 2015.

[62] Alexander Guterman, Chi-Kwong Li, and Peter Šemrl. Some general techniques on linear preserver problems. *Linear Algebra Appl.*, 315(1-3):61–81, 2000.

[63] Thomas Callister Hales. A proof of the Kepler conjecture. *Ann. of Math. (2)*, 162(3):1065–1185, 2005.

[64] Thomas Callister Hales, et al. A formal proof of the Kepler conjecture. *Forum Math. Pi*, 5: art. e2, 29 pp., 2017.

[65] Thomas Callister Hales and Samuel P. Ferguson. A formulation of the Kepler conjecture. *Disc. Comput. Geom.*, 36:21–69, 2006.

[66] Richard Wesley Hamming. Error detecting and error correcting codes. *Bell Sys. Tech. J.*, 29(2):147–160, 1950.

[67] Frank Hansen. Functions of matrices with nonnegative entries. *Linear Algebra Appl.*, 166:29–43, 1992.

[68] Henry Helson, Jean-Pierre Kahane, Yitzhak Katznelson, and Walter Rudin. The functions which operate on Fourier transforms. *Acta Math.*, 102(1):135–157, 1959.

[69] Gustav Herglotz. Über Potenzreihen mit positiven, reellen Teil im Einheitskreis. *Ber. Verhandl. Königlich–Sächsischen Gesell. Wiss. Leipzig*, Math.-Phys. Klasse, Bd. 63:501–511, 1911.

[70] Charles Hermite. Extraits de lettres de M. Ch. Hermite à M. Jacobi sur différents objects de la théorie des nombres. *J. reine angew. Math.*, 40:261–315, 1850.

[71] Alfred Hero and Bala Rajaratnam. Large-scale correlation screening. *J. Amer. Statist. Assoc.*, 106(496):1540–1552, 2011.

[72] Alfred Hero and Bala Rajaratnam. Hub discovery in partial correlation graphs. *IEEE Trans. Inform. Theory*, 58(9):6064–6078, 2012.

[73] Carl Samuel Herz. Fonctions opérant sur les fonctions définies-positives. *Ann. Inst. Fourier (Grenoble)*, 13(1):161–180, 1963.

[74] Fumio Hiai. Monotonicity for entrywise functions of matrices. *Linear Algebra Appl.*, 431(8):1125–1146, 2009.

[75] David Hilbert. Mathematical problems. *Bull. Amer. Math. Soc.*, 8:437–479, 1902.

[76] David Hilbert. *Grundzüge einer allgemeinen Theorie der linearen Integralrechnungen.* Vol. 11, Teubner, Leipzig und Berlin, xxiv+282 pp., 1912.

[77] Edmund Hlawka. Zur Geometrie der Zahlen. *Math. Z.*, 49:285–312, 1943.

[78] Nikolaus Hofreiter. Über Extremformen. *Monat. Math. Phys.*, 40:129–152, 1933.

[79] Roger Alan Horn. The theory of infinitely divisible matrices and kernels. *Trans. Amer. Math. Soc.*, 136:269–286, 1969.

[80] Tanvi Jain. Hadamard powers of some positive matrices. *Linear Algebra Appl.*, 528:147–158, 2017.

[81] Grigory A. Kabatiansky and Vladimir Iosifovich Levenshtein. On bounds for packings on a sphere and in space. *Problemy Peredachi Informatsii*, 14(1):3–25, 1978.

[82] Johannes Kepler. *Strena seu de nive sexangula.* Gottfried. Tampach, Frankfurt, 1611.

[83] Apoorva Khare. *Matrix analysis and entrywise positivity preservers.* London Math. Soc. Lecture Note Ser. 471, Cambridge University Press, 2022. Also Vol. 82, TRIM Series, Hindustan Book Agency, 2022.

[84] Apoorva Khare. Smooth entrywise positivity preservers, a Horn–Loewner master theorem, and symmetric function identities. *Trans. Amer. Math. Soc.*, 375(3):2217–2236, 2022.

[85] Apoorva Khare and Siddhartha Sahi. From Cauchy's determinant formula to bosonic and fermionic immanant identities. *Eur. J. Combin.*, 110: art.# 103683, 16 pp., 2023; erratum, *Eur. J. Combin.* 112: art.# 103743, 2 pp., 2023.

[86] Apoorva Khare and Terence Tao. On the sign patterns of entrywise positivity preservers in fixed dimension. *Amer. J. Math.*, 143(6):1863–1929, 2021.

[87] Boaz Klartag. Lattice packing of spheres in high dimensions using a stochastically evolving ellipsoid. *Preprint*, arXiv:math.MG/2504.05042, 2025.

[88] Aleksandr Nikolayevich Korkin and Yegor Ivanovich Zolotarev. Sur les formes quadratiques positives quaternaires. *Math. Ann.*, 5:581–583, 1872.

[89] Aleksandr Nikolayevich Korkin and Yegor Ivanovich Zolotarev. Sur les formes quadratiques positives. *Math. Ann.*, 6:366–389, 1873.

[90] Aleksandr Nikolayevich Korkin and Yegor Ivanovich Zolotarev. Sur les formes quadratiques. *Math. Ann.*, 11:242–292, 1877.

[91] Mark Grigor'evich Krein. Hermitian-positive kernels on homogeneous spaces: I, II. *Ukrain. Mat. Žurnal* 1:64–98, 1949; and *Ukrain. Mat. Žurnal* 2:10–59, 1950.

[92] Michael Krivelevich, Simon Litsyn, and Alexander Vardy. A lower bound on the density of sphere packings via graph theory. *Int. Math. Res. Not.*, 2004(43):2271–2279, 2004.

[93] Casimir Kuratowski. Quelques problèmes concernant les espaces métriques non-séparables. *Fund. Math.*, 25(1):534–545, 1935.

[94] Joseph-Louis Lagrange. Recherches d'Arithmétique. *Nouveaux mém. de l'acad. sci. des belles–lettres de Berlin*, 695–795, 1773,1775.

[95] John Leech. Notes on sphere packings. *Canadian J. Math.*, 19:251–267, 1967.

[96] Vladimir Iosifovich Levenshtein. On bounds for packings in n-dimensional Euclidean space. *Dokl. Akad. Nauk SSSR*, 245(6):1299–1303, 1979.

[97] Ai Li and Steve Horvath. Network neighborhood analysis with the multi-node topological overlap measure. *Bioinformatics*, 23(2):222–231, 2007.

[98] Chi-Kwong Li and Stephen Pierce. Linear preserver problems. *Amer. Math. Monthly*, 108(7):591–605, 2001.

[99] Charles Loewner. Über monotone Matrixfunktionen. *Math. Z.*, 38(1):177–216, 1934.

[100] Marvin David Marcus. All linear operators leaving the unitary group invariant. *Duke Math. J.*, 26(1):155–163, 1959.

[101] Maximilian Mathias. Über positive Fourier-Integrale. *Math. Z.*, 16:103–125, 1923.

[102] Karl Menger. Untersuchungen über allgemeine Metrik. *Math. Ann.*: *Part I*: Vol. 100:75–163, 1928; *Part II*: Vol. 103:466–501, 1930.

[103] Karl Menger (Ed.). *Ergebnisse eines Mathematischen Kolloquiums.* Leipzig und Berlin, 1933.

[104] James Mercer. Functions of positive and negative type and their connection with the theory of integral equations. *Phil. Trans. Royal Soc. London, Ser. A*, 209:415–446, 1909.

[105] Hermann Minkowski. Diskontinuitätsbereich für arithmetische Äquivalenz. *J. reine angew. Math.*, 129:220–274, 1905.

[106] Lajos Molnár. *Selected preserver problems on algebraic structures of linear operators and on function spaces.* Vol. 1895 of Lecture Notes in Mathematics, Springer-Verlag, Berlin, Heidelberg, 2007.

[107] Louis Joel Mordell. The definite quadratic forms in eight variables with determinant unity. *J. Math. Pures Appl.*, Ser. 9, Jubilé de M. J. Hadamard, 17(1-4):41–46, 1938.

[108] Louis Joel Mordell. Observation on the minimum of a positive quadratic form in eight variables. *J. London Math. Soc.*, 19(73):3–6, 1944.

[109] Claus Müller. Über die ganzen Lösungen der Wellengleichung. *Math. Ann.*, 124:235–264, 1951.

[110] Claus Müller. *Spherical harmonics.* Vol. 17 of Lecture Notes in Mathematics, Springer-Verlag, Berlin, Heidelberg, 1966.

[111] Oleg R. Musin. The kissing number in four dimensions. *Ann. of Math. (2)*, 168(1):1–32, 2008.

[112] Phong Q. Nguyen. Hermite's constant and lattice algorithms. Chapter 2, pp. 19–69, in: *The LLL algorithm: survey and applications* (Phong Nguyen, Brigitte Vallée, Eds.), Information Security and Cryptography (ISC) series, xiv+496 pp., Springer Berlin, Heidelberg, 2010.

[113] Andrew Michael Odlyzko and Neil James Alexander Sloane. New bounds on the number of unit spheres that can touch a unit sphere in n dimensions. *J. Combin. Th. Ser. A*, 26(2):210–214, 1979.

[114] Vern I. Paulsen and Mrinal Raghupathi. *An Introduction to the Theory of Reproducing Kernel Hilbert Spaces.* Cambridge studies in advanced mathematics 152, Cambridge University Press, Cambridge, 2016.

[115] Florian Pfender and Günter M. Ziegler. Kissing numbers, sphere packings, and some unexpected proofs. *Notices Amer. Math. Soc.*, 51(8):873–883, 2004.

[116] Allan Pinkus. Strictly positive definite functions on a real inner product space. *Adv. Comput. Math.*, 20:263–271, 2004.

[117] Georg Pólya and Gabor Szegő. *Aufgaben und Lehrsätze aus der Analysis. Band II: Funktionentheorie, Nullstellen, Polynome Determinanten, Zahlentheorie.* Springer-Verlag, Berlin, 1925.

[118] Frigyes Riesz. Sur certains systèmes singuliers d'équations intégrales. *Ann. Sci. Ecole Norm. Sup.*, ser. 3, 28:33–62, 1911.

[119] Claude Ambrose Rogers. Existence theorems in the geometry of numbers. *Ann. of Math.*, 48(4):994–1002, 1947.

[120] Claude Ambrose Rogers. The packing of equal spheres. *Proc. London Math. Soc.*, s3-8(4):609–620, 1958.

[121] Claude Ambrose Rogers. *Packing and covering.* Vol. 54 in: Cambridge Tracts in Mathematics and Mathematical Physics, viii+111 pp., Cambridge University Press, Cambridge, 1964.

[122] Walter Rudin. Positive definite sequences and absolutely monotonic functions. *Duke Math. J*, 26(4):617–622, 1959.

[123] Bernard Russo and Henry Abel Dye, Jr. A note on unitary operators in C^*-algebras. *Duke Math. J.*, 33(2):413–416, 1966.

[124] Naser Talebizadeh Sardari and Masoud Zargar. New upper bounds for spherical codes and packings. *Math. Ann.*, 389:3653–3703, 2024.

[125] 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.*, 36(3):724–732, 1935.

[126] Isaac Jacob Schoenberg. Metric spaces and positive definite functions. *Trans. Amer. Math. Soc.*, 44(3):522–536, 1938.
[127] Isaac Jacob Schoenberg. Positive definite functions on spheres. *Duke Math. J.*, 9(1):96–108, 1942.
[128] Issai Schur. Bemerkungen zur Theorie der beschränkten Bilinearformen mit unendlich vielen Veränderlichen. *J. reine angew. Math.*, 140:1–28, 1911.
[129] Kurt Schütte and Bartel Leendert van der Waerden. Das Problem der dreizehn Kugeln. *Math. Ann.*, 125:325–334, 1952.
[130] Ludwig August Seeber. Untersuchungen über die Eigenschaften der positiven ternären quadratischen Formen. *Math. Abh.*, 1 Bd., Freiburg, 1831.
[131] Claude Shannon. A mathematical theory of communication. *Bell Sys. Tech. J.*, 27(3):379–423, 1948.
[132] Harold Shapiro. Operator theory and harmonic analysis. In: *Twentieth century harmonic analysis – a celebration* (J.S. Byrnes, Ed.), pp. 31–56, NATO Sci. Ser., vol. 33, 2001.
[133] Barry Simon. *Loewner's Theorem on Monotone Matrix Functions*. Vol. 354 in: Grundlehren der mathematischen Wissenschaften, xi+459 pp., Springer Nature Switzerland, 2019.
[134] James Singer. A theorem in finite projective geometry and some applications to number theory. *Trans. Amer. Math. Soc.*, 43(3):377–385, 1938.
[135] Henry John Stephen Smith. On the orders and genera of quadratic forms containing more than three indeterminates. *Proc. Roy. Soc. London*, 16:197–208, 1868.
[136] Ingo Steinwart. On the influence of the kernel on the consistency of support vector machines. *J. Mach. Learn. Res.*, 2(1):67–93, 2002.
[137] James Drewry Stewart. Positive definite functions and generalizations, an historical survey. *Rocky Mountain J. Math.*, 6(3):409–434, 1976.
[138] Pieter Merkus Lambertus Tammes. On the origin of number and arrangement of the places of exit on the surface of pollen-grains. *Rec. Trv. Bot. Neerl.*, 27(1):1–84, 1930.
[139] Axel Thue. Über die dichteste Zusammenstellung von kongruenten Kreisen in einer Ebene. *Christiania Videnskabs-Selskabets Skrifter*, I. Math.- Naturv. Klasse, 1:1–9, 1910.
[140] Otto Toeplitz. Über die fourier'sche entwickelung positiver funktionen. *Rend. Circ. Mat. Palermo*, 32:191–192, 1911.
[141] Stephanie Vance. Improved sphere packing lower bounds from Hurwitz lattices. *Adv. Math.*, 227(5):2144–2156, 2011.
[142] Vladimir Naumovich Vapnik. *The nature of statistical learning theory*. Information Science and Statistics, Springer, New York, 1999.
[143] Harkrishan Lal Vasudeva. Positive definite matrices and absolutely monotonic functions. *Indian J. Pure Appl. Math.*, 10(7):854–858, 1979.
[144] Akshay Venkatesh. A note on sphere packings in high dimension. *Int. Math. Res. Not.*, 2013(7):1628–1642, 2013.
[145] Maryna S. Viazovska. The sphere packing problem in dimension 8. *Ann. of Math. (2)*, 185(3):991–1015, 2017.
[146] Prateek Kumar Vishwakarma. Positivity preservers forbidden to operate on diagonal blocks. *Trans. Amer. Math. Soc.*, 376(8):5261–5279, 2023.
[147] Hsien-Chung Wang. Two-point homogeneous spaces. *Ann. of Math. (2)*, 55(1):177–191, 1952.
[148] Hans Sylvain Witsenhausen. Minimum dimension embedding of finite metric spaces. *J. Combin. Th. Ser. A*, 42(2):184–199, 1986.
[149] Aaron Daniel Wyner. Capabilities of bounded discrepancy decoding. *Bell Systems Tech. J.*, 44(6):1061–1122, 1965.
[150] Bin Zhang and Steve Horvath. A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.*, 4(1):Art. 17, 45 pp., 2005.
[151] Chuanming Zong. *Sphere packings*. In: Universitext, xiv+242 pp., Springer, New York, 1999.

(A. Khare) Department of Mathematics, Indian Institute of Science, Bangalore 560012, India and Analysis & Probability Research Group, Bangalore 560012, India
Email address: `khare@iisc.ac.in`