# Finite free convolutions of polynomials

Adam W. Marcus[1] · Daniel A. Spielman[2] · Nikhil Srivastava[3]

## Abstract

We study three convolutions of polynomials in the context of free probability theory. We prove that these convolutions can be written as the expected characteristic polynomials of sums and products of unitarily invariant random matrices. The symmetric additive and multiplicative convolutions were introduced by Walsh and Szegö in different contexts, and have been studied for a century. The asymmetric additive convolution, and the connection of all of them with random matrices, is new. By developing the analogy with free probability, we prove that these convolutions produce real rooted polynomials and provide strong bounds on the locations of the roots of these polynomials.

✉  Daniel A. Spielman
    daniel.spielman@yale.edu

    Adam W. Marcus
    adam.marcus@epfl.ch

    Nikhil Srivastava
    nikhil@math.berkeley.edu

[1]  École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

[2]  Yale University, New Haven, CT, USA

[3]  UC Berkeley, Berkeley, CA, USA

🖄 Springer

# 1 Introduction

We study three convolutions on polynomials that are inspired by free probability theory. Instead of capturing the limiting spectral distributions of ensembles of random matrices, we show that these capture the expected characteristic polynomials of random matrices in a fixed dimension. We develop the analogy with Free Probability by proving that Voiculescu's $R$ and $S$-transforms can be used to prove upper bounds on the extreme roots of these polynomials. Two of the convolutions have been classically studied. The third, and the connection of all of them with random matrices, is new. We begin by defining the three convolutions and stating the algebraic identities that establish this connection, as well as basic results regarding their real-rootedness properties.

## 1.1 Algebraic identities and real rootedness

### Symmetric additive convolution

**Definition 1.1** For complex univariate polynomials

$$p(x) = \sum_{i=0}^{d} x^{d-i}(-1)^i a_i \quad \text{and} \quad q(x) = \sum_{i=0}^{d} x^{d-i}(-1)^i b_i$$

of degree at most $d$, the *dth symmetric additive convolution* of $p$ and $q$ is defined as:

$$p(x) \boxplus_d q(x) \stackrel{\text{def}}{=} \sum_{k=0}^{d} x^{d-k}(-1)^k \sum_{i+j=k} \frac{(d-i)!(d-j)!}{d!(d-k)!} a_i b_j$$

$$= \frac{1}{d!} \sum_{k=0}^{d} D^k p(x) D^{d-k} q(0)$$

$$= \frac{1}{d!} \sum_{k=0}^{d} D^k q(x) D^{d-k} p(0), \tag{1}$$

where $D$ denotes differentiation with respect to $x$. Note that we have defined a *sequence* of convolutions parameterized by $d$, and in general $p \boxplus_c q \neq p \boxplus_d q$, even if both $p$ and $q$ have degree less than $c$ and $d$; we will discuss this point more in Sect. 1.4.

Observe that the definition above has the compact form:

$$(p \boxplus_d q)(x) = \hat{p}(D)\hat{q}(D)x^d, \tag{2}$$

where $\hat{p}$ and $\hat{q}$ are the unique polynomials satisfying $\hat{p}(D)x^d = p(x)$ and $\hat{q}(D)x^d = q(x)$. This reveals that $\boxplus_d$ is symmetric, bilinear in its arguments, and commutes with differentiation and translation; i.e.,

$$(Dp(x)) \boxplus_d q(x) = D(p(x) \boxplus_d q(x)) \quad \text{and} \quad p(x-t) \boxplus_d q(x) = (p \boxplus_d q)(x-t).$$

Note that the identity element for $\boxplus_d$ is the polynomial $x^d$.

For a square $d \times d$ complex matrix $M$, we define

$$\chi_x(M) \overset{\text{def}}{=} \det(xI - M),$$

its characteristic polynomial in the variable $x$. We show that for monic polynomials, the operation $\boxplus_d$ can be realized as an expected characteristic polynomial of a sum of random matrices.

**Theorem 1.2** *If $p(x) = \chi_x(A)$ and $q(x) = \chi_x(B)$ for $d \times d$ complex normal matrices $A, B$, then*

$$p(x) \boxplus_d q(x) = \underset{Q}{\mathbb{E}} \chi_x \left( A + QBQ^* \right), \tag{3}$$

*where the expectation is over a random unitary matrix $Q$ from the Haar measure on $U(d)$.*

In fact, by unitary invariance the right hand side depends only on $\chi_x(A)$ and $\chi_x(B)$ and not on the further details of $A$ and $B$, so we may take them to be any normal matrices with the same characteristic polynomials.

The convolution (1) was studied by Walsh [1], who proved results including the following theorem (see also [2] and [3, Theorem 5.3.1]).

**Theorem 1.3** *If $p$ and $q$ are real rooted polynomials of degree $d$, then so is $p \boxplus_d q$. Moreover,*

$$maxroot\,(p \boxplus_d q) \leq maxroot\,(p) + maxroot\,(q)\,.$$

In Theorem 1.12 we strengthen this bound on the maximum root. Our result is much tighter in the case that most of the roots of $p$ and $q$ are far from their maximum roots.

**Symmetric multiplicative convolution**

**Definition 1.4** For complex univariate polynomials

$$p(x) = \sum_{i=0}^{d} x^{d-i}(-1)^i a_i \quad \text{and} \quad q(x) = \sum_{i=0}^{d} x^{d-i}(-1)^i b_i$$

of degree at most $d$, the *$d$th symmetric multiplicative convolution* of $p$ and $q$ is defined as:

$$p(x) \boxtimes_d q(x) \overset{\text{def}}{=} \sum_{i=0}^{d} x^{d-i}(-1)^i \frac{a_i b_i}{\binom{d}{i}}. \tag{4}$$

It is clear that $\boxtimes_d$ is also bilinear, though it does not commute with differentiation or translation.

The following compact differential form of (4), analogous to (2), was discovered by B. Mirabelli [4] who has kindly allowed us to include it here: if $p(x) = P(xD)(x-1)^d$ and $q(x) = Q(xD)(x-1)^d$, then

$$(p \boxtimes_d q)(x) = P(xD)Q(xD)(x-1)^d. \tag{5}$$

Note that every polynomial of degree at most $d$ can be written as $P(xD)(x-1)^d$, though it is not as obvious as in the additive case. The appearance of the polynomial $(x-1)^d$ is explained by the fact that it is the identity element for $\boxtimes_d$, i.e., $p(x) \boxtimes_d (x-1)^d = p(x)$ for every $p$ of degree at most $d$.

We show that for monic polynomials the operation $\boxtimes_d$ can be realized as an expected characteristic polynomial of a *product* of random matrices.

**Theorem 1.5** *If $p(x) = \chi_x(A)$ and $q(x) = \chi_x(B)$ for $d \times d$ complex normal matrices $A$, $B$, then*

$$p(x) \boxtimes_d q(x) = \mathbb{E}_Q \chi_x \left( AQBQ^* \right), \tag{6}$$

*where the expectation is over a Haar unitary Q.*

The identity element $(x-1)^d$ thereby corresponds to taking $B = I$ in the above formula.

This convolution was studied by Szegö [5], who proved the following theorem.

**Theorem 1.6** *If $p$ and $q$ have only nonnegative real roots, then the same is true of $p \boxtimes_d q$. Moreover,*

$$\mathsf{maxroot}\,(p \boxtimes_d q) \leq \mathsf{maxroot}\,(p)\,\mathsf{maxroot}\,(q)\,.$$

We give a quantitatively stronger bound in Theorem 1.13.

### Asymmetric additive convolution

**Definition 1.7** For complex univariate polynomials

$$p(x) = \sum_{i=0}^{d} x^{d-i}(-1)^i a_i, \quad \text{and} \quad q(x) = \sum_{i=0}^{d} x^{d-i}(-1)^i b_i,$$

of degree at most $d$, we define the *dth asymmetric additive convolution* of $p$ and $q$ to be:

$$p(x) \boxplus_d q(x) = \sum_{k=0}^{d} x^{d-k}(-1)^k \sum_{i+j=k} \left( \frac{(d-i)!(d-j)!}{d!(d-k)!} \right)^2 a_i b_j. \tag{7}$$

This is equivalent to the expression

$$p(x) \boxplus_d q(x) = \left(\frac{1}{d!}\right)^2 \sum_{i=0}^{d} ((d-i)!)^2 b_i (DxD)^i p(x),$$

which may also be written for $p(x) = P(DxD)x^d$ and $q(x) = Q(DxD)x^d$ as

$$p(x) \boxplus_d q(x) = P(DxD)Q(DxD)x^d,$$

the latter being an observation of [4].

We are not aware of previous studies of this convolution. We show that if $p(x)$ and $q(x)$ are monic *with all roots real and nonnegative*, then $p \boxplus_d q$ can be realized as an expected characteristic polynomial.

**Theorem 1.8** *For two monic polynomials $p(x) = \chi_x(AA^*)$ and $q(x) = \chi_x(BB^*)$ with $A, B$ any $d \times d$ square complex matrices,*

$$p(x) \boxplus_d q(x) = \mathop{\mathbb{E}}_{R,Q} \chi_x \left((A + RBQ)(A + RBQ)^*\right). \tag{8}$$

*where the expectation is taken over independent Haar unitaries $R$ and $Q$.*

Theorem 1.8 is a consequence of Theorem 2.14, which we prove in Sect. 2.2.2.

We prove the following real-rootedness theorem in Sect. 3.

**Theorem 1.9** *If $p$ and $q$ have only nonnegative real roots, then the same is true of $p \boxplus_d q$.*

We obtain bounds on the roots of $p \boxplus_d q$ in Theorem 1.14.

**Remark 1.10** In an earlier version of this paper [6], we *defined* the operations $\boxplus_d, \boxtimes_d$, and $\boxplus_d$ in terms of random matrices using the formulas (3), (6), (7), and stated the formulas appearing in Definitions 1.1, 1.4, and 1.7 as theorems by showing that they were equivalent. We have chosen to reverse this presentation since $\boxplus_d, \boxtimes_d$ were in fact already defined by Walsh and Szegö, albeit in a different context, and their basic properties such as bilinearity are more immediate from the purely algebraic definitions.

## 1.2 Motivation and related results

Before describing our analytic results on the locations of roots of the convolutions in the next section, we explain the motivations for studying them in the context of several other areas of mathematics.

*Free probability* Free probability theory (see e.g. [7–9]) studies among other things the large $d$ limits of random matrices such as those in (1.1), (1.4), and (1.7). In particular, it allows one to calculate the limiting spectral distribution of a sum or product of two unitarily invariant random matrix ensembles in terms of the limiting spectral distributions of the individual ensembles. For both the sum and the product, free probability provides a transform of the moments of the spectra of the individual matrices

from which one can easily derive the transform of the moments of the limiting spectra of the resulting matrices—these are known as the *R*- and *S*-transforms, which may be viewed as generating functions for certain polynomials in the moments (known as free cumulants) which are linear in the convolutions. We show in Theorems 1.12, 1.13, and 1.14 (discussed in Sect. 1.3) that for our "finite free convolutions" the same transforms provide *upper bounds* on the roots of the corresponding expected polynomials in *finite dimensions*, when evaluated at specific points on the real line.

Following the definitions in this paper, [10,11] have shown that by taking appropriate limits our finite free convolutions yield the standard free convolutions in free probability theory. Thus, expected characteristic polynomials provide an alternative "discretization" of free convolutions from the typical one involving random matrices.

The paper [12] shows that the real zeros of $p \boxplus_d q$ and $p \boxtimes_d q$ may be interpreted as the $\beta \to \infty$ limit of certain generalizations of $\beta-$ensembles in random matrix theory.

*Combinatorics* The original motivation for this work is the method of interlacing families of polynomials [13–15], which reduces various combinatorial problems concerning eigenvalues to problems of bounding the roots of the expected characteristic polynomials of certain random matrices. In particular, the paper [15] studies bipartite random regular graphs, whose expected characteristic polynomials turn out to be of the type appearing in (1.7). The bound of Theorem 1.14 on the roots of these polynomials then implies the existence of bipartite Ramanujan graphs of every size and every degree (a result that was later turned into a polynomial time algorithm by Cohen [16]). Hall, Puder, and Sawin [17] used some of the techniques in this paper to prove the related result that every bipartite graph has a $k-$cover which is Ramanujan, for every $k \geq 2$, generalizing [13].

**Remark 1.11** (Renumbered Theorems) The papers [15–17] cite theorems in the original arxiv version of this work [6], which are numbered differently from the ones in this paper. In particular, Theorems 1.7 and 1.8 of [6] correspond to Theorems 1.12 and 1.14 of this paper.

*Representation theory* As shown in Sect. 2, the unitary group may be replaced by the orthogonal group or the group of signed permutations in Theorems 1.2, 1.5, and 1.8 without changing the expected characteristic polynomial and therefore any of their statements. The ability to compute the average of a matrix function over the group of unitary matrices by instead computing an average over some smaller group of matrices is a phenomenon we refer to as *quadrature* (see [15] for more details).

The proofs given Sect. 2 are different from those that appeared in the first version [6] of this paper. The original proofs treat each of the convolutions as (specifically) being an average over the unitary group, and then by explicit calculations show that one can (in some cases) replace the unitary group with the signed permutation matrices and get the same result. After posting that preprint, we were informed [17] that our quadrature results were actually a manifestation of well-studied concepts in representation theory (the Peter-Weyl theorem); the work [17] gave a more general sufficient condition for a subgroup of the unitary group to have this property .

One subgroup that has been used specifically in an application is the $n \times n$ matrices corresponding to the standard representation of $S_{n+1}$ (the symmetric group on $n + 1$

elements). The fact that this group is a valid quadrature group plays a pivotal role in the results on Ramanujan graphs [15] and [17] mentioned above.

*Geometry of polynomials* Theorem 1.3 implies that if $p(x)$ is real-rooted of degree $d$, then the linear transformation $p \boxplus_d (\cdot)$ preserves real-rootedness of all real polynomials of degree at most $d$. Leake and Ryder [18] observe that a partial converse is also true: every differential operator $T : \mathbb{R}_{\leq d}[x] \to \mathbb{R}_{\leq d}[x]$ preserving real-rootedness can be written as $T(q) = p \boxplus_d q$ for some $p$ of degree at most $d$. Thus, our bounds on the extreme roots of the additive convolution imply bounds on how much any such operator can perturb the roots of its real rooted inputs. They also generalize our analytic bounds on the roots of symmetric additive convolutions (Theorem 1.12) by showing that they are a special case of submodularity relation.

*Random matrix theory* The expected characteristic polynomials of symmetric Gaussian random matrices are Hermite polynomials (see, e.g. [19, Theorem 4.1]). If $R$ is an $d$-by-$d$ matrix of independent Gaussian random variables of variance 1, then

$$\mathbb{E} \chi_x \left( \frac{R + R^T}{\sqrt{2}} \right) = H_d(x),$$

where

$$H_d(x) = e^{-D^2/2} x^d$$

is the $d$th Hermite polynomial. By applying Theorem 1.2, but taking the expectation over orthogonal matrices, or by applying the formula (2), we may conclude that for positive $a$ and $b$ and $c = \sqrt{a^2 + b^2}$,

$$a^d H_d(x/a) \boxplus_d b^d H_d(x/b) = c^d H_d(x/c).$$

Similarly, the expected characteristic polynomial of $RR^T$ is the $d$th Laguerre polynomial [20, Sect. 9]

$$L_d(x) = (1 - D)^d x^d.$$

Thus, both Theorem 1.8 and the definition (7) can be used to show that for positive $a$ and $b$ and $c = a + b$,

$$a^d L_d(x/a) \boxplus\!\!\!\boxplus_d b^d L_d(x/b) = c^d L_d(x/c).$$

## 1.3 Transforms and root bounds

In free probability, each of the three convolutions comes equipped with a natural transform of probability measures. We define analogous transforms on polynomials and use them to bound the extreme roots of the convolutions of polynomials.

We will identify a vector $(\lambda_1, \ldots, \lambda_d)$ with the discrete distribution that takes each value $\lambda_i$ with probability $1/d$. The Cauchy/Hilbert/Stieltjes transform of such a distribution is the function

$$\mathcal{G}_\lambda (x) = \frac{1}{d} \sum_{i=1}^{d} \frac{1}{x - \lambda_i}.$$

Given a polynomial $p$ with roots $\lambda_1, \ldots, \lambda_d$, we similarly define

$$\mathcal{G}_p (x) := \mathcal{G}_\lambda (x).$$

We will prove theorems about the inverse of the Cauchy transform, which we define for real $w > 0$ by

$$\mathcal{K}_p (w) := \max \{ x : \mathcal{G}_p (x) = w \}.$$

For a real rooted polynomial $p$, and thus for real $\lambda_1, \ldots, \lambda_d$, this is the value of $x$ that is larger than all the $\lambda_i$ for which $\mathcal{G}_p (x) = w$. Since $\mathcal{G}_p (x)$ is decreasing above the largest root of $p$, the maximum is uniquely attained for each $w > 0$.

As $\mathcal{G}_p (x) = \frac{1}{d} \frac{p'(x)}{p(x)}$,

$$\mathcal{G}_p (x) = w \quad \Longleftrightarrow \quad p(x) - \frac{1}{wd} p'(x) = 0.$$

This tells us that

$$\mathcal{K}_p (w) = \mathsf{maxroot} \left( U_{1/wd}\, p \right),$$

where we define

$$U_\alpha p(x) := p(x) - \alpha D p(x).$$

Voiculescu's R-transform of the probability distribution that is uniform on $\lambda$ is given by $\mathcal{R}_\lambda (w) = \mathcal{K}_\lambda (w) - 1/w$ (though in free probability the inversion is typically done at the level of power series). We use the same notation to define a transform on polynomials

$$\mathcal{R}_p (w) := \mathcal{K}_p (w) - 1/w.$$

If $\lambda$ and $\mu$ are compactly supported probability distributions on the reals, then their free convolution $\lambda \boxplus \mu$ satisfies [7]:

$$\mathcal{R}_{\lambda \boxplus \mu} (w) = \mathcal{R}_\lambda (w) + \mathcal{R}_\mu (w).$$

For our finite additive convolution, we obtain an analogous *inequality* for $w > 0$.

**Theorem 1.12** *For $w > 0$ and real-rooted polynomials $p$ and $q$ of degree $d$,*

$$\mathcal{R}_{p \boxplus_d q} (w) \leq \mathcal{R}_p (w) + \mathcal{R}_q (w), \tag{9}$$

*with equality if and only if $p(x)$ or $q(x)$ has the form $(x - \lambda)^d$.*

We will often write (9) as:

$$\mathsf{maxroot}\,(U_\alpha(p \boxplus_d q)) + d\alpha \leq \mathsf{maxroot}\,(U_\alpha p) + \mathsf{maxroot}\,(U_\alpha q),$$

where $\alpha = 1/wd$.

To bound the roots of the finite multiplicative convolution, we employ a variant of Voiculescu's $S-$transform. We first define a variant of the moment transform, which we write as a power series in $1/z$ instead of in $z$:

$$\widetilde{\mathcal{M}}_p(z) := z\mathcal{G}_p(z) - 1 = \frac{1}{d}\sum_{i=1}^{d}\sum_{j\geq 1}\left(\frac{\lambda_i}{z}\right)^j.$$

For a polynomial $p$ having only nonnegative real roots and a $z > 0$,

$$z > \mathsf{maxroot}\,(p) \quad \Longleftrightarrow \quad \widetilde{\mathcal{M}}_p(z) < \infty.$$

We define the inverse of this transform, $\widetilde{\mathcal{M}}_p^{(-1)}(w)$, to be the largest $z$ so that $\widetilde{\mathcal{M}}_p(z) = w$, and

$$\widetilde{\mathcal{S}}_p(w) = \frac{w}{w+1}\widetilde{\mathcal{M}}_p^{(-1)}(w).$$

This is the reciprocal of the usual $S$-transform. We prove the following bound on this transformation in Sect. 4.2.

**Theorem 1.13** *For degree $d$ polynomials $p$ and $q$ having only nonnegative real roots and $w > 0$,*

$$\widetilde{\mathcal{S}}_{p\boxtimes_d q}(w) \leq \widetilde{\mathcal{S}}_p(w)\,\widetilde{\mathcal{S}}_q(w),$$

*with equality only when $p$ or $q$ has only one distinct root.*

One can ask whether an inequality similar to Theorem 1.13 could hold in more generality (i.e., for a larger collection of polynomials than just those with nonnegative roots). While this seems possible, in this paper we will restrict to the case that *both* polynomials have nonnegative roots (see Remark 4.10 in Sect. 4.2).

To define the relevant transform for the asymmetric additive convolution, we define $\mathbb{S}$ to be the linear map taking a polynomial $p(x)$ to $p(x^2)$. If $p$ has only positive real roots $\lambda_i$, then $\mathbb{S}p$ has roots $\pm\sqrt{\lambda_i}$. If $\lambda$ is a probability distribution supported on the nonnegative reals, then we use $\mathbb{S}\lambda$ to denote the corresponding symmetric probability distribution on $\pm\sqrt{\lambda_i}$. If $\lambda$ and $\mu$ are compactly supported probability distributions on the positive reals, then Benaych-Georges [21] showed that their appropriately defined rectangular convolution $\lambda \boxplus \mu$ satisfies:

$$\mathcal{R}_{\mathbb{S}\lambda\boxplus\mathbb{S}\mu}(w) = \mathcal{R}_{\mathbb{S}\lambda}(w) + \mathcal{R}_{\mathbb{S}\mu}(w).$$

In Sect. 4.3 we derive an analog of this result in the form of the following inequality.

**Theorem 1.14** *For degree d polynomials p and q having only nonnegative real roots and $w > 0$,*

$$\mathcal{R}_{\mathbb{S}(p \boxplus_d q)}(w) \leq \mathcal{R}_{\mathbb{S}p}(w) + \mathcal{R}_{\mathbb{S}q}(w).$$

**Remark 1.15** The formulas above are stated only for polyomials of degree exactly $d$, but they may be applied to polynomials of degree at most $d$ by first applying the degree-reduction formulas outlined in the next section.

### 1.4 Polynomials of different degrees

The operation $p \boxplus_d q$ is defined above for pairs of polynomials $p$ and $q$ of degree at most $d$, but if one or both of the polynomials has degree $c < d$, it may be written in terms of the lower degree operation $\boxplus_c$.

**Lemma 1.16** (Degree Reduction for $\boxplus_d$) *Suppose $\deg(p) \leq d$ and $\deg(q) \leq d - 1$. Then:*

$$p(x) \boxplus_d q(x) = (D/d) p(x) \boxplus_{d-1} q(x) \tag{10}$$

**Proof** By (2), the differential operator $p(x) \boxplus_d (\cdot) : \mathbb{R}_{d-1}(x) \to \mathbb{R}_{d-1}[x]$ is equal to $\hat{p}(D)$ where $\hat{p}(D)x^d = p(x)$. Differentiating, we have

$$\hat{p}(D)x^{d-1} = (D/d)\hat{p}(D)x^d = (Dp/d)(x),$$

so we must have $p(x) \boxplus_d q(x) = (Dp/d)(x) \boxplus_{d-1} q(x)$ whenever $\deg(q) \leq d - 1$. □

This relationship between convolutions of different grades turns out to be a key tool in the analytic proofs in Sect. 4, which induct on the degrees of the polynomials. We prove similar degree reduction formulas for $\boxtimes_d$ and $\boxminus_d$ in Lemmas 4.9 and 4.16, but have chosen to present them in context along with the inductive proofs in Sect. 4.

It turns out that the operation $\boxplus_d$ can be naturally defined for polynomials of degree strictly greater than $d$ via the random matrix identity (3). The key observation is that the expected characteristic polynomial of a random restriction of a $d \times d$ matrix is proportional to a derivative of its characteristic polynomial.

**Lemma 1.17** *If $a > d$, A is an $a \times a$ matrix and Q is a random $d \times a$ complex matrix with orthonormal rows (i.e., sampled from the Haar measure on the complex $d \times a$ Stiefel manifold $\mathbb{C}_a^d$), then*

$$\mathbb{E}_Q \chi_x\left(QAQ^*\right) = \frac{d!}{a!} D^{a-d} q(x).$$

We prove this lemma in Sect. 2. Lemma 1.17 yields the following corollary, which may be viewed as a generalization of the definition of $\boxplus_d$ to polynomials of degree greater than $d$.

**Corollary 1.18** *If $A$ and $B$ are $a \times a$ and $b \times b$ Hermitian matrices with $a, b \geq d$ and characteristic polynomials $p(x)$ and $q(x)$ respectively, and $R, Q$ are uniformly random from $\mathbb{C}_a^d$ and $\mathbb{C}_b^d$ respectively, then*

$$p(x) \boxplus_d q(x) \overset{\text{def}}{=} \mathbb{E}_{Q,R} \chi_x \left( RAR^* + QBQ^* \right) = \frac{d!}{a!} D^{a-d} p(x) \boxplus_d \frac{d!}{b!} D^{b-d} q(x).$$

$$(11)$$

*Proof* Since the formula (1) is bilinear in the characteristic polynomials $\chi_x(A)$ and $\chi_x(B)$, we have for fixed $R$:

$$\mathbb{E}_{Q \in \mathbb{C}_b^d} \chi_x \left( RAR^* + QBQ^* \right)$$
$$= \mathbb{E}_{Q \in \mathbb{C}_b^d} \mathbb{E}_{U \in \mathbb{C}_d^d} \chi_x \left( RAR^* + UQBQ^*U^* \right) \quad \text{by left invariance of } Q$$
$$= \mathbb{E}_{Q \in \mathbb{C}_b^d} \chi_x \left( RAR^* \right) \boxplus_d \chi_x \left( QBQ^* \right) \quad \text{by (1)}$$
$$= \chi_x \left( RAR^* \right) \boxplus_d \mathbb{E}_{Q \in \mathbb{C}_b^d} \chi_x \left( QBQ^* \right) \quad \text{by bilinearity of } \boxplus_d$$
$$= \chi_x \left( RAR^* \right) \boxplus_d \frac{c!}{d!} D^{b-d} \chi_x(B) \quad \text{by Lemma 1.17.}$$

Averaging over $R$ and invoking bilinearity and Lemma 1.17 once more finishes the proof. $\square$

Note that the definition (11) is consistent with Lemma 1.16, e.g. if $b = d$ then the right hand side of (11) is equal to $p(x) \boxplus_a q(x)$. As differentiation preserves real-rootedness, the generalized definition of $\boxplus_d$ preserves real-rootedness of polynomials of all degrees, and bounds on their roots may be obtained from our results by reducing to the equal degree case by differentiating sufficiently many times.

While Lemma 1.17 can be used in conjunction with $\boxtimes_d$ and $\boxplus\!\boxplus_d$ just as easily due to their bilinearity, this does not correspond to the appropriate degree-reduction operators (see Lemmas 4.9 and 4.16) for those cases, so it does not lead to a satisfying generalization of the definitions to higher degrees.

## 1.5 Notation and organization

Let $\mathbb{P}(d)$ be the family of real rooted polynomials of degree exactly $d$ with positive leading coefficient, and let $\mathbb{P}$ be the union over $d$ of $\mathbb{P}(d)$. Let $\mathbb{P}^+(d)$ be the subset of these polynomials having only nonnegative roots. We let $\mathbb{P}^+$ be the union of $\mathbb{P}^+(d)$ over all $d \geq 1$.

For a function $f(x)$, we write the derivative as $Df(x)$. For a number $\alpha$, we let $U_\alpha$ be the operator that maps $f$ to $f - \alpha Df$. That is, $U_\alpha$ is multiplication by $1 - \alpha D$.

## 2 Equivalence of convolutions and $\mathbb{E}\chi$

The goal of this section is to prove Theorems 1.2, 1.5, and 1.8 relating the three convolutions to random matrices. While we have so far only considered averages over unitary matrices, it turns out that one can average over various other collections of matrices and get the same formula. In Sect. 2.1, we will define a property that we call *minor-orthogonality* and then in Sect. 2.2 we will show that the coefficients we are interested can be computed using an average over any collection of minor-orthogonal matrices. Also in Sect. 2.1, we will show that the collection of $n \times n$ signed permutation matrices (under a uniform distribution) is minor-orthogonal, and then we will use this to show that the orthogonal matrices (under the Haar measure) is minor-orthogonal.

There are some advantages to being able to express the convolutions as averages over different collections of matrices; in particular, a formula that is easily derived by replacing a unitary average by one over signed permutation matrices will be used in the proof of Lemma 4.25.

### 2.1 Minor-orthogonality

We will write $[n]$ to denote the set $\{1, \ldots, n\}$ and for a set $S$, we write $\binom{S}{k}$ to denote the collection of subsets of $S$ that have exactly $k$ elements. When our sets contain integers (which they always will), we will consider the set to be ordered from smallest to largest. Hence, for example, if $S$ contains the elements $\{2, 5, 3\}$, then we will write

$$S = \{s_1, s_2, s_3\} \quad \text{where} \quad s_1 = 2, s_2 = 3, s_3 = 5.$$

Now let $S = \{s_1, \ldots, s_k\} \in \binom{[n]}{k}$. For a set $W \in \binom{[k]}{j}$ with $j \leq k$, we will write

$$W(S) = \{s_i : i \in W\}.$$

Lastly, for a set of integers $S$, we will write

$$\|S\|_1 = \sum_{s \in S} s.$$

**Example 2.1** For $W = \{1, 3\}$ and $S = \{2, 4, 5\}$ we have

$$W(S) = \{2, 5\} \wedge \|W\|_1 = 1 + 3 = 4 \wedge \|S\|_1 = 2 + 4 + 5 = 11.$$

Let $m, n$ be positive integers. Given an $m \times n$ matrix $A$ and sets $S \subseteq [m], T \subseteq [n]$ with $|S| = |T|$, we will write the $(S, T)$-*minor of* $A$ as

$$[A]_{S,T} = \det\left(\{a_{i,j}\}_{i \in S, j \in T}\right)$$

By definition, we will set $[A]_{\varnothing,\varnothing} = 1$. A well-known theorem of Cauchy and Binet relates the minor of a product to the product of minors ( [22]):

**Theorem 2.2** *For integers $m, n, p, k$ and $m \times n$ matrix $A$ and $n \times p$ matrix $B$, we have*

$$[AB]_{S,T} = \sum_{|U| \in \binom{[n]}{k}} [A]_{S,U}[B]_{U,T} \tag{12}$$

*for any sets $S \in \binom{[m]}{k}$ and $T \in \binom{[p]}{k}$.*

**Definition 2.3** We will say that an $m \times n$ random matrix $R$ is *minor-orthogonal* if for all integers $k, \ell \le \max\{m, n\}$ and all sets $S, T, U, V$ with $|S| = |T| = k$ and $|U| = |V| = \ell$, we have

$$\mathbb{E}_R \left\{ [R]_{S,T}[R^*]_{U,V} \right\} = \frac{1}{\binom{\max\{m,n\}}{k}} \delta_{\{S=V\}} \delta_{\{T=U\}}.$$

Given a minor-orthogonal matrix $R$ it is easy to see from the definition that

1. $R^*$ is minor orthogonal
2. any submatrix that preserves the largest dimension of $R$ is minor orthogonal

**Lemma 2.4** *If $R$ is minor-orthogonal and $Q$ is a fixed matrix for which $QQ^* = I$, then $QR$ is minor-orthogonal. If $Q^*Q = I$ then $RQ$ is minor orthogonal.*

**Proof** For any sets $S, T$ with $|S| = |T| = k$, we have

$$[QR]_{S,T} = \sum_{|W|=k} [Q]_{S,W}[R]_{W,T}$$

so for $|U| = |V| = \ell$, we have

$$\mathbb{E}_R \left\{ [QR]_{S,T}[(QR)^*]_{U,V} \right\} = \mathbb{E}_R \left\{ \sum_{|W|=k} \sum_{|Z|=\ell} [Q]_{S,W}[R]_{W,T}[R^*]_{U,Z}[Q^*]_{Z,V} \right\}$$

$$= \sum_{|W|=k} \sum_{|Z|=\ell} [Q]_{S,W}[Q^*]_{Z,V} \frac{1}{\binom{\max\{m,n\}}{k}} \delta_{\{W=Z\}} \delta_{\{T=U\}}$$

$$= \sum_{|W|=k} \frac{1}{\binom{\max\{m,n\}}{k}} [Q]_{S,W}[Q^*]_{W,V} \delta_{\{T=U\}}$$

$$= \frac{1}{\binom{\max\{m,n\}}{k}} \delta_{\{S=V\}} \delta_{\{T=U\}},$$

where the last line comes from the fact that $[I]_{S,V} = \delta_{\{S=V\}}$.

The other case $RQ$ follows by repeating the argument with $R^*$ and noting that $RQ = (Q^*R^*)^*$. $\qquad\square$

**Definition 2.5** A *signed permutation matrix* is a matrix that can be written $EP$ where $E$ is a diagonal matrix with $\pm 1$ diagonal entries and $P$ is a permutation matrix.

**Lemma 2.6** *A uniformly random $n \times n$ signed permutation matrix is minor-orthogonal.*

**Proof** We can write a uniformly random signed permutation matrix $Q$ as $Q = E_\chi P_\pi$ where $P_\pi$ is a uniformly random permutation matrix and $E_\chi$ is a uniformly random diagonal matrix with $\chi \in \{\pm 1\}^n$ on the diagonal (and the two are independent). Hence for $|S| = |T| = k$ and $|U| = |V| = \ell$, we have

$$\mathbb{E}_Q \left\{ [Q]_{S,T} [Q^*]_{U,V} \right\} = \mathbb{E}_{\chi,\pi} \left\{ [E_\chi P_\pi]_{S,T} [P_\pi^* E_\chi]_{U,V} \right\}$$

$$= \sum_{|W|=k} \sum_{|Z|=\ell} \mathbb{E}_{\chi,\pi} \left\{ [E_\chi]_{S,W} [P_\pi]_{W,T} [P_\pi^*]_{U,Z} [E_\chi]_{Z,V} \right\}.$$

$$= \mathbb{E}_{\chi,\pi} \left\{ [E_\chi]_{S,S} [P_\pi]_{S,T} [P_\pi^*]_{U,V} [E_\chi]_{V,V} \right\}$$

$$\mathbb{E}_\chi \left\{ \prod_{i \in S} \chi_i \prod_{j \in V} \chi_j \right\} \mathbb{E}_\pi \left\{ [P_\pi]_{S,T} [P_\pi^*]_{U,V} \right\}.$$

where the penultimate line uses the fact that a diagonal matrix $X$ satisfies $[X]_{A,B} = 0$ whenever $A \neq B$. Now the $\chi_i$ are uniformly distributed $\{\pm 1\}$ random variables, so

$$\mathbb{E}_\chi \left\{ \prod_{i \in S} \chi_i \prod_{j \in V} \chi_j \right\} = \delta_{\{S=V\}}$$

and so we have

$$\mathbb{E}_Q \left\{ [Q]_{S,T} [Q^*]_{U,V} \right\} = \mathbb{E}_\pi \left\{ [P_\pi]_{S,T} [P_\pi^*]_{U,V} \right\} \delta_{\{S=V\}}$$

$$= \mathbb{E}_\pi \left\{ [P_\pi]_{S,T} [P_\pi]_{S,U} \right\} \delta_{\{S=V\}}$$

Furthermore, $[P_\pi]_{S,T} = 0$ except when $T = \pi(S)$, so in order for both $[P_\pi]_{S,T}$ and $[P_\pi]_{S,U}$ to be nonzero simultaneously requires $U = T$. In the case that $U = T = \pi(S)$, $[P_\pi]_{S,T} = \pm 1$, and so we have

$$\mathbb{E}_Q \left\{ [Q]_{S,T} [Q^*]_{U,V} \right\} = \mathbb{E}_\pi \left\{ [P_\pi]_{S,T}^2 \right\} \delta_{\{S=V\}} \delta_{\{T=U\}}$$

$$= \mathbb{E}_\pi \left\{ \delta_{\{\pi(S)=T\}} \right\} \delta_{\{S=V\}} \delta_{\{T=U\}}$$

The probability that a permutation of length $n$ maps a set $S$ to a set $T$ with $|S| = |T| = k$ is

$$\frac{k!(n-k)!}{n!} = \frac{1}{\binom{n}{k}}$$

and so for $|S| = |T| = k$, we have

$$\mathbb{E}_Q \left\{ [Q]_{S,T} [Q^*]_{U,V} \right\} = \frac{1}{\binom{n}{k}} \delta_{\{S=V\}} \delta_{\{T=U\}}$$

as required. The case $m > n$ follows by considering $R^*$ instead of $R$. □

**Corollary 2.7** *An $m \times n$ Haar random matrix from the Stiefel manifold $\mathbb{C}_n^m$ is minor-orthogonal.*

**Proof** Let $R$ be such a random matrix, and assume $m \leq n$. As a signed permutation matrix is unitary, $RQ$ is also Haar distributed for any fixed signed permutation matrix $Q$. By Lemma 2.4 it is also minor-orthogonal. Hence

$$\mathbb{E}_R \left\{ [R]_{S,T} [R^*]_{U,V} \right\} = \mathbb{E}_R \left\{ [RQ]_{S,T} [(RQ)^*]_{U,V} \right\}$$

and so

$$\mathbb{E}_R \left\{ [R]_{S,T} [R^*]_{U,V} \right\} = \mathbb{E}_{R,Q} \left\{ [R]_{S,T} [R^*]_{U,V} \right\} = \mathbb{E}_{R,Q} \left\{ [RQ]_{S,T} [(RQ)^*]_{U,V} \right\},$$

where $Q$ is a uniform signed permutation independent from $R$. By Lemma 2.6, $Q$ is minor-orthogonal and so, for fixed $R$, Lemma 2.4 implies that $RQ$ is minor-orthogonal. So

$$\mathbb{E}_R \left\{ \mathbb{E}_Q \left\{ [RQ]_{S,T} [(RQ)^*]_{U,V} \right\} \right\} = \mathbb{E}_R \left\{ \frac{1}{\binom{n}{k}} \delta_{\{S=V\}} \delta_{\{T=U\}} \right\} = \frac{1}{\binom{n}{k}} \delta_{\{S=V\}} \delta_{\{T=U\}}$$

as required. □

### 2.2 Formulas

We begin this section by mentioning a well-known formula for the determinants of a sum of matrices (see [23]):

**Theorem 2.8** *For integers $k \leq n$, $n \times n$ matrices $A$, $B$, and sets $S, T \in \binom{[n]}{k}$, we have*

$$[A + B]_{S,T} = \sum_{i=0}^{k} \sum_{U,V \in \binom{[k]}{i}} (-1)^{\|U\|_1 + \|V\|_1} [A]_{U(S),V(T)} [B]_{\overline{U}(S),\overline{V}(T)}, \quad (13)$$

*where $\overline{U} = [k] \setminus U$.*

We denote the coefficient of $(-1)^k x^{d-k}$ of the characteristic polynomial of a $d$-dimensional matrix $A$ by $\sigma_k(A)$, which we recall is the $k$th elementary symmetric function of the eigenvalues of $A$. We will repeatedly use the fact that

$$\sigma_k(A) = \sum_{|S|=k} [A]_{S,S}. \quad (14)$$

**Lemma 2.9** *For integers $m \leq n$, let $B$ be an $n \times n$ matrix and let $R$ be a $m \times n$ minor-orthogonal matrix independent from $B$. Then for all $|S| = |T| = k$, we have*

$$\mathbb{E}_R \left\{ [RBR^*]_{S,T} \right\} = \delta_{\{S=T\}} \frac{\sigma_k(B)}{\binom{n}{k}}$$

*Proof* By Theorem 2.2,

$$
\begin{aligned}
\mathbb{E}_R \left\{ [RBR^*]_{S,T} \right\} &= \sum_{X,Y \in \binom{[n]}{k}} \mathbb{E}_R \left\{ [R]_{S,X}[B]_{X,Y}[R^*]_{Y,T} \right\} \\
&= \sum_{X,Y \in \binom{[n]}{k}} [B]_{X,Y} \mathbb{E}_R \left\{ [R]_{S,X}[R]_{T,Y} \right\} \\
&= \sum_{X,Y \in \binom{[n]}{k}} [B]_{X,Y} \frac{1}{\binom{n}{k}} \delta_{\{X=Y\}} \delta_{\{S=T\}} \\
&= \sum_{X \in \binom{[n]}{k}} [B]_{X,X} \frac{1}{\binom{n}{k}} \delta_{\{S=T\}}.
\end{aligned}
$$

$\square$

The above lemma yields a quick proof of Lemma 1.17.

*Proof of Lemma 1.17* Let $A$ be an $a \times a$ matrix and let $Q$ be Haar distributed on $\mathbb{C}_a^d$. The $k^{th}$ coefficient of $\mathbb{E}\chi_x(QAQ^*)$ is equal to

$$
\begin{aligned}
\mathbb{E}_Q \left\{ \sigma_k(QAQ^*) \right\} &= \sum_{|S|=k} \mathbb{E}_Q \left\{ [QAQ^*]_{S,S} \right\} \\
&= \sum_{|S|=k} \frac{\sigma_k(A)}{\binom{a}{k}} \quad \text{by Lemma 2.9 and minor-orthogonality of } Q \\
&= \frac{\binom{d}{k}}{\binom{a}{k}},
\end{aligned}
$$

which is precisely the $k^{th}$ coefficient of $\frac{d!}{a!} D^{a-d} \chi_x(Q)$.  $\square$

### 2.2.1 Symmetric additive and multiplicative convolutions

Using Lemma 2.9, we can easily prove Theorems 1.2 and 1.5 by showing equality of each coefficient as per (14).

**Theorem 2.10** (Implies Theorem 1.2) *Let $A$ and $B$ be $d \times d$ matrices, and let $R$ be a random $d \times d$ minor-orthogonal matrix. Then*

$$\mathbb{E}_R \left\{ \sigma_k(A + RBR^*) \right\} = \sum_{i=0}^{k} \frac{\binom{d-i}{k-i}}{\binom{d}{k-i}} \sigma_i(A)\sigma_{k-i}(B).$$

**Proof** By (14), Theorem 2.8, and then Lemma 2.9, we have

$$
\mathbb{E}_R \left\{ \sigma_k(A + RBR^*) \right\}
$$
$$
= \sum_{S \in \binom{[d]}{k}} \mathbb{E}_R \left\{ [A + RBR^*]_{S,S} \right\}
$$
$$
= \sum_{S \in \binom{[d]}{k}} \sum_{i=0}^{k} \sum_{U,V \in \binom{[k]}{i}} (-1)^{\|U\|_1 + \|V\|_1} [A]_{U(S),V(S)} \mathbb{E}_R \left\{ [RBR^*]_{\overline{U}(S), \overline{V}(S)} \right\}
$$
$$
= \sum_{S \in \binom{[d]}{k}} \sum_{i=0}^{k} \sum_{U,V \in \binom{[k]}{i}} (-1)^{\|U\|_1 + \|V\|_1} [A]_{U(S),V(S)} \delta_{\{\overline{U}(S) = \overline{V}(S)\}} \frac{\sigma_{k-i}(B)}{\binom{d}{k-i}}
$$
$$
= \sum_{i=0}^{k} \frac{\sigma_{k-i}(B)}{\binom{d}{k-i}} \sum_{S \in \binom{[d]}{k}} \sum_{U \in \binom{[k]}{i}} [A]_{U(S),U(S)}
$$

where the last equality uses the fact that $\overline{U}(S) = \overline{V}(S)$ if and only if $U = V$. Finally,

$$
\sum_{S \in \binom{[d]}{k}} \sum_{U \in \binom{[k]}{i}} [A]_{U(S),U(S)} = \binom{d-i}{k-i} \sum_{V \in \binom{[d]}{i}} [A]_{V,V}
$$

as $\binom{d-i}{k-i}$ is the number of times a set $V$ appears as $V = U(S)$ for some $S$ and some $U$. That is, the number of ways we can add elements to a set of size $V$ to obtain a set of size $k$. $\qquad \square$

Using Theorem 2.10 and Lemma 2.6 we can derive another useful formula for the symmetric additive convolution, this time as a function of the *roots*. It states that $p(x) \boxplus_d q(x)$ is the average of all polynomials you can form by adding the roots of $p$ and $q$ pairwise.

**Theorem 2.11** *For $p(x) = \prod_{i=0}^{d}(x - a_i)$ and $q(x) = \prod_{i=0}^{d}(x - b_i)$ we have*

$$
p(x) \boxplus_d q(x) = \frac{1}{d!} \sum_{\sigma} \prod_{i=1}^{d} (x - a_i - b_{\sigma(i)})
$$

*where the sum is over permutations $\sigma$ of $[d]$.*

**Proof** Let $A$ be the diagonal matrix with diagonal elements $\{a_i\}$ and let $B$ be the diagonal matrix with diagonal elements $\{b_i\}$. By Theorem 2.10, we have

$$
p(x) \boxplus_d q(x) = \det(xI - A) \boxplus_d \det(xI - B)
$$
$$
= \mathbb{E}_R \left\{ \det\left(xI - A - RBR^*\right) \right\}.
$$

where the expectation can be taken over any minor-orthogonal random matrix $R$. By Lemma 2.6, we can take $R$ to be a (uniformly random) signed permutation matrix. Since $A$ and $B$ are diagonal, it is easy to compute (for each fixed value of $R$)

$$\det\left(xI - A - RBR^*\right) = \prod_i (x - a_i - b_{\sigma(i)})$$

where $\sigma$ is the permutation part of $R$ (all of the signs cancel). Averaging over these gives the result. $\qquad\square$

**Theorem 2.12** (Implies Theorem 1.5) *Let $A$ and $B$ be $d \times d$ matrices, and let $R$ be an $d \times d$ minor-orthogonal matrix. Then*

$$\mathbb{E}_R\left\{\sigma_k(ARBR^*)\right\} = \frac{\sigma_k(A)\sigma_k(B)}{\binom{d}{k}}.$$

**Proof** By Theorem 2.2 and then Lemma 2.9, we have

$$\begin{aligned}
\mathbb{E}_R\left\{\sigma_k(ARBR^*)\right\} &= \sum_{S \in \binom{[d]}{k}} \mathbb{E}_R\left\{[ARBR^*]_{S,S}\right\} \\
&= \sum_{S,T \in \binom{[d]}{k}} [A]_{S,T}\mathbb{E}_R\left\{[RBR^*]_{T,S}\right\} \\
&= \sum_{S,T \in \binom{[d]}{k}} [A]_{S,T}\delta_{\{T=S\}}\frac{\sigma_k(B)}{\binom{d}{k}} \\
&= \frac{\sigma_k(A)\sigma_k(B)}{\binom{d}{k}}.
\end{aligned}$$

$\qquad\square$

### 2.2.2 Asymmetric additive convolution

The proof of the asymmetric convolution is a bit more involved, due to the appearance of a second random matrix.

**Lemma 2.13** *Let $B$ be a $d \times d$ matrix and let $Q$ and $R$ be $d \times d$ independently random minor-orthogonal matrices. Then*

$$\mathbb{E}_{Q,R}\left\{[QBR^*]_{S,T}[(QBR^*)^*]_{U,V}\right\} = \frac{\delta_{\{T=U\}}\delta_{\{S=V\}}}{\binom{d}{k}\binom{d}{k}}\sigma_k(BB^*)$$

*for any $|S| = |T| = k$ and $|U| = |V| = \ell$.*

***Proof*** By Theorem 2.2 we have

$$[QBR^*]_{S,T}[(QBR^*)^*]_{U,V}$$
$$= \sum_{|W|=|Z|=k} \sum_{|W'|=|Z'|=\ell} [Q]_{S,W}[B]_{W,Z}[R^*]_{Z,T}[R]_{U,W'}[B^*]_{W',Z'}[Q^*]_{Z',V}$$

where

$$\mathbb{E}_Q\left\{[Q]_{S,W}[Q^*]_{Z',V}\right\} = \frac{1}{\binom{d}{k}}\delta_{\{S=V\}}\delta_{\{W=Z'\}}$$

and

$$\mathbb{E}_R\left\{[R^*]_{Z,T}[R]_{U,W'}\right\} = \frac{1}{\binom{d}{\ell}}\delta_{\{T=U\}}\delta_{\{Z=W'\}}.$$

Hence

$$\mathbb{E}_{Q,R}\left\{[QBR^*]_{S,T}[(QBR^*)^*]_{U,V}\right\} = \sum_{|W|=|Z|=k}\frac{1}{\binom{d}{k}}\frac{1}{\binom{d}{k}}[B]_{W,Z}[B^*]_{Z,W}\delta_{\{T=U\}}\delta_{\{S=V\}}$$
$$= \sum_{|W|=k}\frac{1}{\binom{d}{k}}\frac{1}{\binom{d}{k}}[BB^*]_{W,W}\delta_{\{T=U\}}\delta_{\{S=V\}}$$
$$= \frac{1}{\binom{d}{k}}\frac{1}{\binom{d}{k}}\sigma_k(BB^*)\delta_{\{T=U\}}\delta_{\{S=V\}}$$

$\square$

**Theorem 2.14** (Implies Theorem 1.8) *Let A and B be $d \times d$ matrices and let Q and R be $d \times d$ minor-orthogonal matrices that are independent from A, B and each other. Then*

$$\mathbb{E}_{Q,R}\left\{\sigma_k((A+QBR^*)(A+QBR^*)^*)\right\} = \sum_i \frac{\binom{d-i}{k-i}\binom{d-i}{k-i}}{\binom{d}{k-i}\binom{d}{k-i}}\sigma_i(AA^*)\sigma_{k-i}(BB^*).$$

***Proof*** Let $|U| = k$. Then by Theorem 2.2 we have

$$[(A+QBR^*)(A+QBR^*)^*]_{U,U} = \sum_{|V|=k}[A+QBR^*]_{U,V}[(A+QBR^*)^*]_{V,U}.$$

By Theorem 2.8,

$$[A+QBR^*]_{U,V} = \sum_{i=0}^k \sum_{|W|=|Z|=i}(-1)^{\|W\|_1+\|Z\|_1}[A]_{W(U),Z(V)}[QBR^*]_{\overline{W}(U),\overline{Z}(V)}$$

and

$$[(A + QBR^*)^*]_{V,U}$$
$$= \sum_{i=0}^{k} \sum_{|W'|=|Z'|=i} (-1)^{\|W'\|_1 + \|Z'\|_1} [A^*]_{W'(V), Z'(U)} [(QBR^*)^*]_{\overline{W'}(V), \overline{Z'}(U)}.$$

Applying Lemma 2.13 then gives

$$\mathbb{E}_{Q,R} \left\{ [QBR^*]_{\overline{W}(U), \overline{Z}(V)} [(QBR^*)^*]_{\overline{W'}(V), \overline{Z'}(U)} \right\}$$
$$= \frac{\delta_{\{\overline{W'}(V) = \overline{Z}(V)\}} \delta_{\{\overline{W}(U) = \overline{Z'}(U)\}}}{\binom{d}{k-i} \binom{d}{k-i}} \sigma_{k-i}(BB^*)$$
$$= \frac{\delta_{\{W'=Z\}} \delta_{\{W=Z'\}}}{\binom{d}{k-i} \binom{d}{k-i}} \sigma_{k-i}(BB^*)$$

where again we use the fact that $A(X) = B(X)$ if and only if $A = B$. Hence

$$\mathbb{E}_{Q,R} \left\{ [(A + QBR^*)(A + QBR^*)^*]_{U,U} \right\}$$
$$= \frac{1}{\binom{d}{k-i}} \frac{1}{\binom{d}{k-i}} \sigma_{k-i}(BB^*) \sum_{|V|=k} \sum_{i=0}^{k} \sum_{|W|=|Z|=i} [A]_{W(U), Z(V)} [A^*]_{Z(V), W(U)}.$$

Similar to Theorem 2.10, we have

$$\sum_{|U|=|V|=k} \sum_{|W|=|Z|=i} [A]_{W(U), Z(V)} [A^*]_{Z(V), W(U)} = \theta_{i,k} \sum_{|S|=|T|=i} [A]_{S,T} [A^*]_{T,S}$$
$$= \theta_{i,k} \sum_{|S|=i} [AA^*]_{S,S}$$
$$= \theta_{i,k} \sigma_i(AA^*),$$

where $\theta_{i,k}$ is the number of ways to complete $S$ to a set of size $k$ *and* complete $T$ to a set of size $k$. Since $S \subseteq [d]$ and $T \subseteq [d]$, this is precisely $\binom{d-i}{k-i} \binom{d-i}{k-i}$, completing the proof. $\qquad \square$

## 3 Real rootedness of the asymmetric additive convolution

We will use the theory of stable polynomials to prove Theorem 1.9 (see e.g. [24] for an introduction). Stability is a multivariate generalization of real-rootedness which is preserved under a rich and well-understood class of linear transformations, and our approach is to realize $p \boxplus_d q(x)$ as a univariate restriction of a multivariate stable polynomial. For this theorem, we will require Hurwitz stable polynomials. We recall

that a multivariate polynomial $p(z_1, \ldots, z_m) \in \mathbb{R}[z_1, \ldots, z_m]$ is *Hurwitz stable* if it is identically zero or if whenever the real part of $z_i$ is positive for all $i$, $p(z_1, \ldots, z_m) \neq 0$.

Instead of proving Theorem 1.9 directly, we prove the following theorem from which it follows by substituting $-x$ for $x$. We define $\mathbb{P}^-(d)$ to be the subset of polynomials in $\mathbb{P}(d)$ having only nonpositive roots and $\mathbb{P}^-$ to be the union over $\mathbb{P}^-(d)$ for $d \geq 1$.

**Theorem 3.1** *Let*

$$p(x) = \sum_{i=0}^{d} x^{d-i} a_i \quad and \quad q(x) = \sum_{i=0}^{d} x^{d-i} b_i$$

*be in* $\mathbb{P}^-(d)$. *Then,*

$$r(x) \stackrel{\text{def}}{=} \sum_{k=0}^{d} x^{d-k} \sum_{i+j=k} \left( \frac{(d-i)!(d-j)!}{d!(d-i-j)!} \right)^2 a_i b_j$$

*is also in* $\mathbb{P}^-(d)$.

We will use the following result to prove that a polynomial is in $\mathbb{P}^-$.

**Lemma 3.2** *Let* $r(x)$ *be a polynomial such that* $h(x, y) = r(xy)$ *is Hurwitz stable. Then,* $r \in \mathbb{P}^-$.

**Proof** Let $\zeta$ be any root of $r$. If $\zeta$ is neither zero or negative, then it has a square root with positive real part. Setting both $x$ and $y$ to this square root would contradict the Hurwitz stability of $h$. □

We will prove that $r(x)$ is in $\mathbb{P}^-$ by constructing a Hurwitz stable polynomial and applying Lemma 3.2. To this end, we need a few tools for constructing Hurwitz stable polynomials.

The first is elementary.

**Claim 3.3** *If* $p(x) \in \mathbb{P}^-$, *then the polynomial* $f(x, y) = p(xy)$ *is Hurwitz stable.*

**Proof** If both $x$ and $y$ have positive real part, then $xy$ cannot be a nonpositive real, and thus cannot be a root of $p$. □

The second tool is the following result of Borcea and Brändén, which is a consequence of Corollary 5.9 of [25].

**Proposition 3.4** (Polarization) *Let*

$$p(x, y) = \sum_{i=0}^{d} \sum_{j=0}^{d} c_{i,j} x^i y^j$$

*be a Hurwitz stable polynomial. For each integer $i$, let $\sigma_i^x$ be the $i$th elementary symmetric polynomial in the variables $x_1, \ldots, x_d$, and let $\sigma_i^y$ be the corresponding polynomial in $y_1, \ldots, y_d$. Then, the polynomial*

$$P(x_1, \ldots, x_d, y_1, \ldots, y_d) \overset{\text{def}}{=} \sum_{i=0}^{d} \sum_{j=0}^{d} c_{i,j} \frac{\sigma_i^x \sigma_j^y}{\binom{d}{i}\binom{d}{j}}$$

*is Hurwitz stable.*

The polynomial $P$ is called the *polarization* of $p$. We remark that $P(x, \ldots, x, y, \ldots, y) = p(x, y)$.

The last result we need is due to Lieb and Sokal [26] (see also [27, Theorem 8.4]).

**Theorem 3.5** *Let $P(z_1, \ldots, z_d)$ and $Q(z_1, \ldots, z_d)$ be Hurwitz stable polynomials. Let $D_i^z$ denote the derivative with respect to $z_i$. Then,*

$$Q(D_1^z, \ldots, D_d^z) P(z_1, \ldots, z_d)$$

*is Hurwitz stable.*

**Proof of Theorem 3.1** Define $f(x, y) = p(xy)$ and $g(x, y) = (xy)^d q(1/xy)$. Let $F(x_1, \ldots, x_d, y_1, \ldots, y_d)$ be the polarization of $f(x, y)$ in the variables $x_1, \ldots, x_d, y_1, \ldots, y_d$. Let $G(x_1, \ldots, x_d, y_1, \ldots, y_d)$ be the analogous polarization of $g(x, y)$.

Let $\sigma_i^x$ be the $i$th elementary symmetric function in $x_1, \ldots, x_d$, and let $\delta_i^x$ be the $i$th elementary symmetric function in $D_1^x, \ldots, D_d^x$. Define $\sigma_i^y$ and $\delta_i^y$ analogously. Then,

$$F(x_1, \ldots, x_d, y_1, \ldots, y_d) = \sum_{i=0}^{d} a_i \frac{\sigma_{d-i}^x \sigma_{d-i}^y}{\binom{d}{i}^2},$$

and

$$G(D_1^x, \ldots, D_d^x, D_1^y, \ldots, D_d^y) = \sum_{i=0}^{d} b_i \frac{\delta_i^x \delta_i^y}{\binom{d}{i}^2}.$$

Define

$$H(x_1, \ldots, x_d, y_1, \ldots, y_d) = G(D_1^x, \ldots, D_d^x, D_1^y, \ldots, D_d^y) F(x_1, \ldots, x_d, y_1, \ldots, y_d).$$

We know from Theorem 3.5 that $H$ is Hurwitz stable. Define

$$h(x, y) = H(x, \ldots, x, y, \ldots, y).$$

It is immediate that $h$ is Hurwitz stable too. We will prove that $h(x, y) = r(xy)$, which by Lemma 3.2 implies that $r$ is in $\mathbb{P}^-$.

It will be convenient to know that

$$\delta_i^x \sigma_j^x = \begin{cases} \binom{d+i-j}{i} \sigma_{j-i}^x & \text{if } i \leq j \\ 0 & \text{otherwise.} \end{cases}$$

We may now compute

$$
\begin{aligned}
H(x_1, \ldots x_d, y_1, \ldots, y_d) &= \sum_{i=0}^{d} b_i \frac{\delta_i^x \delta_i^y}{\binom{d}{i}^2} \sum_{j=0}^{d} a_j \frac{\sigma_{d-j}^x \sigma_{d-j}^y}{\binom{d}{j}^2} \\
&= \sum_{i=0}^{d} \sum_{j:i \leq d-j} \frac{b_i}{\binom{d}{i}^2} \frac{a_j}{\binom{d}{j}^2} \delta_i^x \delta_i^y \sigma_{d-j}^x \sigma_{d-j}^y \\
&= \sum_{i+j \leq d} \frac{a_j b_i}{\binom{d}{i}^2 \binom{d}{j}^2} \delta_i^x \delta_i^y \sigma_{d-j}^x \sigma_{d-j}^y \\
&= \sum_{i+j \leq d} \frac{a_j b_i}{\binom{d}{i}^2 \binom{d}{j}^2} \binom{d+i-(d-j)}{i}^2 \sigma_{d-i-j}^x \sigma_{d-i-j}^y \\
&= \sum_{i+j \leq d} a_j b_i \frac{\binom{i+j}{i}^2}{\binom{d}{i}^2 \binom{d}{j}^2} \sigma_{d-i-j}^x \sigma_{d-i-j}^y \\
&= \sum_{k=0}^{d} \sum_{i+j=k} a_j b_i \frac{\binom{k}{i}^2}{\binom{d}{i}^2 \binom{d}{j}^2} \sigma_{d-k}^x \sigma_{d-k}^y.
\end{aligned}
$$

So,

$$
\begin{aligned}
h(x, y) &= \sum_{k=0}^{d} \sum_{i+j=k} a_j b_i \left( \frac{\binom{d}{k}\binom{k}{i}}{\binom{d}{i}\binom{d}{j}} \right)^2 x^{d-k} y^{d-k}. \\
&= \sum_{k=0}^{d} \sum_{i+j=k} a_j b_i \left( \frac{(d-i)!(d-j)!}{d!(d-i-j)!} \right)^2 x^{d-k} y^{d-k}.
\end{aligned}
$$

So, $r(xy) = h(x, y)$ and therefore must have only nonpositive real roots. $\qquad \square$

## 4 Transform bounds

In this section we prove Theorems 1.12, 1.13, and 1.14. All of our transform bounds are proved using the following lemma. It allows us to pinch together two of the roots of a polynomial without changing the value of the Cauchy transform at a particular

point. Through judicious use of this lemma, we are able to reduce statements about arbitrary polynomials to statements about polynomials with just one root.

**Lemma 4.1** (Pinching) *Let $\alpha > 0$, $d \geq 2$, and let $p(x) \in \mathbb{P}(d)$ have at least two distinct roots. Write $p(x) = \prod_{i=1}^{d}(x - \lambda_i)$ where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$ and $\lambda_1 > \lambda_k$ for some k. Then there exist real $\mu$ and $\rho$ so that $p(x) = \widehat{p}(x) + \widetilde{p}(x)$, where*

$$\widetilde{p}(x) = (x - \mu)^2 \prod_{i \notin \{1,k\}} (x - \lambda_i) \in \mathbb{P}(d) \quad and$$

$$\widehat{p}(x) = (x - \rho) \prod_{i \notin \{1,k\}} (x - \lambda_i) \in \mathbb{P}(d - 1),$$

*and*

a. $\mathsf{maxroot}\,(U_\alpha \widetilde{p}) = \mathsf{maxroot}\,(U_\alpha \widehat{p}) = \mathsf{maxroot}\,(U_\alpha p)$,
b. $\lambda_1 > \mu > \lambda_k$, and
c. $\rho > \lambda_1$. In particular, if $d \geq 3$ then $\widehat{p}$ has at least two distinct roots.

**Proof** Let $t = \mathsf{maxroot}\,(U_\alpha p)$ and set

$$\mu = t - \frac{2}{1/(t - \lambda_1) + 1/(t - \lambda_k)}.$$

We have chosen $\mu$ so that

$$\frac{2}{t - \mu} = \frac{1}{t - \lambda_1} + \frac{1}{t - \lambda_k},$$

which implies

$$\frac{D\widetilde{p}(t)}{\widetilde{p}(t)} = \frac{Dp(t)}{p(t)} = \frac{1}{\alpha},$$

and thus $\mathsf{maxroot}\,(U_\alpha \widetilde{p}) = t$. Our choice of $\mu$ also guarantees that $t - \mu$ is the harmonic average of $t - \lambda_1$ and $t - \lambda_k$. Thus, $\mu$ must lie strictly between $\lambda_1$ and $\lambda_k$, which implies part *b*. As the harmonic mean of distinct numbers is less than their average, $t - \mu < (1/2)(2t - (\lambda_1 + \lambda_k))$, which implies that

$$\mu > (\lambda_1 + \lambda_k)/2. \tag{15}$$

We have

$$\widehat{p}(x) = p(x) - \widetilde{p}(x) = \left((2\mu - (\lambda_1 + \lambda_k))x - (\mu^2 - \lambda_1\lambda_k)\right) \prod_{i \notin \{1,k\}} (x - \lambda_i).$$

This and inequality (15) imply that $\widehat{p}(x) \in \mathbb{P}(d - 1)$. As $U_\alpha$ is linear, we also have $(U_\alpha \widehat{p})(t) = 0$. To finish the proof of part *a*, we need to show that the maximum root

of $U_\alpha \widehat{p}$ is less than $t$. The one root of $\widehat{p}$ that is not a root of $p$ is

$$\rho \overset{\text{def}}{=} \frac{\mu^2 - \lambda_1 \lambda_k}{2\mu - (\lambda_1 + \lambda_k)}.$$

To see that $t > \rho$, compute

$$t - \rho = \frac{(\lambda_1 - \mu)(\mu - \lambda_k)}{2\mu - \lambda_1 - \lambda_k},$$

which we know is greater than 0 because of (15) and the fact that $\mu$ is between $\lambda_1$ and $\lambda_k$. This completes the proof of part $a$.

To prove part $c$, note that $2\mu - (\lambda_1 + \lambda_k) > 0$, and

$$(2\mu - (\lambda_1 + \lambda_k))(\rho - \lambda_1) = \mu^2 - 2\lambda_1\mu + \lambda_1^2 = (\mu - \lambda_1)^2 > 0.$$

$\square$

The following lemma provides one of the facts we exploit about the decomposition $p(x) = \widetilde{p}(x) + \widehat{p}(x)$

**Lemma 4.2** *Let $f, g, h$ be real rooted polynomials with positive leading coefficients such that $f = g + h$. Then*

$$maxroot\,(f) \leq \max\,\{maxroot\,(g)\,, maxroot\,(h)\} \tag{16}$$

*with equality if and only if*

$$maxroot\,(f) = maxroot\,(g) = maxroot\,(h)\,. \tag{17}$$

**Proof** Note that equality in (17) clearly implies equality in (16). Now, assume by way of contradiction that (16) is false, and let $x = maxroot\,(f)$. Since $g$ and $h$ have positive leading terms, $g(x), h(x) > 0$. Thus, $f(x) = g(x) + h(x) > 0$, a contradiction . $\square$

### 4.1 Symmetric additive convolution

We now prove the upper bound on the $R$-transform of $p \boxplus_d q$.

**Theorem 4.3** (Restatement of Theorem 1.12) *For $p, q \in \mathbb{P}(d)$ and $\alpha > 0$,*

$$maxroot\,(U_\alpha(p \boxplus_d q)) + d\alpha \leq maxroot\,(U_\alpha p) + maxroot\,(U_\alpha q)\,,$$

*with equality only if $p(x)$ or $q(x)$ has the form $(x - \lambda)^d$.*

Theorem 1.2 and Lemma 1.16 tell us that if $q(x) \in \mathbb{P}(d)$ and $p(x) = x^{d-1}$, then $p(x) \boxplus_{d-1} q(x) = (1/d)Dq(x)$. As $maxroot\,(U_\alpha x^{d-1}) = (d-1)\alpha$, the following lemma may be viewed as a restriction of Theorem 4.3 to the case that $q(x) = x^{d-1}$.

**Lemma 4.4** *For $\alpha \geq 0$, $d \geq 2$, and $p \in \mathbb{P}(d)$,*

$$\mathsf{maxroot}\,(U_\alpha Dp) \leq \mathsf{maxroot}\,(U_\alpha p) - \alpha. \tag{18}$$

*with equality if and only if $p = (x - \lambda)^d$ for some real number $\lambda$.*

**Proof** If $p = (x - \lambda)^d$, then $\mathsf{maxroot}\,(U_\alpha p) = \lambda + d\alpha$ and $\mathsf{maxroot}\,(U_\alpha Dp) = \lambda + (d - 1)\alpha$, giving equality in (18).

We now prove the rest of the lemma by induction on $d$, with $d = 2$ being the base case. To establish the lemma in the case that $d = 2$ and $p$ has roots $\lambda_1 > \lambda_2$, note that

$$r \stackrel{\text{def}}{=} \mathsf{maxroot}\,(U_\alpha Dp) + \alpha = 2\alpha + (\lambda_1 + \lambda_2)/2$$

and

$$U_\alpha p(x) = x^2 - (\lambda_1 + \lambda_2 + 2\alpha)x + \alpha(\lambda_1 + \lambda_2) + \lambda_1\lambda_2.$$

As this polynomial has a positive leading term, the fact that $\mathsf{maxroot}\,(U_\alpha p) > r$ follows from

$$U_\alpha p(r) = -(\lambda_1 - \lambda_2)^2/4 < 0.$$

For a real rooted polynomial $p$, define

$$\phi(p) = \mathsf{maxroot}\,(U_\alpha p) - \alpha - \mathsf{maxroot}\,(U_\alpha Dp)\,.$$

We will prove by induction on $d$ that $\phi(p) > 0$ for all polynomials $p \in \mathbb{P}(d)$ that have more than one root.

Assume by way of contradiction that there exists a monic (without loss of generality, since $\phi$ is independent of scaling) polynomial $p \in \mathbb{P}(d)$ with at least two distinct roots for which $\phi(p) \leq 0$. Let $[-R, R]$ be an interval containing all of the roots of $p$, and define $\mathbb{P}(d)[-R, R]$ to be all monic polynomials in $\mathbb{P}(d)$ with all roots in this interval. Since $[-R, R]^d$ is a compact set and $\phi$ is a continuous function of the roots of $p$, there is a monic polynomial $p_0 \in \mathbb{P}(d)[-R, R]$ at which $\phi$ obtains its minimum. Let $p_0$ be such a polynomial, so $\phi(p_0) \leq \phi(p) \leq 0$. We may assume that $p_0$ has at least two distinct roots, because it is true if $\phi(p_0) < 0$ whereas if $\phi(p_0) = 0$ we may assume $p_0 = p$.

Lemma 4.1 implies that there exist polynomials $\widehat{p}$ and $\widetilde{p}$ with $p_0 = \widehat{p} + \widetilde{p}$ such that

1. $\widehat{p}$ and $\widetilde{p}$ have positive first coefficients
2. $\widehat{p}$ has degree $d - 1$, and if $d \geq 3$ it has at least two distinct roots.
3. $\widetilde{p} \in \mathbb{P}(d)[-R, R]$.
4. $\mathsf{maxroot}\,(U_\alpha \widetilde{p}) = \mathsf{maxroot}\,(U_\alpha \widehat{p}) = \mathsf{maxroot}\,(U_\alpha p_0)$

By linearity we have

$$U_\alpha D\widetilde{p} + U_\alpha D\widehat{p} = U_\alpha Dp_0.$$

If $\mathsf{maxroot}\,(U_\alpha D p_0) \leq \mathsf{maxroot}\,(U_\alpha D \widehat{p})$, then

$$\phi(p_0) = \mathsf{maxroot}\,(U_\alpha p_0) - \alpha - \mathsf{maxroot}\,(U_\alpha D p_0)$$
$$\geq \mathsf{maxroot}\,(U_\alpha \widehat{p}) - \alpha - \mathsf{maxroot}\,(U_\alpha D \widehat{p})$$
$$= \phi(\widehat{p})$$
$$> 0,$$

by the inductive hypothesis, as $\widehat{p}$ has degree $d-1$. As this would contradict our assumption that $\phi(p) \leq 0$, we may assume $\mathsf{maxroot}\,(U_\alpha D p_0) > \mathsf{maxroot}\,(U_\alpha D \widehat{p})$ and apply Lemma 4.2 to conclude $\mathsf{maxroot}\,(U_\alpha D p_0) < \mathsf{maxroot}\,(U_\alpha D \widetilde{p})$. This implies

$$\phi(p_0) = \mathsf{maxroot}\,(U_\alpha p_0) - \alpha - \mathsf{maxroot}\,(U_\alpha D p_0)$$
$$> \mathsf{maxroot}\,(U_\alpha \widehat{p}) - \alpha - \mathsf{maxroot}\,(U_\alpha D \widetilde{p})$$
$$= \phi(\widetilde{p}),$$

contradicting the minimality of $p_0$. Thus, we may conclude that $\phi(p) > 0$ for all $p \in \mathbb{P}(d)$ with at least two roots. □

**Lemma 4.5** *For $\alpha \geq 0$, $q = (x - \lambda)^d$ for some real $\lambda$ and $p \in \mathbb{P}(d)$,*

$$maxroot\,(U_\alpha(p \boxplus_d q)) = maxroot\,(U_\alpha p) + maxroot\,(U_\alpha q) - \alpha d.$$

***Proof*** We can prove this either by manipulating the identity in (1) and those following it, or by going through the defintion (1.1). To pursue the latter route, let $A$ be a Hermitian matrix whose characteristic polynomials is $p$ and let $B = \lambda I$. We then have

$$p(x) \boxplus_d q(x) = \mathbb{E}_Q \chi_x\,(A + \lambda I) = p(x - \lambda).$$

Thus,

$$\mathsf{maxroot}\,(U_\alpha(p \boxplus_d q)) = \lambda + \mathsf{maxroot}\,(U_\alpha p)\,.$$

On the other hand, $\mathsf{maxroot}\,(U_\alpha q) = \lambda + \alpha d$. □

**Lemma 4.6** *If $p \in \mathbb{P}(d)$ for $d \geq 3$ and $Dp$ has just one root, then $p$ has just one root.*

***Proof*** If $Dp = (x - \lambda)^{d-1}$, then $p$ can be written in the form $(x - \lambda)^d + c$ for some constant $c$. If $d \geq 3$ and $c$ were a constant other than zero, then this polynomial would have at least two complex roots. □

Our proof of Theorem 4.3 will be very similar to our proof of Lemma 4.4.

**Proof of Theorem 4.3** Lemma 4.5 proves the theorem in the case that either $p$ or $q$ can be written in the form $(x - \lambda)^d$. So, we will prove that if neither $p$ nor $q$ is of the form $(x - \lambda)^d$ then

$$\text{maxroot}\,(U_\alpha(p \boxplus_d q)) + d\alpha < \text{maxroot}\,(U_\alpha p) + \text{maxroot}\,(U_\alpha q). \qquad (19)$$

We will prove this by induction on $d$, the maximum degree of $p$ and $q$. The base case $d = 1$ is handled by Lemma 4.5. Assume $d \geq 2$ and fix a polynomial $q \in \mathbb{P}(d)$ with at least two roots. For any polynomial $p$ in $\mathbb{P}(d)$, define

$$\phi(p) = \text{maxroot}\,(U_\alpha p) + \text{maxroot}\,(U_\alpha q) - d\alpha - \text{maxroot}\,(U_\alpha(p \boxplus_d q)).$$

As before, assume for contradiction that there exists a monic polynomial $p$ with at least two roots for which $\phi(p) \leq 0$. Let $[-R, R]$ be an interval containing the roots of $p$ and let $p_0$ minimize $\phi$ over all monic polynomials whose roots are contained in this interval. We may assume that $p_0$ has at least two roots because Lemma 4.5 says it must if $\phi(p_0) < 0$, and otherwise we may take $p_0 = p$.

Thus, we can apply Lemma 4.1 to $p_0$ to obtain polynomials $\widehat{p} \in \mathbb{P}(d - 1)$ and $\widetilde{p} \in \mathbb{P}(d)$ such that

a. $p_0 = \widehat{p} + \widetilde{p}$, which by the linearity of $\boxplus_d$ and $U_\alpha$ implies

$$U_\alpha p_0 \boxplus_d q = U_\alpha \widehat{p} \boxplus_d q + U_\alpha \widetilde{p} \boxplus_d q;$$

b. the roots of $\widetilde{p}$ lie inside $[-R, R]$, and so $\phi(\widetilde{p}) \geq \phi(p_0)$; and
c. $\text{maxroot}\,(U_\alpha \widetilde{p}) = \text{maxroot}\,(U_\alpha \widehat{p}) = \text{maxroot}\,(U_\alpha p_0)$.

By Lemma 1.16

$$\text{maxroot}\,(U_\alpha \widehat{p} \boxplus_d q) = \text{maxroot}\,(U_\alpha \widehat{p} \boxplus_{d-1} Dq).$$

As the degree of $Dq$ is less than $d$, and $Dq$ has at least two distinct roots by Lemma 4.6 we may apply our inductive hypothesis to conclude that

$$\begin{aligned}
\text{maxroot}\,(U_\alpha \widehat{p} \boxplus_{d-1} Dq) &\leq \text{maxroot}\,(U_\alpha \widehat{p}) + \text{maxroot}\,(U_\alpha Dq) - (d-1)\alpha \\
&< \text{maxroot}\,(U_\alpha \widehat{p}) + \text{maxroot}\,(U_\alpha q) - d\alpha \quad \text{by Lemma 4.4} \\
&= \text{maxroot}\,(U_\alpha p_0) + \text{maxroot}\,(U_\alpha q) - d\alpha \quad \text{by property (c)} \\
&\leq \text{maxroot}\,(U_\alpha p_0 \boxplus_d q),
\end{aligned}$$

as $\phi(p_0) \leq 0$. Thus, property $(a)$ above and Lemma 4.2 imply that

$$\text{maxroot}\,(U_\alpha \widetilde{p} \boxplus_d q) > \text{maxroot}\,(U_\alpha p_0 \boxplus_d q).$$

As $\text{maxroot}\,(U_\alpha \widetilde{p}) = \text{maxroot}\,(U_\alpha p_0)$, this implies $\phi(\widetilde{p}) < \phi(p_0)$, a contradiction. Thus, (19) holds when both polynomials have at least two roots.

$\square$

## 4.2 Symmetric multiplicative convolution

In this section we prove the following upperbound on the variant of the $S-$transform of $p \boxtimes_d q$, defined in Sect. 1.3.

**Theorem 4.7** (Restatement of Theorem 1.13) *For $p, q \in \mathbb{P}^+(d)$ having only nonnegative real roots and $w > 0$,*

$$\widetilde{\mathcal{S}}_{p \boxtimes_d q}(w) \leq \widetilde{\mathcal{S}}_p(w) \widetilde{\mathcal{S}}_q(w),$$

*with equality only when $p$ or $q$ has only one distinct root.*

We begin by considering the case in which $p = (x - \lambda)^d$. We then have that

$$\widetilde{\mathcal{M}}_p(z) = \sum_{j \geq 1} (\lambda/z)^j = \frac{\lambda}{z - \lambda}.$$

Thus,

$$\widetilde{\mathcal{M}}_p^{(-1)}(w) = \frac{1 + w}{w} \lambda,$$

and

$$\widetilde{\mathcal{S}}_p(w) = \lambda.$$

**Lemma 4.8** *If $\lambda > 0$, $p(x) = (x - \lambda)^d$ and $q(x) \in \mathbb{P}^+(d)$, then for all $w \geq 0$*

$$\widetilde{\mathcal{S}}_{p \boxtimes_d q}(w) = \widetilde{\mathcal{S}}_p(w) \widetilde{\mathcal{S}}_q(w).$$

***Proof*** For $p(x) = (x - \lambda)^d$, one may use either the definition (1.4) or Theorem 1.5 to compute

$$p(x) \boxtimes_d q(x) = \lambda^d q(x/\lambda).$$

As, $\widetilde{\mathcal{M}}_{q(x/\lambda)}(\lambda z) = \widetilde{\mathcal{M}}_{q(x)}(z)$,

$$\widetilde{\mathcal{S}}_{p(x) \boxtimes_d q(x)}(w) = \lambda \widetilde{\mathcal{S}}_{q(x)}(w).$$

$\square$

The finite multiplicative convolution of polynomials of different degrees may be computed by taking the *polar derivative with respect to 0* of the polynomial of higher degree. We recall that the polar derivative at 0 of a polynomial $p$ of degree $d$ is given by $dp - x Dp$ (see [2, p. 44]).

**Lemma 4.9** (Degree Reduction for $\boxtimes_d$) *For $p(x) \in \mathbb{P}(d)$ and $q(x) \in \mathbb{P}(k)$, for $k < d$,*

$$p(x) \boxtimes_d q(x) = \frac{1}{d}(dp(x) - xDp(x)) \boxtimes_{d-1} q(x).$$

**Proof** Follows from equation (5) by an elementary computation. □

**Remark 4.10** If we let $R$ be the operation on polynomials in $\mathbb{P}^+(d)$ that maps $p(x)$ to $x^d p(1/x)$. The polar derivative of a degree $d$ polynomial may be expressed in terms of $R$ by

$$dp - xDp = RDRp.$$

In particular, the polar derivative has a discontinuity at $\infty$ that occurs as a root of $DRp$ passes 0, causing technical issues when considering polynomials $p(x)$ with both positive and negative roots (especially given that the root we are concerned in is the largest one). This prevents our proof method (which we chose to highlight the parallel between the additive and multiplicative case) from generalizing to a larger collection of polynomials. We therefore leave the possibility of a generalization as an open problem.

**Claim 4.11** *For $p = (x - \lambda)^d$,*

$$xDp - dp = \lambda d(x - \lambda)^{d-1}.$$

*For $p \in \mathbb{P}^+(d)$, $(xD - d)p \in \mathbb{P}^+(d-1)$ and*

$$\text{maxroot}(p) \geq \text{maxroot}(xDp - dp). \tag{20}$$

*In the special case of $p \in \mathbb{P}^+(2)$ with distinct roots, strict inequality holds:*

$$\text{maxroot}(p) > \text{maxroot}(xDp - 2p). \tag{21}$$

**Proof** The first part is a simple calculation. Inequality (20) follows from the fact that $p \in \mathbb{P}^+(d)$ implies that $Rp \in \mathbb{P}^+(d)$ and the fact that the roots of $DRp$ interlace those of $Rp$. To see that $(xD - d)p \in \mathbb{P}^+(d-1)$, observe that its lead coefficient is positive.

The last claim follows by noting that $Rp$ is quadratic polynomial with distinct roots and so $DRp$ strictly interlaces $Rp$. □

As we did with the symmetric additive convolution, we relate the $\widetilde{\mathcal{M}}$-transform to the maximum root of a polynomial. We have

$$\widetilde{\mathcal{M}}_p(z) = w \quad \Longleftrightarrow \quad \text{maxroot}\left(\left(1 - \frac{xD}{d(w+1)}\right)p(x)\right) = 0.$$

We therefore define the operator $V_w$ by

$$V_w p(x) = \left(1 - \frac{xD}{d(w+1)}\right) p(x),$$

which gives

$$\widetilde{\mathcal{M}}_p(z) = w \quad \Longleftrightarrow \quad \text{maxroot}(V_w p) = z.$$

Note that the polar derivative is $d V_0$.

Our proof of Theorem 4.7 will also employ the following consequence of Lemma 4.1.

**Corollary 4.12** *Let $w > 0$, $d \geq 2$, and let $p(x) \in \mathbb{P}^+(d)$ have at least two distinct roots. Then there exist $\widetilde{p} \in \mathbb{P}^+(d)$ and $\widehat{p} \in \mathbb{P}^+(d-1)$ so that $p(x) = \widehat{p}(x) + \widetilde{p}(x)$, the largest root of $\widetilde{p}$ is at most the largest root of $p$,*

$$\text{maxroot}(V_w p) = \text{maxroot}(V_w \widetilde{p}) = \text{maxroot}(V_w \widehat{p}),$$

*and if $d \geq 3$ then $\widehat{p}$ has two distinct roots.*

**Proof** To derive this from Lemma 4.1, let $t = \text{maxroot}(V_w p)$ and set

$$\alpha = \frac{t}{d(w+1)}.$$

The polynomials $\widetilde{p}$ and $\widehat{p}$ constructed in Lemma 4.1 now satisfy

$$\text{maxroot}(U_\alpha \widehat{p}) = \text{maxroot}(U_\alpha \widetilde{p}) = t = \text{maxroot}(V_w \widetilde{p}) = \text{maxroot}(V_w \widehat{p}),$$

as desired. □

**Proof of Theorem 4.7** We proceed by induction on $d$, the maximum degree of $p$ and $q$. The theorem is true for $d = 1$ by Lemma 4.8. As we have already shown that equality holds when one of $p$ or $q$ has just one root, we need to show that when both $p$ and $q$ have at least two distinct roots:

$$\text{maxroot}(V_w p \boxtimes_d q) < \frac{w}{1+w} \text{maxroot}(V_w p) \, \text{maxroot}(V_w q).$$

Fix $q \in \mathbb{P}^+(d)$ with at least two distinct roots, and for $p \in \mathbb{P}^+$ define:

$$\phi(p) \stackrel{\text{def}}{=} \frac{w}{1+w} \text{maxroot}(V_w p) \, \text{maxroot}(V_w q) - \text{maxroot}(V_w p \boxtimes_d q).$$

As before, we assume (for contradiction) that there exists a monic $p$ with two distinct roots and $\phi(p) \leq 0$. Choose an interval $[0, R]$ containing the roots of $p$, and let $p_0$ minimize $\phi$ over all degree $d$ monic polynomials with roots in this interval. Observe

that we can choose $p_0$ having two distinct roots: if $\phi(p_0) < 0$ this is implied by Lemma 4.8 and if $\phi(p_0) = 0$ we can take $p = p_0$. Thus we may apply Corollary 4.12 to obtain polynomials $\widetilde{p} \in \mathbb{P}^+(d)$ and $\widehat{p} \in \mathbb{P}^+(d-1)$ with $p_0 = \widetilde{p} + \widehat{p}$,

$$\mathsf{maxroot}\,(V_w\,p_0) = \mathsf{maxroot}\,(V_w\,\widetilde{p}) = \mathsf{maxroot}\,(V_w\,\widehat{p})\,, \qquad (22)$$

and $\mathsf{maxroot}\,(\widetilde{p}) \le \mathsf{maxroot}\,(p_0)$. By Lemma 4.2, we have

$$\mathsf{maxroot}\,(V_w(p_0 \boxtimes_d q)) \le \max\{\mathsf{maxroot}\,(V_w(\widehat{p} \boxtimes_d q))\,, \mathsf{maxroot}\,(V_w(\widetilde{p} \boxtimes_d q))\}\,,$$

with equality only if all three are equal. However, noting that $\widehat{p}$ and $(xD - d)q = -RDRq$ have two distinct roots whenever $d \ge 3$:

$$
\begin{aligned}
&\mathsf{maxroot}\,(V_w(\widehat{p} \boxtimes_d q))\\
&\quad = \mathsf{maxroot}\,(V_w\widehat{p} \boxtimes_{d-1}((xD - d)q)) \quad \text{by Lemma 4.9}\\
&\quad \le \frac{w}{1+w}\mathsf{maxroot}\,(V_w\widehat{p})\,\mathsf{maxroot}\,(V_w(xD - d)q)) \quad \text{by induction, strict for } d \ge 3\\
&\quad \le \frac{w}{1+w}\mathsf{maxroot}\,(V_w\widehat{p})\,\mathsf{maxroot}\,(V_wq)) \quad \text{by Claim 4.11, strict for } d = 2\\
&\quad = \frac{w}{1+w}\mathsf{maxroot}\,(V_w\,p_0)\,\mathsf{maxroot}\,(V_wq)) \quad \text{by (22)}\\
&\quad \le \mathsf{maxroot}\,(V_w(p_0 \boxtimes_d q))\,,
\end{aligned}
$$

since $\phi(p_0) \le 0$. Since at least one of the inequalities above is strict for all $d \ge 2$, we must have $\mathsf{maxroot}\,(V_w(p_0 \boxtimes_d q)) < \mathsf{maxroot}\,(V_w(\widetilde{p} \boxtimes_d q))$, which implies $\phi(\widetilde{p}) < \phi(p_0)$, contradicting the minimality of $p_0$. $\qquad\square$

### 4.3 Asymmetric additive convolution

In this section we prove the rectangular analogue of Theorem 1.12.

**Theorem 4.13** (Restatement of Theorem1.14) *Let $p(x)$ and $q(x)$ be in $\mathbb{P}^+(d)$ for $d \ge 1$. Then for all $\alpha \ge 0$,*

$$\mathsf{maxroot}\,(U_\alpha\mathbb{S}(p \boxplus_d q)) \le \mathsf{maxroot}\,(U_\alpha\mathbb{S}p) + \mathsf{maxroot}\,(U_\alpha\mathbb{S}q) - 2\alpha d,$$

*with equality if and only if $p$ or $q$ equals $x^d$.*

We remark that if $q(x) = x^d$, then $p \boxplus_d q = p$, and

$$U_\alpha\mathbb{S}q = U_\alpha x^{2d} = x^{2d-1}(x - 2d\alpha),$$

so

$$\mathsf{maxroot}\,(U_\alpha\mathbb{S}q) = 2\alpha d.$$

This is why Theorem 4.13 holds with equality when $q(x) = x^d$.

The following lemma tells us that it suffices to prove Theorem 4.13 in the case that $\alpha = 1$.

**Lemma 4.14** *For a real-rooted polynomial $p(x)$,*

$$maxroot\,(U_\alpha p(x)) = \frac{1}{\alpha} maxroot\,(U_1 p(\alpha x)).$$

**Proof** Let $q(x) = p(\alpha x)$, so

$$U_1 q(x) = p(\alpha x) - \alpha p'(\alpha x).$$

Let

$$w = \alpha\text{–max}\,(p(x)) = maxroot\,(U_\alpha p) \quad\Longleftrightarrow\quad p(w) - \alpha p'(w) = 0.$$

Then,

$$(U_1 q)(w/\alpha) = p(w) - \alpha p'(w) = 0.$$

$\square$

Our proof of Theorem 4.13 will use the following lemma to pinch together roots of $p$ to reduce the analysis to a few special cases.

**Corollary 4.15** *Let $\alpha > 0$, $d \geq 2$, and let $p(x) \in \mathbb{P}^+(d)$ have at least two distinct roots. Then there exist $\widetilde{p} \in \mathbb{P}^+(d)$ and $\widehat{p} \in \mathbb{P}^+(d-1)$ so that $p(x) = \widehat{p}(x) + \widetilde{p}(x)$, the largest root of $\widetilde{p}$ is at most the largest root of $p$, $\widehat{p}$ has a root larger than 0, and*

$$maxroot\,(U_\alpha \mathbb{S}\widetilde{p}) = maxroot\,(U_\alpha \mathbb{S}\widehat{p}) = maxroot\,(U_\alpha \mathbb{S}p) \qquad (23)$$

**Proof** Let $t = maxroot\,(U_\alpha \mathbb{S}p)$, so

$$maxroot\,((1 - 2\alpha t D)p) = \sqrt{t}.$$

Apply Lemma 4.1 with $2\alpha t$ in the place of $\alpha$ to construct the polynomials $\widetilde{p}$ and $\widehat{p}$. They satisfy

$$maxroot\,(U_\alpha \widehat{p}) = maxroot\,(U_\alpha \widetilde{p}) = \sqrt{t},$$

which implies (23). $\square$

We will build up to the proof of Theorem 4.13 by first handling three special cases:

- When $p(x) = (x - \lambda)^d$ and $q(x) = x^{d-1}$. That is, we consider $DxD(x - \lambda)^d$ (Lemma 4.18).

- When $p(x) \in \mathbb{P}^+(d)$ and $q(x) = x^{d-1}$. That is, we consider $DxDp(x)$ (Lemma 4.19).
- When $p(x) = (x - \lambda)^d$ and $q(x) = (x - \mu)^d$ (Lemma 4.20).

As with the other convolutions, we may compute the asymetric additive convolution of two polynomials by first applying an operation to the polynomial of higher degree. In this case it is $DxD$, also known as the "Laguerre Derivitive".

**Lemma 4.16** (Degree Reduction for $⊞_d$) *Let $p \in \mathbb{P}^+(d)$ and let $q \in \mathbb{P}^+(k)$ for $k < d$. Then,*

$$p ⊞_d q = (1/d^2)(DxDp) ⊞_{d-1} q.$$

**Proof** Follows from Theorem 1.8. □

The following characterization of the Laguerre derivitive also follows from Theorem 1.8.

**Claim 4.17** *If $q(x) = x^{d-1}$, then*

$$p ⊞_d q = DxDp.$$

**Lemma 4.18** *For $\alpha \geq 0$, $\lambda \geq 0$, $d \geq 2$ and $p(x) = (x - \lambda)^d$,*

$$maxroot\,(U_\alpha \mathbb{S} Dx Dp) \leq maxroot\,(U_\alpha \mathbb{S} p) - 2\alpha,$$

*with equality only if $\lambda = 0$.*

**Proof** By Lemma 4.14, it suffices to consider the case of $\alpha = 1$. As $\mathbb{S} p(x) = (x^2 - \lambda)^d$,

$$U_1 \mathbb{S} p(x) = (x^2 - 2\lambda d - \lambda)(x^2 - \lambda)^{d-1}.$$

So, the largest root of this polynomial is the largest root of

$$r_\lambda(x) \overset{\text{def}}{=} (x^2 - 2\lambda d - \lambda).$$

We may also compute

$$U_1 \mathbb{S} Dx D(x^2 - \lambda)^d = q_\lambda(x)(x^2 - \lambda)^{d-2},$$

where

$$q_\lambda(x) \overset{\text{def}}{=} dx^4 - 2d(d-1)x^3 - (d+1)\lambda x^2 + 4(d-1)\lambda x + \lambda^2.$$

We now prove that

$$maxroot\,(q_\lambda) \leq maxroot\,(r_\lambda) - 2,$$

with equality only if $\lambda = 0$.

We first argue that $q_\lambda(x)$ is real rooted. This follows from that fact that it is a factor of $U_1 \mathbb{S} Dx D(x - \lambda)^d$. For $\lambda \geq 0$ all of the roots of $Dx D(x - \lambda)^d$ are nonnegative, and so applying $\mathbb{S}$ to it yields a polynomial with all real roots.

We now compute

$$\mathsf{maxroot}\,(r_\lambda) = d + \sqrt{d^2 + \lambda}.$$

Define

$$\mu_\lambda = d + \sqrt{d^2 + \lambda} - 2.$$

Elementary algebra gives

$$q_\lambda(\mu_\lambda) = 4\lambda - 8d\sqrt{d^2 + \lambda} + 8d^2 = (2d - 2\sqrt{d^2 + \lambda})^2.$$

So, $q_\lambda(\mu_\lambda) \geq 0$, with equality only when $\lambda = 0$. With just a little more work, we will show that $\mu_\lambda$ is an upper bound on the roots of $q_\lambda$ for all $\lambda$.

For $q_\lambda$ to have a root larger than $\mu_\lambda$, it would have to have two roots larger than $\mu_\lambda$. When $\lambda = 0$, the polynomial $q_\lambda$ has one root at $\mu_0$ and a root at 0 with multiplicity 3. As $q_\lambda$ is real rooted for all $\lambda \geq 0$ and the roots of $q_\lambda$ are continuous functions of its coefficients, and thus of $\lambda$, we can conclude that for small $\lambda$ all but one of the roots of $q_\lambda$ must be near 0. Thus, for sufficiently small $\lambda$, $q_\lambda$ can have at most one root greater than $\mu_\lambda$, and so it must have none. As the largest root of $q_\lambda$ and $\mu_\lambda$ are continuous function of $\lambda$, $\mathsf{maxroot}\,(q_\lambda) > \mu_\lambda$ for all sufficiently small $\lambda$. As $q_\lambda(\mu_\lambda) > 0$ for all $\lambda \geq 0$, we can conclude that $\mathsf{maxroot}\,(q_\lambda) > \mu_\lambda$ for all $\lambda \geq 0$. $\qquad\square$

To see that Lemma 4.18 is equivalent Theorem 4.13 in the case of $q = x^{d-1}$, note that for $q(x) = x^{d-1}$,

$$U_\alpha \mathbb{S} q(x) = U_\alpha q(x^2) = x^{2(d-1)} - \alpha Dx^{2(d-1)} = x^{2d-3}(x - 2(d-1)\alpha).$$

The equivalence now follows from Claim 4.17 and the fact that the the largest root of this polynomial is $2(d-1)\alpha$.

**Lemma 4.19** *For $\alpha \geq 0$, $d \geq 2$ and $p \in \mathbb{P}^+(d)$,*

$$\mathsf{maxroot}\,(U_\alpha \mathbb{S} Dx Dp) \leq \mathsf{maxroot}\,(U_\alpha \mathbb{S} p) - 2\alpha,$$

*with equality only if $p(x) = x^d$.*

**Proof** For every $p \in \mathbb{P}^+$, define

$$\phi(p) = \mathsf{maxroot}\,(U_\alpha \mathbb{S} p) - \mathsf{maxroot}\,(U_\alpha \mathbb{S} Dx Dp) - 2\alpha.$$

We will show that $\phi(p) \geq 0$ for every polynomial $p \in \mathbb{P}^+$ of degree at least 2, with equality only when $p = x^d$.

Our proof will be by induction on the degree of $p$. Assume that there exists a polynomial $p \in \mathbb{P}^+(d)$ with $\phi(p) < 0$, let $[0, R]$ be an interval containing the roots of $p$, and let $p_0$ minimize $\phi$ over polynomials with roots in that interval. By Lemma 4.18, $p_0$ must have at least 2 distinct roots, and so we can apply Corollary 4.15 to obtain polynomials $\widehat{p}$ and $\widetilde{p}$.

Let $x = \mathsf{maxroot}\,(U_\alpha \mathbb{S} Dx\,Dp)$. If $d = 2$, then $\widehat{p}$ has degree 1 and so $U_\alpha \mathbb{S} Dx\,D\widehat{p}$ equals the lead coefficient of $\widehat{p}$, which implies

$$U_\alpha \mathbb{S} Dx\,D\widehat{p}(x) > 0. \tag{24}$$

For $d \geq 3$, we can then assume by induction that $\phi(\widehat{p}) > 0$, which then implies (24) as well. Combining this with Lemma 4.2, we get that $\phi(p_0) > \phi(\widetilde{p})$ which contradicts the minimality of $p_0$.

$\square$

In Sect. 4.4, we establish the following special case of Theorem 4.13.

**Lemma 4.20** *For $\lambda, \mu > 0$, and $d \geq 1$, let $p(x) = (x - \lambda)^d$ and $q(x) = (x - \mu)^d$. Then for all $\alpha \geq 0$,*

$$\mathsf{maxroot}\,(U_\alpha \mathbb{S}(p \boxplus_d q)) < \mathsf{maxroot}\,(U_\alpha \mathbb{S}p) + \mathsf{maxroot}\,(U_\alpha \mathbb{S}q) - 2\alpha d.$$

We now use Lemma 4.20 to prove Theorem 4.13 through a variation of the pinching argument employed in the proof of Lemma 4.18.

***Proof of Theorem 4.13*** We will prove this by induction on the maximum degree of $p$ and $q$, which we call $d$. Our base case of $d = 1$ is handled by Lemma 4.20.

Assume, without loss of generality, that the degree of $p$ is at least the degree of $q$. If the degree of $p$ is larger than the degree of $q$, then we may prove the hypothesis by

$$\begin{aligned}
\mathsf{maxroot}\,(U_\alpha \mathbb{S}(p \boxplus_d q)) &= \mathsf{maxroot}\,(U_\alpha \mathbb{S}((Dx\,Dp) \boxplus_{d-1} q)) \quad \text{(by Lemma 4.16)} \\
&\leq \mathsf{maxroot}\,(U_\alpha \mathbb{S}(Dx\,Dp)) + \mathsf{maxroot}\,(U_\alpha \mathbb{S}q) - 2\alpha(d - 1) \quad \text{(by induction)} \\
&\leq \mathsf{maxroot}\,(U_\alpha \mathbb{S}p) + \mathsf{maxroot}\,(U_\alpha \mathbb{S}q) - 2\alpha d \quad \text{(by Lemma 4.18).}
\end{aligned}$$

Lemma 4.18 also tells us that equality is only achieved when $p = x^d$.

We now consider the case in which both $p$ and $q$ have degree $d$. For polynomials $p$ and $q$ in $\mathbb{P}^+(d)$, define

$$\phi(p, q) = \mathsf{maxroot}\,(U_\alpha \mathbb{S}p) + \mathsf{maxroot}\,(U_\alpha \mathbb{S}q) - 2\alpha d - \mathsf{maxroot}\,(U_\alpha \mathbb{S}(p \boxplus_d q)).$$

We will prove that $\phi(p, q) \geq 0$ for all such polynomials.

Assume (for contradiction) that there exist polynomials $p, q$ with $\phi(p, q) < 0$ and let $[0, R]$ be an interval containing the roots of $p$ and $q$. Again, $\phi$ is a continuous function (this time on the compact set $[0, R]^{2d}$) so let $p_0, q_0$ be a minimizer. If *both* $p_0$ and $q_0$ have at most 1 distinct root, then Lemma 4.20 implies $\phi(p_0, q_0) \geq 0$, with equality only if one of them equals $x^d$ (a contradiction). Hence we can assume without

loss of generality that $p_0$ has least 2 distinct roots and so Corollary 4.12 provides polynomials $\widehat{p}$ and $\widetilde{p}$ with

$$\text{maxroot}\,(U_\alpha \mathbb{S} p_0) = \text{maxroot}\,(U_\alpha \mathbb{S} \widetilde{p}) = \text{maxroot}\,(U_\alpha \mathbb{S} \widehat{p})$$

and by Lemma 4.2

$$\begin{aligned}
\text{maxroot}\,&(U_\alpha \mathbb{S}(p_0 \boxplus_d q_0)) \\
&\leq \max\{\text{maxroot}\,(U_\alpha \mathbb{S}(\widehat{p} \boxplus_d q_0))\,, \text{maxroot}\,(U_\alpha \mathbb{S}(\widetilde{p} \boxplus_d q_0))\}
\end{aligned}$$

which in turn implies $\phi(p_0, q_0) \geq \min \phi(\widehat{p}, q_0), \phi(\widetilde{p}, q_0)$ with equality if and only if all are equal. Again equality cannot occur, since $\phi(p_0, q_0) < 0$ by assumption and $\phi(\widehat{p}, q_0) \geq 0$ by the inductive hypothesis and $\phi(p_0, q_0) > \phi(\widehat{p}, q_0)$ cannot occur for the same reason. But this implies $\phi(p_0, q_0) > \phi(\widetilde{p}, q_0)$, which contradicts the minimality of the pair $(p_0, q_0)$.

$\square$

## 4.4 Ultraspherical polynomials

This section is devoted to the proof of Lemma 4.20. It is a consequence of the following lemma.

**Lemma 4.21** *For $d \geq 0$ and positive $\lambda$ and $\mu$,*

$$\begin{aligned}
\text{maxroot}\,&\left(U_\alpha \mathbb{S}((x - \lambda)^d \boxplus_d (x - \mu)^d)\right) \\
&< \text{maxroot}\,\left(U_\alpha (\mathbb{S}(x - \lambda)^d \boxplus_{2d} \mathbb{S}(x - \mu)^d))\right).
\end{aligned}$$

Lemma 4.20 then follows from Theorem 4.3. We will prove Lemma 4.21 by showing that the polynomial on the left is a scaled Chebyshev polynomial of the second kind, and that the polynomial on the right is a Legendre polynomial with the same scaling. We then appeal to known relations between the roots of these polynomials.

The Cheybshev and Legendre polynomials are both Ultraspherical (also called Gegenbauer) polynomials. These are special cases of Jacobi polynomials. It is known that their roots all lie between $-1$ and $1$ and that they are symmetric about zero (see Theorem 3.3.1 of [28]).

The Ultraspherical polynomials with parameter $\alpha$ are defined by the generating function

$$\sum_n C_n^{(\alpha)}(x) t^n = \frac{1}{(1 - 2xt + t^2)^\alpha}. \tag{25}$$

Two special instances of these polynomials are

1. the Legendre polynomials: $P_d(x) = C_d^{(1/2)}(x)$, and
2. the Chebyshev polynomials of the second kind: $U_d(x) = C_d^{(1)}(x)$.

Stieltjes [29] (see Theorem 6.21.1 of [28]) established the following relation between the zeros of the Chebyshev and Legendre polynomials.

**Theorem 4.22** *Let $\alpha_1 > \alpha_2 > \cdots > \alpha_{\lfloor d/2 \rfloor}$ be the positive roots of $U_d(x)$ and let $\beta_1 > \beta_2 > \cdots > \beta_{\lfloor d/2 \rfloor}$ be the positive roots of $P_d(x)$. Then, $\alpha_i < \beta_i$ for all $i$.*

The relationship between the asymmetric additive convolution and the Chebyshev polynomials will make use of the generating function (25). To aid in this endeavor, we recall the following well-known generalization of the *binomial theorem* (see, for example, [30]).

**Theorem 4.23** *The function $(1 + z)^{-k}$ has the formal power series expansion*

$$\frac{1}{(1 + z)^k} = \sum_{i=0}^{\infty} \binom{k + i - 1}{i} (-z)^i.$$

**Lemma 4.24** *For $d \geq 0$ and positive $\lambda$ and $\mu$,*

$$\mathbb{S}((x - \lambda)^d \boxplus_d (x - \mu)^d) = (\lambda\mu)^{d/2} U_d \left( \frac{x - (\lambda + \mu)}{2\sqrt{\lambda\mu}} \right).$$

***Proof*** By (25), we have

$$\sum_d U_d(x) \, t^d = \frac{1}{1 - 2xt + t^2}$$

and so

$$\sum_d \sqrt{\lambda\mu}^d U_d \left( \frac{x - \lambda - \mu}{2\sqrt{\lambda\mu}} \right) t^d = \frac{1}{1 - (x - \lambda - \mu)t + \lambda\mu t^2}$$

$$= \frac{1}{(1 - \lambda t)(1 - \mu t) - xt}$$

$$= \frac{1}{(1 - \lambda t)(1 - \mu t)} \frac{1}{1 - \frac{xt}{(1 - \lambda t)(1 - \mu t)}}$$

$$= \sum_{k \geq 0} \frac{x^k t^k}{(1 - \lambda t)^{k+1}(1 - \mu t)^{k+1}}$$

$$= \sum_{i,j,k \geq 0} \binom{k + i}{i} \binom{k + j}{j} x^k (-\lambda)^i (-\mu)^j t^{i+j+k}$$

so the coefficient of $t^d$ can be written (setting $\ell = d - k$)

$$\sum_{\ell=0}^{d} x^{d-\ell} \sum_{i+j=\ell} \binom{d - j}{i} \binom{d - i}{j} (-\lambda)^i (-\mu)^j.$$

On the other hand, the formula for the asymmetric convolution also gives us

$$(x - \lambda)^d \boxplus_d (x - \mu)^d = \sum_{k=0}^{d} x^{d-k} \sum_{i+j=k} \left( \frac{(d-i)!(d-j)!}{d!(d-i-j)!} \right)^2 \binom{d}{i}\binom{d}{j}(-\mu)^i(-\lambda)^j$$

$$= \sum_{k=0}^{d} x^{d-k} \sum_{i+j=k} \frac{(d-i)!(d-j)!}{i!j!(d-i-j)!(d-i-j)!}(-\mu)^i(-\lambda)^j$$

$$= \sum_{k=0}^{d} x^{d-k} \sum_{i+j=k} \binom{d-i}{j}\binom{d-j}{i} x^i (-\lambda)^i (-\mu)^j.$$

□

The relationship between the symmetric additive convolution and the Legendre polynomials can be established by applying Theorem 2.11

**Lemma 4.25** *For $d \geq 0$ and positive $\lambda$ and $\mu$,*

$$(\mathbb{S}(x - \lambda)^d) \boxplus_{2d} (\mathbb{S}(x - \mu)^d) = \frac{(4\sqrt{\lambda\mu})^d}{\binom{2d}{d}} P_d \left( \frac{x^2 - \lambda - \mu}{2\sqrt{\lambda\mu}} \right)$$

**Proof** We start by recalling one of the well-known formulas for Legendre polynomials [28]:

$$P_d(x) = \frac{1}{2^d} \sum_i \binom{d}{i}^2 (x-1)^i (x+1)^{d-i}.$$

On the one hand we have

$$P_d \left( \frac{x^2 - \lambda - \mu}{2\sqrt{\lambda\mu}} \right) = \frac{1}{2^d} \sum_i \binom{d}{i}^2 \left( \frac{x^2 - \lambda - \mu}{2\sqrt{\lambda\mu}} - 1 \right)^i \left( \frac{x^2 - \lambda - \mu}{2\sqrt{\lambda\mu}} + 1 \right)^{d-i}$$

$$= \frac{1}{(4\sqrt{\lambda\mu})^d} \sum_i \binom{d}{i}^2 (x^2 - (\sqrt{\lambda} + \sqrt{\mu})^2)^i (x^2 - (\sqrt{\lambda} - \sqrt{\mu})^2)^{d-i}.$$

On the other hand, we can use Theorem 2.11 to calculate $(x^2 - \lambda)^d \boxplus_{2d} (x^2 - \mu)^d$. There are four possible root sums that will appear: $\{\pm(\sqrt{\lambda} + \sqrt{\mu}), \pm(\sqrt{\lambda} - \sqrt{\mu})\}$. Furthermore, it is easy to check that

1. the $\pm(\sqrt{\lambda} + \sqrt{\mu})$ terms appear the same number of times in every pairing
2. the $\pm(\sqrt{\lambda} - \sqrt{\mu})$ terms appear the same number of times in every pairing
3. the probability of having $i$ copies of $(\sqrt{\lambda} + \sqrt{\mu})$ and $d - i$ copies of $(\sqrt{\lambda} - \sqrt{\mu})$ is $\binom{d}{i}^2 / \binom{2d}{d}$

Hence we have

$$[(x^2 - \lambda)^d \boxplus_{2d} (x^2 - \mu)^d] = \frac{1}{\binom{2d}{d}} \sum_{i=0}^{d} \binom{d}{i}^2$$
$$(x^2 - (\sqrt{\lambda} + \sqrt{\mu})^2)^i (x^2 - (\sqrt{\lambda} - \sqrt{\mu})^2)^{d-i}.$$

$\square$

***Proof of Lemma 4.21*** We prove this in the case that $d$ is even. The proof with $d$ odd is similar, except that there is one extra common term for the root at 0. Let $1 > \alpha_1 > \alpha_2 > \cdots > \alpha_{d/2}$ be the positive roots of $U_d(x)$ and let $\alpha_{d+1-i} = -\alpha_i$ be the negative roots. Similarly, let $1 > \beta_1 > \beta_2 > \cdots > \beta_d$ be the roots of $P_d(x)$. As $(\lambda + \mu)/2\sqrt{\lambda\mu} \geq 1$, the roots of

$$\mathbb{S}((x - \lambda)^d \boxplus_d (x - \mu)^d)$$

are

$$\pm \left(\sqrt{\lambda\mu}\beta_i + (\lambda + \mu)\right)^{1/2}$$

and the roots of

$$(\mathbb{S}(x - \lambda)^d) \boxplus_{2d} (\mathbb{S}(x - \mu)^d)$$

are

$$\pm \left(\sqrt{\lambda\mu}\alpha_i + (\lambda + \mu)\right)^{1/2}.$$

By Theorem 4.22 the largest root of the first polynomial is larger than the largest root of the second, and for every $t$ larger than that root,

$$\mathcal{G}_t \left(\mathbb{S}((x - \lambda)^d \boxplus_d (x - \mu)^d)\right) = \frac{1}{2d} \sum_{i=1}^{d} \frac{1}{t - \left(\sqrt{\lambda\mu}\alpha_i + (\lambda + \mu)\right)^{1/2}}$$
$$+ \frac{1}{t + \left(\sqrt{\lambda\mu}\alpha_i + (\lambda + \mu)\right)^{1/2}}$$
$$= \frac{1}{2d} \sum_{i=1}^{d} \frac{2t}{t^2 - \sqrt{\lambda\mu}\alpha_i - (\lambda + \mu)}$$
$$= \frac{1}{2d} \sum_{i=1}^{d/2} \frac{2t}{t^2 - (\lambda + \mu) - \sqrt{\lambda\mu}\alpha_i}$$
$$+ \frac{2t}{t^2 - (\lambda + \mu) + \sqrt{\lambda\mu}\alpha_i}$$

$$= \frac{1}{2d} \sum_{i=1}^{d/2} \frac{4t(t^2 - (\lambda + \mu))}{(t^2 - (\lambda + \mu))^2 - \lambda\mu\alpha_i^2}$$

$$< \frac{1}{2d} \sum_{i=1}^{d/2} \frac{4t(t^2 - (\lambda + \mu))}{(t^2 - (\lambda + \mu))^2 - \lambda\mu\beta_i^2}$$

$$= \mathcal{G}_t \left( (\mathbb{S}(x - \lambda)^d) \boxplus_{2d} (\mathbb{S}(x - \mu)^d) \right),$$

where the last equality follows by reversing the previous reasoning. This implies that

$$\mathcal{K}_{\mathbb{S}((x-\lambda)^d \boxplus_d (x-\mu)^d)}(w) < \mathcal{K}_{(\mathbb{S}(x-\lambda)^d)\boxplus_{2d}(\mathbb{S}(x-\mu)^d)}(w),$$

where $w = 1/\alpha d$. □

## Declaration

## References

1. Walsh, J.: On the location of the roots of certain types of polynomials. Trans. Am. Math. Soc. **24**(3), 163–180 (1922)
2. Marden, M.: Geometry of Polynomials, vol. 3. American Mathematical Society, Providence, Rhode Island (1966)
3. Rahman, Q.I., Schmeisser, G.: Analytic Theory of Polynomials. Clarendon Press, Oxford (2002)
4. Mirabelli, B.: Hermitian, Non-Hermitian and Multivariate Finite Free Probability. PhD thesis, Princeton University (2020)
5. Szegö, G.: Bemerkungen zu einem Satz von J H Grace über die Wurzeln algebraischer Gleichungen. Math. Z. **13**(1), 28–55 (1922)
6. Marcus, A., Spielman, D.A., Srivastava, N.: Finite free convolutions of polynomials. arXiv preprint arXiv:1504.00350v1 (2015)
7. Voiculescu, D.V.: Free Probability Theory. American Mathematical Society, Providence, Rhode Island (1997)
8. Nica, A., Speicher, R.: Lectures on the Combinatorics of Free Probability, vol. 13. Cambridge University Press, Cambridge (2006)
9. Anderson, G.W., Guionnet, A., Zeitouni, O.: An Introduction to Random Matrices, vol. 118. Cambridge University Press, Cambridge (2010)
10. Marcus, A.W.: Polynomial convolutions and (finite) free probability. arXiv preprint arXiv:2108.07054 (2015)

11. Arizmendi, O., Perales, D.: Cumulants for finite free convolution. J. Combin. Theory Ser. A **155**, 244–266 (2018)
12. Gorin, V., Marcus, A.W.: Crystallization of random matrix orbits. Int. Math. Res. Not. (2018). https://doi.org/10.1093/imrn/rny052
13. Marcus, A.W., Spielman, D.A., Srivastava, N.: Interlacing families I: Bipartite Ramanujan graphs of all degrees. Ann. Math. **182**(1), 307–325 (2015)
14. Marcus, A.W., Spielman, D.A., Srivastava, N.: Interlacing families II: mixed characteristic polynomials and the Kadison–Singer problem. Ann. Math. **182**(1), 327–350 (2015)
15. Marcus, A.W., Spielman, D.A., Srivastava, N.: Interlacing families IV: Bipartite Ramanujan graphs of all sizes. In: 2015 IEEE 56th Annual Symposium on Foundations of Computer Science (FOCS), pp. 1358–1377 ( 2015). IEEE
16. Cohen, M.B.: Ramanujan graphs in polynomial time. In: 2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS), pp. 276–281 ( 2016). IEEE
17. Hall, C., Puder, D., Sawin, W.F.: Ramanujan coverings of graphs. Adv. Math. **323**, 367–410 (2018)
18. Leake, J., Ryder, N.: On the further structure of the finite free convolutions. arXiv preprint arXiv:1811.06382 (2018)
19. Dumitriu, I., Edelman, A.: Matrix models for beta ensembles. J. Math. Phys. **43**(11), 5830–5847 (2002)
20. Edelman, A.: Eigenvalues and condition numbers of random matrices. SIAM J. Matrix Anal. Appl. **9**(4), 543–560 (1988)
21. Benaych-Georges, F.: Rectangular random matrices, related convolution. Probab. Theory Relat. Fields **144**(3–4), 471–515 (2009)
22. Horn, R.A., Johnson, C.R.: Matrix Analysis. Cambridge University Press, Cambridge (2012)
23. Marcus, M.: Determinants of sums. Coll. Math. J. **21**(2), 130–135 (1990)
24. Wagner, D.: Multivariate stable polynomials: theory and applications. Bull. Am. Math. Soc. **48**(1), 53–84 (2011)
25. Brändén, P.: Polynomials with the half-plane property and matroid theory. Adv. Math. **216**(1), 302–320 (2007)
26. Lieb, E.H., Sokal, A.D.: A general Lee–Yang theorem for one-component and multicomponent ferromagnets. Commun. Math. Phys. **80**(2), 153–179 (1981). https://doi.org/10.1007/BF01213009
27. Borcea, J., Brändén, P.: The Lee–Yang and Pólya-Schur programs. II. Theory of stable polynomials and applications. Commun. Pure Appl. Math. **62**( 12), 1595–1631 ( 2009)
28. Szegö, G.: Orthogonal Polynomials, vol. 23. American Mathematical Society, Providence, Rhode Island (1939)
29. Stieltjes, T.J.: Sur les racines de l'équation $x_n = 0$. Acta Math. **9**(1), 385–400 (1887)
30. Wilf, H.S.: Generatingfunctionology. AK Peters/CRC Press, Boca Raton, Florida (2005)