

A RECTANGULAR ADDITIVE CONVOLUTION FOR POLYNOMIALS

Aurelien Gribinski¹ and Adam W. Marcus*²

¹*Princeton University, Princeton, New Jersey, U.S.A.*

aurelien.gribinski@epfl.ch

²*École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland*

adam.marcus@epfl.ch

Submitted: Apr 1, 2021; Accepted: Feb 15, 2022; Published: Mar 31, 2022

© The authors. Released under the CC BY license (International 4.0).

Abstract. Motivated by the study of singular values of random rectangular matrices, we define and study the rectangular additive convolution of polynomials with nonnegative real roots. Our definition directly generalizes the asymmetric additive convolution introduced by Marcus, Spielman and Srivastava (2015), and our main theorem gives the corresponding generalization of the bound on the largest root from that paper. The main tool used in the analysis is a differential operator derived from the “rectangular Cauchy transform” introduced by Benaych-Georges (2009). The proof is inductive, with the base case requiring a new nonasymptotic bound on the Cauchy transform of Gegenbauer polynomials which may be of independent interest.

Keywords. Polynomial convolutions, finite free probability, Gegenbauer polynomials

Mathematics Subject Classifications. 26C10, 33C45

1. Introduction

This paper introduces the rectangular additive convolution to the theory of finite free probability. The motivation for a finite analogue of free probability came from a series of works that used expected characteristic polynomials to study certain combinatorial problems in linear algebra [14, 16, 17]. It is well known that the characteristic polynomial of a Hermitian matrix A is the real-rooted polynomial $\chi_A(x) = \det [x\mathbb{I} - A]$, and in [16] the authors showed that one could analyze the effect of certain random perturbations on A by studying the effect of certain differential operators applied to $\chi_A(x)$. In particular, they showed that any valid bounds on the largest root of the transformed polynomials could be translated into statements about the randomly perturbed

*Research done under the support of NSF CAREER grant DMS-1552520 and a Von Neumann Fellowship at the Institute of Advanced Study, NSF grant DMS-1128155.

matrix. In [14], the same authors showed that these differential operators were actually a special case of a more general class of “polynomial convolutions” and they introduced a technique for deriving bounds on the largest root under certain convolutions. One interesting property of the bounds that came from this technique was that they had quite similar form to known identities from free probability. This idea was strengthened by the realization that the major tools used in proving these bounds had striking similarities to tools in free probability.

The connection was formalized in [13], where it was shown that the inequalities derived for two of the convolutions studied in [14] — the symmetric additive and multiplicative convolutions — converge to the R - and S -transform identities of free probability (respectively). Since the release of [13], a number of advances have been made in understanding the relationship between free probability and polynomial convolutions, most notably the work of Arizmendi and Perales [1] in developing a combinatorial framework for finite free probability using finite free cumulants (the approach in [13] is primarily analytic).

The purpose of this paper is to introduce a convolution on polynomials that generalizes the third convolution studied in [14] — what is there called the *asymmetric additive convolution* — and to prove the corresponding bound on the largest root. The method of proof will be similar to the one used in [14], however there will be a number of added complications. Those familiar with [14] may recall that all of the inequalities proved there utilized various levels of induction to reduce to a small set of “base cases.” One of the difficulties in dealing with the asymmetric additive convolution (as opposed to the other two convolutions) was the fact that the corresponding base case was highly nontrivial. Rather, it required a bound on the Cauchy transform of Chebyshev polynomials that was both unknown at the time and not particularly easy to derive.

We will encounter the same issue: replacing the analysis on Chebyshev polynomials will be an analysis on (the more general) Gegenbauer polynomials. To establish the bound, we will prove a number of inequalities relating nonasymptotic properties of Gegenbauer polynomials (which appear to be unknown) with the corresponding asymptotic properties (many of which are known) and these could be of interest in their own right (see Section 1.2 for the location of results).

Many of the ideas required in generalizing the various constructs in [14] to the ones used in this paper were inspired by the work of Florent Benaych-Georges, in particular [3] where the appropriate transforms for calculating the rectangular additive convolution of two freely independent rectangular operators were introduced (hence the name of our convolution). We remark on this connection briefly in Section 1.3, but in general have written the paper in a way that assumes no previous knowledge of free probability.

1.1. Previous work

The primary predecessors of this work are [14], where other polynomial convolutions were introduced, and [3], where the free probability version of the rectangular additive convolution was introduced (see 1.3 for more discussion on the relation to [3]). The original purpose of [14] was to develop a generic way to bound the largest root of a real-rooted polynomial when certain differential operators were applied. Such bounds are useful in tandem with the “method

of interlacing polynomials” first introduced in [15]. One of the main inequalities in [14] is the following:

Theorem 1.1 (Marcus, Spielman, Srivastava). *Let p and q be real rooted polynomials of degree at most d . Then*

$$\maxroot(U_\alpha[p \boxplus_d q]) \leq \maxroot(U_\alpha p) + \maxroot(U_\alpha q) - \alpha d$$

The operator \boxplus_d here is what is called the *symmetric additive convolution* in [14] and U_α is the differential operator $1 - \alpha\partial$. The U_α acts to smoothen the roots of the polynomials, making the convolution more predictable. Theorem 1.1 is used in [16] to prove an asymptotically tight version of *restricted invertibility*, a theorem first introduced by Bourgain and Tzafriri that has seen a wide variety of uses in mathematics (see [19]).

A considerably more difficult inequality in [14] concerned what the authors called the *asymmetric additive convolution*. In the notation of this paper — see (2.1) and (3.1) — this inequality reads

Theorem 1.2 (Marcus, Spielman, Srivastava). *Let p and q be polynomials of degree at most d with nonnegative real roots. Then*

$$\maxroot(U_\alpha \mathbb{S}[p \boxplus_d^0 q]) \leq \maxroot(U_\alpha \mathbb{S}p) + \maxroot(U_\alpha \mathbb{S}q) - 2\alpha d.$$

for all $\alpha \leq 0$. Furthermore, equality holds if and only if $p(x) = x^d$ or $q(x) = x^d$.

Theorem 1.2 was then used in [17] to prove the existence of bipartite Ramanujan graphs of all degrees and all sizes. The main theorem of this paper (Theorem 1.3) is a generalization of Theorem 1.2. In particular, we show that the generalized convolution \boxplus_d^n (defined in Section 2) satisfies a similar inequality:

Theorem 1.3. *Let p and q be polynomials of degree at most d with nonnegative real roots. Then*

$$\Theta_\alpha^n(p \boxplus_d^n q) \leq \Theta_\alpha^n(p) + \Theta_\alpha^n(q) - (n + 2d)\alpha \tag{1.1}$$

for all $\alpha > 0$ and $n \geq 0$, where

$$\Theta_\alpha^n(p) := \sqrt{n^2\alpha^2 + [\maxroot(W_\alpha^n p)]^2}.$$

Furthermore, equality holds if and only if $p(x) = x^d$ or $q(x) = x^d$.

One obvious difference between the two theorems is the replacement of the U_α operator with the more general W_α^n operator, defined in Section 3 (the differences between the two are discussed in Remark 3.1). Much of the added difficulty in the proof of Theorem 1.3 (with respect to the proof of Theorem 1.2) is the quadratic nature of the W_α^n operator.

The other obvious difference is that the bound in (1.1) no longer relates the maximum root of the convolution to that of the original polynomials in a linear way (unless $n = 0$). We have yet to come up with an intuitive explanation of why this is the correct form, apart from it coming up naturally in the work of Benaych-Georges (see Section 1.3). However, it is worth noting

that the form of W_α^n (and Θ_α^n) only plays a significant role in Section 5 (the base cases). The inductive steps in Section 4 can be adapted to work with a much larger family of operators (and quantities) with the appropriate monotonicity properties (as was pointed out by an anonymous referee). One corollary of the results in Section 3, however is that, for fixed n , the quantity $\Theta_\alpha^n(p)$ is increasing in α . Given that $\Theta_0^n(p) = \text{maxroot}(p)$, this will provide the type of bound we are hoping for: one that quantitatively bound the effects of \boxplus_d^n on the largest root (with respect to the input polynomials).

In a different direction, Leake and Ryder showed that Theorem 1.1 was actually a special case of a more general submodularity inequality [10] (note that they use the notation \boxplus^d as opposed to the \boxplus_d notation introduced in [14]).

Theorem 1.4 (Leake, Ryder). *For any real rooted polynomials p, q, r of degree at most d , one has*

$$\text{maxroot}([p \boxplus_d q \boxplus_d r]) + \text{maxroot}(r) \leq \text{maxroot}(p \boxplus_d r) + \text{maxroot}(q \boxplus_d r)$$

Theorem 1.1 is then the case $r(x) = x^d - \frac{\alpha}{d}x^{d-1}$. The proof method of [10] is similar to that of [14] (and therefore this paper as well) however it is unclear whether it is possible to generalize any of the other convolutions in a similar way.

1.2. Structure

The paper is structured as follows: in the remaining introductory sections, we attempt to provide some concrete motivation for studying the rectangular additive convolution. In Section 2, we define the rectangular additive convolution \boxplus_d^n and prove some of the basic properties that it satisfies. Of particular importance in application is Theorem 2.3, which shows that \boxplus_d^n preserves nonnegative rootedness. In Section 3, we introduce two equivalent methods — the H -transform and W_α^n operator — that we will use to measure the effect of \boxplus_d^n on the largest root (with respect to the input polynomials).

The proof of Theorem 1.3 will be presented in Section 5, however it will require a number of lemmas and simplifications that will need to be proved along the way. Section 4 contains two “induction” lemmas that will allow us to reduce the proof of Theorem 1.3 to a subset of polynomials we call basic polynomials (see Section 2 for definitions). One of the main ingredients in the inductions is a “pinching” technique similar to the one used in [14] which is introduced in Section 4.1.

Given these induction lemmas, the primary difficulty remaining will be in proving various “base cases” of Theorem 1.3. These will be addressed in Section 5 modulo one major assumption: a bound on the Cauchy transform of a certain class of orthogonal polynomials (the *Gegenbauer polynomials*). The missing bound will then be proved in Section 6, along with a number of other nonasymptotic inequalities concerning Gegenbauer polynomials that may be of independent interest. Section 6 has been (to a large extent) quarantined from the rest of the paper in an attempt to allow readers with a primary interest in orthogonal polynomials to find it accessible without needing to read other parts of the paper.

In Section 7, we attempt to give some added understanding regarding the action of the (somewhat opaque) operator W_α^n through a set of informative examples. Finally, in Section 8, we end with some open problems.

1.3. Motivation for this work

Before beginning, we would like to give some motivation for the study of this particular convolution. In particular, Section 1.3.2 and Section 1.3.3 will discuss the relationship to a field first initiated by Voiculescu known as “free probability” [22]. The discussion of free probability will be restricted to these sections, and the remainder of the paper should be accessible to those unfamiliar with this area. However, for those interested in (or at least aware of) the connections of polynomial convolutions to linear algebra (in particular, expected characteristic polynomials) may find these sections useful for understanding the motivations for some of what will appear. For those interesting in learning more about this subject (and, in particular, the large role that combinatorics plays), we suggest starting with [18].

1.3.1 As an operation on singular values

We start by recalling the relationship between the symmetric additive convolution from [14] and eigenvalues of matrices. There is a natural correspondence between any degree d real-rooted polynomial p and the class of $d \times d$ Hermitian matrices H for which

$$p(x) = \det [x\mathbb{I} - H].$$

It was shown in [14] that the symmetric additive convolution of two degree d real-rooted polynomials is (again) a degree d real-rooted polynomial. Hence the symmetric additive convolution can be viewed as a binary operation on these classes of matrices. This may seem coincidental, but it was shown in [13] that the symmetric additive convolution can be reproduced by the actual addition of matrices; that is, for all Hermitian H and K , there exists a unitary matrix U for which

$$\det [x\mathbb{I} - H] \boxplus_d \det [x\mathbb{I} - K] = \det [x\mathbb{I} - (H + U^*KU)].$$

Hence, for example, the symmetric additive convolution must satisfy Horn’s inequalities [8] (precisely where in the Horn polytope the convolution lies is an interesting open question).

In a similar spirit, a polynomial in \mathcal{P}_d^+ can be written as $\det [x\mathbb{I} - AA^*]$ for some $d \times (d+n)$ matrix A (where we are free to choose $n \geq 0$). In Theorem 2.3, we show that the rectangular additive convolution of two polynomials in \mathcal{P}_d^+ is another polynomial in \mathcal{P}_d^+ . Hence, similar to above, the rectangular additive convolution can be seen as a binary operation on the classes of matrices with a given set of singular values. The one small issue in that the roots of the polynomial $\det [x\mathbb{I} - AA^*]$ are actually the *squares* of the singular values of A , and one can view the appearance of the \mathbb{S} operator in the definition of W_α^n as a correction to this issue (see Remark 3.1).

While this relationship to matrices is a primary motivation for defining (and studying) this particular convolution, we will see that (similar to [14]) that it is beneficial to ignore this relationship when proving bounds. In particular, one part of the induction we will use requires us to

be able to apply the rectangular additive convolution to polynomials of different degrees, which would correspond to adding matrices of different sizes in the view discussed above. Hence we will refrain from using this view for the majority of this paper, instead working strictly in the realm of polynomials. The two exceptions to this are Remark 3.1 and Section 7, where we will use the connection to matrices to help explain certain aspects of our analysis that may not be obvious when viewing the rectangular additive convolution completely in terms of polynomials. In particular, none of the main results will require any previous knowledge of free probability (or finite free probability).

1.3.2 Inspiration from free probability

Much of the connection between polynomial convolutions and free probability stems from Remark 3.1; in many ways, both theories consist of the only reasonable way to define a unitarily invariant binary operations on matrices. Hence the fact that polynomial convolutions turn into free convolutions (in the appropriate limit) is not surprising (proving that this is the case is less straight-forward [13]). While the concepts in [14, 16, 17] were developed without any knowledge of this connection, more recent work (including this paper) has benefited greatly from this relationship.

The connection is perhaps best seen by recalling that the primary tool used in [14] to understand the behavior of the symmetric additive convolution was the differential transform

$$U_\alpha p = p - \alpha p'$$

for $\alpha \geq 0$. This is nothing more than a more “polynomial friendly” version of the Cauchy transform (of the uniform distribution on the roots of p):

$$\mathcal{G}_p(x) = \frac{1}{\deg(p)} \frac{p'(x)}{p(x)} = \frac{1}{\deg(p)} \sum_i \frac{1}{x - \lambda_i(p)}$$

where $\lambda_i(p)$ denote the roots of p . In particular, one can check that

$$\maxroot(U_\alpha p) = t \iff \mathcal{G}_p(t) = \frac{1}{\alpha \deg(p)}$$

and so many of the properties of U_α can be derived directly from well known properties of the Cauchy transform. On the other hand, U_α has the advantage of remaining in the realm of real rooted polynomials (it is well known that the operator $(1 - \partial)$ preserves real rootedness, see for example [4]) and this turns out to be useful in the analysis done in [14].

In order to understand the behavior of the rectangular convolution, we will do something similar. Instead of relating to the Cauchy transform, however, we will use a construction inspired by the work of Benaych-Georges [3]. The H -transform defined in this paper is a slight modification of what Benaych-Georges calls a *rectangular Cauchy transform* with much of the modification coming from the fact that the objects of interest in [3] are infinite dimensional operators, and so one must view the relationship between n and d as a ratio. Some *a posteriori* explanations of the differences between U_α and W_α^n are discussed in Remark 3.1, but this is not intended to obscure the fact that all of our *a priori* inspiration came directly from [3].

1.3.3 Finite free probability

The original inspiration for [14, 13] came from the study of expected characteristic polynomials of matrices. In fact the definition of the rectangular additive convolution in Section 2 was originally derived from the following observation [12]:

Theorem 1.5. *Let A and B be $m \times n$ rectangular matrices, with (singular value) characteristic polynomials*

$$p(x) = \det [xI - AA^*] \quad \text{and} \quad q(x) = \det [xI - BB^*]$$

Let dQ and dR denote Haar invariant measures over $U(m)$ and $U(n)$ (the spaces of unitary matrices of size m and n , respectively). Then the coefficients of the polynomial

$$r(x) = \int_{Q,R} \det [xI - (A + QBR)(A + QBR)^*] dQdR$$

are multilinear functions of the coefficients of p and q .

One can check that the exact formula for $r(x)$ in terms of $p(x)$ and $q(x)$, derived in [12], is given by the rectangular additive convolution. In this respect, this paper can be viewed as the continuation of an investigation into the correspondence between expected characteristic polynomials and free probability that has become known as “finite free probability.”

Finite free probability has played an important role in some recent results in computer science [17, 25], and while we will not discuss any applications in this paper, we should mention that this paper is the first in a series of three papers. The second will be an extension of [17] (which showed the existence of bipartite Ramanujan graphs for any number of vertices n and degree d) to the case of biregular, bipartite graphs [6]. The third will be an analogue of the analysis done in [13], showing that the inequality in Theorem 1.3 becomes an equality in the appropriate limit by showing that the individual terms converge to Benaych-Georges’ rectangular R -transform [7]. The third paper, in particular, suggests that the rectangular additive convolution can be viewed as a finite version of the addition of freely independent rectangular matrices. In this respect, the bounds given in this paper show that the roots of $r(x)$ lie inside the support of the free convolution, a property which we will combine with the “method of interlacing polynomials” introduced in [15] to build Ramanujan graphs.

2. The Convolution

For $j \geq 1$, we define \mathcal{P}_j^+ to be the collection of degree j polynomials with real coefficients that have

1. All nonnegative roots
2. At least one root positive
3. The coefficient of x^j positive

and set \mathcal{P}_j^+ to be the (constant) 0 polynomial for all $j \leq 0$. Note that the second property only serves to eliminate the polynomial x^j and that this is the only polynomial that is added by the closure:

$$\overline{\mathcal{P}_j^+} = \mathcal{P}_j^+ \cup \{x^j\}.$$

We will use $\mathcal{P}_{\leq j}^+$ to denote the union $\bigcup_{k=1}^j \mathcal{P}_k^+$. Note that this *does not* include the 0 polynomial.

We will call a polynomial $p \in \mathcal{P}_j^+$ *basic* if it has the form $p(x) = c(x-t)^j$ for some real numbers $c, t > 0$. Otherwise we call it *nonbasic*. Note that, trivially, every polynomial in \mathcal{P}_1^+ is basic.

We define a binary operation on polynomials \boxplus_d^n as follows:

Definition 2.1. For $i, j \leq d$ and $n \geq 0$, we define the *rectangular additive convolution* to be the linear extension of the operation

$$[x^i \boxplus_d^n x^j](y) = \begin{cases} \frac{(n+i)!(n+j)!j!i!}{(d+n)!d!(i+j-d)!(n+i+j-d)!} y^{i+j-d} & \text{when } i+j \geq d \\ 0 & \text{otherwise} \end{cases} \quad (2.1)$$

In particular, if we write

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

then we have

$$[p \boxplus_d^n q](x) = \sum_{\ell} x^{d-\ell} \left(\sum_{i+j=\ell} \frac{(d-i)!(d-j)!}{d!(d-\ell)!} \frac{(n+d-i)!(n+d-j)!}{(n+d)!(n+d-\ell)!} a_i b_j \right). \quad (2.2)$$

There are two special cases worth mentioning: for any polynomial $p \in \mathcal{P}_{\leq d}^+$, we have

1. $[p \boxplus_d^n x^d] = p$, and
2. $[p \boxplus_d^n x^{d-1}] = \frac{1}{(d+n)d} (xp''(x) + (n+1)p'(x))$.

One property that can be derived directly from the definition is the following:

Lemma 2.2. Let $p \in \mathcal{P}_{\leq d}^+$ and $q \in \mathcal{P}_{\leq d-1}^+$. Then

$$[p \boxplus_d^n q] = \frac{1}{d(d+n)} [[p \boxplus_d^n x^{d-1}] \boxplus_{d-1}^n q]$$

Proof. By linearity, it suffices to prove this when p is a monomial of degree at most d and q a monomial of degree at most $d-1$. Using (2.1) on $p(x) = x^j$, we have

$$[x^j \boxplus_d^n x^{d-1}] = \begin{cases} \frac{j(n+j)}{d(n+d)} x^{j-1} & \text{for } 0 < j \leq d \\ 0 & \text{for } j = 0 \end{cases}$$

and then the lemma follows by plugging into (2.1). \square

2.1. Preservation of nonnegative real roots

The main property that we will need in this paper is the fact that the rectangular additive convolution (with the appropriate parameters) preserves the property of having all nonnegative roots.

Theorem 2.3. *Let $1 \leq i \leq j \leq d$. Then for any $p \in \overline{\mathcal{P}_i^+}$ and $q \in \overline{\mathcal{P}_j^+}$, we have $[p \boxplus_d^n q](x) \in \overline{\mathcal{P}_{i+j-d}^+}$.*

The proof of this theorem will rely on a result of Lieb and Sokal [11]. Recall that a polynomial f is called *stable* if $f(z_1, \dots, z_k) \neq 0$ whenever $\Im(z_i) > 0$ for all i . If (in addition) the coefficients of f are real numbers, then f is called *real stable*. A degree d univariate polynomial is stable if and only if it has d real roots. A multivariate polynomial is called *multiaffine* if it has degree at most 1 in each of its variables.

Theorem 2.4 (Lieb–Sokal). *If $f(y_1, \dots, y_n)$ and $g(x_1, \dots, x_n)$ are multiaffine real stable polynomials, then*

$$f(-\partial_1, \dots, -\partial_n)\{g(x_1, \dots, x_n)\}$$

is either the 0 polynomial or is multiaffine real stable.

We will also need the following well-known lemma:

Lemma 2.5. *For any degree j univariate polynomial p with positive leading coefficient, we have $p \in \overline{\mathcal{P}_j^+}$ if and only if $p(xy)$ is real stable.*

Proof. Assume $p(xy)$ is not real stable; that is, $p(xy) = 0$ for some x, y with $\Im(x) > 0$ and $\Im(y) > 0$. Without loss of generality, we can set $x = re^{i\theta_1}$ and $y = Re^{i\theta_2}$ where $0 < \theta_1 \leq \theta_2 < \pi$. Hence $xy = rRe^{i\theta_3}$ is a root of p , and since

$$\theta_3 = \theta_1 + \theta_2 \in (0, 2\pi)$$

this root is not a nonnegative real, so $p \notin \overline{\mathcal{P}_j^+}$.

In the other direction, if $p \notin \overline{\mathcal{P}_j^+}$ then $p(x^2)$ has a non-real root. Since p has real coefficients, the roots come in conjugate pairs so one such root must be in the upper half plane. Setting x and y to be that root shows that $p(xy)$ is not real stable. \square

We will write pol_x^k to denote the map

$$\text{pol}_x^k[x^j] = \frac{\sigma_j(x_1, \dots, x_k)}{\binom{k}{j}}$$

where σ_j is the j^{th} elementary symmetric polynomial on the inputs x_1, \dots, x_k (and is undefined when $j > k$). The linear extension of the map pol_x^i to polynomials is known as the *polarization* operator. It should be clear that $\text{pol}_x^i[p(x)]$ is multiaffine for any polynomial p and any integer i . A theorem of Borcea and Brändén shows that polarization preserves the property of being real stable [4]:

Theorem 2.6. *If $p(x, y_1, \dots, y_n)$ is a real stable polynomial then $\text{pol}_x^i[p]$ is real stable as well.*

Proof of Theorem 2.3. We first note that it suffices to consider the case when $i = j = d$. To see why, we can proceed by induction on d . Since $i = j = 1$ is the only possibility when $d = 1$, the base case will be covered. Now for $d > 1$, if $i < d$ or $j < d$, then we can use Lemma 2.2 to write the same polynomial using a convolution with parameter $d - 1$, and this is real rooted by the induction hypothesis.

To ease notation, we will write

$$\theta_x^i := \sigma_i(x_1, \dots, x_d) \quad \text{and} \quad \theta_y^i := \sigma_i(y_1, \dots, y_{n+d}).$$

and

$$\delta_x^i := \sigma_i(\partial_{x_1}, \dots, \partial_{x_d}) \quad \text{and} \quad \delta_y^i := \sigma_i(\partial_{y_1}, \dots, \partial_{y_{n+d}}).$$

Given polynomials $p, q \in \overline{\mathcal{P}_d^+}$, with $p(x) = \sum_i x^{d-i} a_i$ and $q(x) = \sum_i x^{d-i} b_i$, set

$$f(x_1, \dots, x_d, y_1, \dots, y_{d+n}) = \text{pol}_x^d [\text{pol}_y^{d+n} [y^n p(xy)]] = \sum_i \frac{\theta_x^{d-i}}{\binom{d}{i}} \frac{\theta_y^{n+d-i}}{\binom{n+d}{i}} a_i$$

and

$$g(x_1, \dots, x_d, y_1, \dots, y_{d+n}) = \text{pol}_x^d \left[\text{pol}_y^{d+n} \left[(xy)^d q \left(\frac{1}{xy} \right) \right] \right] = \sum_i \frac{\theta_x^i}{\binom{d}{i}} \frac{\theta_y^i}{\binom{n+d}{i}} b_i$$

Both f and g are multiaffine (by definition of polarization) and real stable (by a combination of Lemma 2.5 and Theorem 2.6). Hence Theorem 2.4 implies that

$$\begin{aligned} h(x_1, \dots, x_d, y_1, \dots, y_{d+n}) \\ = g(-\partial_{x_1}, \dots, -\partial_{x_d}, -\partial_{y_1}, \dots, -\partial_{y_{n+d}}) \{f(x_1, \dots, x_d, y_1, \dots, y_{d+n})\} \end{aligned}$$

is either 0 or real stable. If it is 0, then we are done, so assume not. Since substitution of variables preserves real stability, the bivariate polynomial $r(x, y)$ formed by substituting $x \leftarrow x_i$ and $y \leftarrow y_i$ into h is also real stable. We claim that

$$r(x, y) = y^n [p \boxplus_d^n q](xy)$$

which, by Lemma 2.5, would complete the proof.

The main observation that we will need to compute h is that for any $\ell > 0$,

$$\sigma_i(\partial_{z_1}, \dots, \partial_{z_\ell}) \{ \sigma_j(z_1, \dots, z_\ell) \} = \begin{cases} \binom{\ell+i-j}{i} \sigma_{j-i}(z_1, \dots, z_\ell) & \text{for } i \leq j \\ 0 & \text{otherwise} \end{cases}$$

which can easily be checked by hand. In particular, this gives

$$\delta_x^i \theta_x^{d-j} = \binom{i+j}{i} \theta_x^{d-i-j} \quad \text{and} \quad \delta_y^i \theta_y^{n+d-j} = \binom{i+j}{i} \theta_y^{n+d-i-j}$$

when $i \leq j$ (and 0 otherwise). Applying this to our formula for h , we get

$$\begin{aligned} h(x_1, \dots, x_d, y_1, \dots, y_{d+n}) &= \sum_{i=0}^d b_i (-1)^{2i} \frac{\delta_x^i}{\binom{d}{i}} \frac{\delta_y^i}{\binom{d+n}{i}} \sum_{j=0}^d a_j \frac{\theta_x^{d-j}}{\binom{d}{j}} \frac{\theta_y^{n+d-j}}{\binom{n+d}{j}} \\ &= \sum_{i=0}^d \sum_{j=0}^d a_j b_i \frac{\binom{i+j}{i} \theta_x^{d-i-j}}{\binom{d}{i} \binom{d}{j}} \frac{\binom{i+j}{i} \theta_y^{n+d-i-j}}{\binom{n+d}{j} \binom{n+d}{i}} \end{aligned}$$

Plugging in $x \leftarrow x_i$ and $y \leftarrow y_i$, we then get

$$\begin{aligned} r(x, y) &= \sum_{\ell=0}^d \sum_{i+j=\ell} a_j b_i \frac{\binom{\ell}{i} \binom{d}{\ell} x^{d-\ell}}{\binom{d}{i} \binom{d}{j}} \frac{\binom{\ell}{i} \binom{d+n}{\ell} y^{d+n-\ell}}{\binom{d+n}{j} \binom{d+n}{i}} \\ &= y^n \sum_{\ell=0}^d (xy)^{d-\ell} \left(\sum_{i+j=\ell} a_j b_i \frac{(d-i)!(d-j)!}{d!(d-\ell)!} \frac{(n+d-i)!(n+d-j)!}{(n+d)!(n+d-\ell)!} \right) \\ &= y^n [p \boxplus_d^n q](xy). \end{aligned} \quad \square$$

3. The measuring stick

In this section, we define the W_α^n operator that will be used to measure the effects of the rectangular additive convolution on the largest root of a polynomial. Rather than define it directly, however, it will be useful to first introduce a modification of the H -transform from [3], which will then have W_α^n as its corresponding differential operator. To state the two succinctly, it will help to introduce two other operators:

$$[\mathbb{S}p](x) = p(x^2) \quad \text{and} \quad [\mathbb{V}^n p](x) = x^n p(x). \tag{3.1}$$

Given a polynomial with nonnegative roots, we define the H -transform of p (with parameter n) as

$$\mathcal{H}_p^n(x) = [\log \mathbb{S}p]' [\log \mathbb{S}\mathbb{V}^n p]' = \frac{[\mathbb{S}p]'}{\mathbb{S}p} \left(\frac{2n}{x} + \frac{[\mathbb{S}p]'}{\mathbb{S}p} \right) \tag{3.2}$$

and the corresponding differential operator

$$W_\alpha^n p = [\mathbb{S}p][\mathbb{S}\mathbb{V}^n p] - \alpha^2 [\mathbb{S}p]' [\mathbb{S}\mathbb{V}^n p]'. \tag{3.3}$$

Note that the parameter n is intended to be the same as the parameter n in (2.1). One could consider \mathcal{H}_p^k and $W_\alpha^k p$ for general value of k , but we will only be using these transforms to directly measure the effect of \boxplus_d^n and so there is no reason to consider this more general case.

Remark 3.1. For those familiar with the U_α operator from [14], we would like to point out that there is a natural way to understand the differences between that operator and the W_α^n defined here using the relationship to characteristic polynomials mentioned in Section 1.3.1. As discussed there, the symmetric additive convolution can be viewed as a binary operation on eigenvalues (technically, classes of Hermitian matrices with the same eigenvalues) and, in a similar spirit, the

rectangular additive convolution can be viewed as a binary operation on singular values (technically, classes of matrices with the same dimensions and singular values). This correspondence goes via the characteristic polynomials $\det [x\mathbb{I} - H]$ with H Hermitian (in the eigenvalue case) and $\det [x\mathbb{I} - AA^*]$ with A a $d \times (n+d)$ matrix with $n \geq 0$ (in the singular value case). The second correspondence is not perfect, however, in that the roots of the polynomial $\det [x\mathbb{I} - AA^*]$ are actually the *squares* of the singular values of A .

The first main difference between the U_α and W_α^n operators — the appearance of the \mathbb{S} operator — can be viewed as a necessary compensation for this issue. One concern one might have is that this “correction” effectively creates two copies of each singular value (a positive one and a negative one). Fortunately, our primary interest is in understanding the largest singular value (that is, the largest root of $\mathbb{S}p$), and so the addition of extra negative roots will prove to be irrelevant in our analysis.

The remaining differences between the U_α and W_α^n operators — the quadratic nature and appearance of the \mathbb{V}^n operator (and the parameter n in general) — can likewise be seen as a necessary compensation for a different issue. Given the polynomial $\det [x\mathbb{I} - AA^*]$, there is no way to know the number of columns in the underlying matrix A . One might hope that this is irrelevant, but the analysis in [3] shows that this is not the case. One obvious attempt to correct this would be to instead consider the polynomial $\det [x\mathbb{I} - A^*A]$, but one runs into the same problem concerning the number of rows (and therefore the value of d). The W_α^n operator compensates for this by using *both* polynomials

$$p(x) = \det [x\mathbb{I} - AA^*] \quad \text{and} \quad \mathbb{V}^n p(x) = \det [x\mathbb{I} - A^*A]$$

to be used. The one case where the W_α^n operator would not need both polynomials is of course when $n = 0$. In this case, $\maxroot (W_\alpha^0 p)$ becomes a difference of squares (which can then be factored):

$$\begin{aligned} \maxroot (W_\alpha^0 p) &= \maxroot (p(x^2)^2 - 4x^2\alpha^2 p'(x^2)^2) = \maxroot (p(x^2) - 2\alpha x p'(x^2)) \\ &= \maxroot (U_\alpha \mathbb{S}p). \end{aligned}$$

This is the reason that the analysis of the asymmetric convolution in [14] could be done using the simpler U_α operator (despite being a special case of the results in this paper).

As in [14], we would like to be able to associate the function \mathcal{H}_p^n with the largest root of the polynomial $W_\alpha^n p$ but for this to be well-defined, we first need to show that $W_\alpha^n p$ has real roots. Our proof will use a classical result in real rooted polynomials [4]:

Theorem 3.2 (Hermite–Biehler). *Let f and g be polynomials with real coefficients. Then the following are equivalent:*

1. Every root λ of $f + ig$ satisfies $\Im\{\lambda\} \leq 0$ (here \Im denotes the imaginary part)
2. f and g are real rooted and the polynomial $fg' - gf' \leq 0$ at any point $x \in \mathbb{R}$.

Using this, it is easy to show the following lemma, which will immediately imply what we need:

Lemma 3.3. *Let p and q be real rooted polynomials. Then*

$$p(x)q(x) - p'(x)q'(x)$$

is real rooted.

Proof. It is an easy consequence of Rolle’s theorem that p' is real rooted whenever p is, and then one can check that

$$p(x)p''(x) - p'(x)^2 \leq 0$$

for all x by noticing

$$\frac{p(x)p''(x) - p'(x)^2}{p(x)^2} = (\log p(x))'' = \sum_i -\frac{1}{(x - \lambda_i(p))^2}.$$

Hermite–Biehler then implies that the polynomials

$$p(x) + ip'(x) \quad \text{and} \quad q(x) + iq'(x)$$

have no roots in the upper half plane. Hence their product

$$pq + i(p'q + q'p) - p'q'$$

has no roots in the upper half plane as well. So Hermite–Biehler (in the opposite direction) gives that $pq - p'q'$ is real rooted. \square

We therefore have the following correspondence, which amounts to nothing more than algebraic manipulation:

Corollary 3.4.

$$\mathcal{H}_p^n(t) = \frac{1}{\alpha^2} \iff \text{maxroot}(W_\alpha^n p) = t$$

3.1. Monotonicity Properties

One of the most important properties of the function $\mathcal{G}_p(x)$ when p is real rooted is that it is strictly monotone decreasing on the interval $[\text{maxroot}(p), \infty)$ (something that can be seen directly from the definition). Our new functions will inherit that property:

Lemma 3.5. *For all polynomials p with nonnegative roots, the function $\mathcal{H}_p^n(x)$ is strictly decreasing at any $x > \sqrt{\text{maxroot}(p)}$.*

Proof. Let $t = \text{maxroot}(p)$. Since p has nonnegative roots, both $\mathbb{S}p$ and $\mathbb{S}\mathbb{V}^n p$ are real rooted with

$$\text{maxroot}(\mathbb{S}p) = \text{maxroot}(\mathbb{S}\mathbb{V}^n p) = \sqrt{t}.$$

Hence both $\mathcal{G}_{\mathbb{S}p}(x)$ and $\mathcal{G}_{\mathbb{S}\mathbb{V}^n p}(x)$ are strictly decreasing for $x \geq \sqrt{t}$ and so the product is strictly decreasing as well. \square

This then implies two “inverse” properties for the differential operators:

Corollary 3.6. *Let p be a polynomial with nonnegative roots. Then for any $\alpha > 0$, the quantity*

$$r(\alpha) := \text{maxroot}(W_\alpha^n p)$$

is strictly increasing.

Proof. By Corollary 3.4, we have

$$\mathcal{H}_p^n(r(\alpha)) = \frac{1}{\alpha^2}.$$

Taking derivatives (in α) on both sides gives

$$[\mathcal{H}_p^n]'(r(\alpha))r'(\alpha) = \frac{-1}{\alpha^3} < 0$$

where we know from Lemma 3.5 that $[\mathcal{H}_p^n]'(r(\alpha)) < 0$. Hence it must be that $r'(\alpha) > 0$. \square

The one disadvantage of these new differential operators is their quadratic nature. In the case where we will only need qualitative information about two polynomials, we will be able to reduce to the simpler case:

Lemma 3.7. *Let p, q be polynomials with nonnegative roots. Then for any $x > \sqrt{\text{maxroot}(pq)}$, we have*

$$\mathcal{H}_p^n(x) \leq \mathcal{H}_q^n(x) \iff [\log p]'(x^2) \leq [\log q]'(x^2).$$

Proof. Let

$$w_p = [\log \mathbb{S}p]'(x) = 2x \frac{p'(x^2)}{p(x^2)} \tag{3.4}$$

and similarly for w_q . By definition, we therefore have

$$\mathcal{H}_p^n(x) = \frac{n}{x}w_p + w_p^2 \quad \text{and} \quad \mathcal{H}_q^n(x) = \frac{n}{x}w_q + w_q^2$$

and so the difference is

$$\mathcal{H}_p^n(x) - \mathcal{H}_q^n(x) = (w_p - w_q) \left(\frac{n}{x} + w_p + w_q \right)$$

where for $x > \text{maxroot}(pq)$ we have $\frac{n}{x}, w_p, w_q > 0$. Hence

$$\mathcal{H}_p^n(x) - \mathcal{H}_q^n(x) \quad \text{and} \quad w_p - w_q$$

have the same sign. But by (3.4), we can write

$$w_p - w_q = 2x \frac{p'(x^2)}{p(x^2)} - 2x \frac{q'(x^2)}{q(x^2)}$$

and since $x > 0$, the lemma follows. \square

Using Lemma 3.7, the primary inequality that we will need becomes a simple calculation.

Lemma 3.8. *Let p, q, r be polynomials with nonnegative roots and positive leading coefficients such that $p + q = r$. Furthermore, assume each has at least one positive root. Then for all $n \geq 0$ and all $\alpha > 0$, we have*

$$\min \{ \mathcal{H}_p^n(t), \mathcal{H}_q^n(t) \} \leq \mathcal{H}_r^n(t) \leq \max \{ \mathcal{H}_p^n(t), \mathcal{H}_q^n(t) \}$$

whenever $t > \sqrt{\max\text{root}(pqr)}$. Furthermore, equality holds in one direction if and only if it holds in the other direction.

Proof. By Lemma 3.7, it suffices to show that

$$\min \{ [\log p]'(s), [\log q]'(s) \} \leq [\log r]'(s) \leq \max \{ [\log p]'(s), [\log q]'(s) \}$$

for any s which is greater than $\max\text{root}(pqr)$. Now notice that

$$\frac{r'}{r} = \frac{p' + q'}{r} = \frac{p p'}{r p} + \frac{q q'}{r q}$$

where p/r and q/r are both positive at s (since s is larger than the maximum root of each polynomial, and each polynomial has a positive leading coefficient). So setting $\lambda = p(s)/r(s)$, we have

$$[\log r]'(s) = \lambda [\log p]'(s) + (1 - \lambda) [\log q]'(s)$$

for some $0 < \lambda < 1$. The lemma then follows. □

Of particular importance will be the case when $t = \max\text{root}(W_\alpha^n p)$.

Corollary 3.9. *Let p, q, r satisfy the conditions of Lemma 3.8 with*

$$\beta_p = \max\text{root}(W_\alpha^n p) \quad \text{and} \quad \beta_q = \max\text{root}(W_\alpha^n q) \quad \text{and} \quad \beta_r = \max\text{root}(W_\alpha^n r).$$

Assume further that $\beta_r \geq \sqrt{\max\text{root}(pqr)}$. Then

$$\min \{ \beta_p, \beta_q \} \leq \beta_r \leq \max \{ \beta_p, \beta_q \}$$

with two of the β_k equal if and only if all three of the β_k are equal.

Proof. Set $t = \beta_r$ in Lemma 3.8 and assume (without loss of generality) that

$$\mathcal{H}_p^n(\beta_r) \leq \mathcal{H}_r^n(\beta_r) \leq \mathcal{H}_q^n(\beta_r) \tag{3.5}$$

By Lemma 3.4, we have

$$\mathcal{H}_p^n(\beta_p) = \mathcal{H}_q^n(\beta_q) = \mathcal{H}_q^n(\beta_r) = \frac{1}{\alpha^2}$$

Hence we can rewrite (3.5) as

$$\mathcal{H}_p^n(\beta_r) \leq \mathcal{H}_p^n(\beta_p) \quad \text{and} \quad \mathcal{H}_q^n(\beta_q) \leq \mathcal{H}_q^n(\beta_r)$$

which, by Lemma 3.5 implies that

$$\beta_q \leq \beta_r \leq \beta_p$$

which satisfies the desired conclusion. □

4. Inductions

The goal of this section is to prove two “induction steps” that will be useful in the proof of Theorem 1.3. Let

$$\phi_\alpha^{n,d}(p, q) := \Theta_\alpha^n(p) + \Theta_\alpha^n(q) - \Theta_\alpha^n(p \boxplus_d^n q) - (n + 2d)\alpha \quad (4.1)$$

where

$$\Theta_\alpha^n(p) := \sqrt{n^2\alpha^2 + [\maxroot(W_\alpha^n p)]^2}.$$

Notice that Theorem 1.3 can be restated as saying that $\phi_\alpha^{n,d}(p, q) > 0$. Our first induction will be on the degree d .

Lemma 4.1. *Let $2 \leq j < d$. Assume that, for some fixed $q \in \mathcal{P}_{\leq d-1}^+$ and $\alpha > 0$, we have*

$$(i) \quad \phi_\alpha^{n,d-1}(p, q) > 0 \text{ for all } p \in \mathcal{P}_{j-1}^+$$

$$(ii) \quad \phi_\alpha^{n,d}(p, x^{d-1}) > 0 \text{ for all } p \in \mathcal{P}_j^+$$

Then $\phi_\alpha^{n,d}(p, q) > 0$ for all $p \in \mathcal{P}_j^+$.

Our second induction will be on the number of distinct roots in the polynomial.

Lemma 4.2. *Let $2 \leq j \leq d$. Assume that, for a fixed $q \in \mathcal{P}_{\leq d}^+$ and $\alpha > 0$, we have*

$$(i) \quad \phi_\alpha^{n,d}(p, q) > 0 \text{ for all polynomials } p \in \mathcal{P}_{j-1}^+$$

$$(ii) \quad \phi_\alpha^{n,d}(p, q) > 0 \text{ for all basic polynomials } p \in \mathcal{P}_j^+$$

Then $\phi_\alpha^{n,d}(p, q) > 0$ for all polynomials $p \in \mathcal{P}_j^+$.

The proofs of both lemmas are given in Section 4.2. The proof of Lemma 4.2 will utilize a decomposition of a nonbasic polynomial $p \in \mathcal{P}_j^+$ into two “simpler” polynomials that is proved in Section 4.1. By “simpler”, we mean that one of the polynomial will have lower degree (allowing for induction) and the other polynomial will have two of its roots moved closer together (an operation that is referred to in [14] as “pinching”). Obviously one can always pinch a nonbasic polynomial — the goal will be to find a pinch that decreases $\phi_\alpha^{n,d}(\cdot, q)$, as this would imply that any minimum of $\phi_\alpha^{n,d}(\cdot, q)$ (if it exists) must be a basic polynomial.

Because basic polynomials feature prominently in the computations moving forward, it will be useful to have calculated the following quantity beforehand:

Lemma 4.3. *Let $p(x) = c(x - \lambda)^j$ with $j \geq 1$ and $c, \lambda > 0$. Then*

$$\Theta_\alpha^n(p) = \sqrt{(n + j)^2\alpha^2 + \lambda} + \alpha j.$$

Proof. For p as given, we have

$$W_\alpha^n p(x) = c(x^2 - \lambda)^{2(j-1)} x^{2n} r_\lambda(x^2)$$

where $r_\lambda(x) = (x - \lambda)^2 - 4j(n + j)\alpha^2(x - \lambda) - 4j^2\alpha^2\lambda$. Hence

$$\maxroot(W_\alpha^n p)^2 = \maxroot(r_\lambda) = \lambda + 2(n + j)j\alpha^2 + 2j\alpha\sqrt{(n + j)^2\alpha^2 + \lambda} \quad (4.2)$$

and so

$$\Theta_\alpha^n(p)^2 = n^2\alpha^2 + \lambda + 2(n + j)j\alpha^2 + 2j\alpha\sqrt{(n + j)^2\alpha^2 + \lambda} = \left(\sqrt{(n + j)^2\alpha^2 + \lambda} + \alpha j\right)^2. \quad \square$$

4.1. The Pinch

The main goal of this subsection is to prove Lemma 4.4, which provides the existence of a pinch with the properties that will be useful in the proof of Lemma 4.2.

Lemma 4.4. *Fix $\alpha > 0$, $j \geq 2$, and let $p(x) \in \mathcal{P}_j^+$ be nonbasic. Then there exist $\tilde{p} \in \mathcal{P}_j^+$ and $\hat{p} \in \mathcal{P}_{j-1}^+$ so that*

- a. $p(x) = \hat{p}(x) + \tilde{p}(x)$,
- b. $\text{maxroot}(\tilde{p}) \leq \text{maxroot}(p)$,
- c. $\text{maxroot}(W_\alpha^n \tilde{p}) = \text{maxroot}(W_\alpha^n \hat{p}) = \text{maxroot}(W_\alpha^n p)$
- d. $\text{maxroot}(\hat{p}) \leq \text{maxroot}(W_\alpha^n p)$.

The proof of Lemma 4.4 will occur in two steps. First we will prove a pinching lemma for a “linearized” version of the W -transform (by removing the \mathbb{S} operator), and then we will show how the existence of a linearized pinch implies Lemma 4.4. Note that this is the only section where these linearized operators will appear.

We define the “linearized” versions of the H -transform as:

$$\widehat{\mathcal{H}}_p^n := [\log p]' \log[\mathbb{V}^n p]' = [\log p]' \left(\frac{n}{x} + [\log p]' \right)$$

and note that for two polynomials p, q , $[\log p]'(t) = [\log q]'(t)$ implies $\widehat{\mathcal{H}}_p^n(t) = \widehat{\mathcal{H}}_q^n(t)$. Corresponding to this is the “linearized” version of the W -transform

$$\widehat{W}_\zeta^n p(x) = [p][\mathbb{V}^n p] - \zeta^2 [p]'[\mathbb{V}^n p]'$$

which by Lemma 3.3 is real-rooted as long as p is real rooted. Hence we can say

$$\widehat{\mathcal{H}}_p^n(t) = \frac{1}{\zeta^2} \iff \text{maxroot}(\widehat{W}_\zeta^n p) = t. \tag{4.3}$$

Lemma 4.5 (Linearized pinching). *Let $\zeta > 0$, $j \geq 2$, and let $p(x) = (x - a)(x - b)r(x) \in \mathcal{P}_j^+$ be a monic polynomial with $0 \leq a < b = \text{maxroot}(p)$. Then there exist real numbers $\mu \in (a, b)$ and $\rho > b$ so that the polynomials*

$$\tilde{p}(x) = (x - \mu)^2 r(x) \quad \text{and} \quad \hat{p}(x) = \kappa(x - \rho)r(x)$$

satisfy

- a. $\tilde{p} \in \mathcal{P}_j^+$ and $\hat{p} \in \mathcal{P}_{j-1}^+$ (in particular, $\kappa > 0$)
- b. $p = \tilde{p} + \hat{p}$
- c. $\text{maxroot}(\widehat{W}_\zeta^n p) = \text{maxroot}(\widehat{W}_\zeta^n \tilde{p}) = \text{maxroot}(\widehat{W}_\zeta^n \hat{p})$
- d. $\rho < \text{maxroot}(\widehat{W}_\zeta^n p)$

Proof. Let $t = \maxroot\left(\widehat{W}_\zeta^n p\right)$. Since $\zeta > 0$, we have by Lemma 3.6 that $t > b > a$, so there is no issue in setting

$$\mu = t - \frac{2}{\frac{1}{t-a} + \frac{1}{t-b}}.$$

or (rearranging slightly)

$$\frac{2}{t-\mu} = \frac{1}{t-a} + \frac{1}{t-b}. \quad (4.4)$$

Note that (4.4) gives $(t-\mu)$ as the harmonic average of $(t-a)$ and $(t-b)$ (where $a \neq b$). Using the fact that $t > b > a$, this implies that $(t-b) < (t-\mu) < (t-a)$ or (equivalently) $a < \mu < b$. Furthermore,

$$[\log(x-c)(x-d)r(x)]'(t) = \frac{r'(t)}{r(t)} + \frac{1}{t-c} + \frac{1}{t-d}$$

so (4.4) also shows that $[\log p]'(t) = [\log \widehat{p}]'(t)$.

Now define $\widehat{p} = p - \widetilde{p}$. Since p and \widetilde{p} have the same leading coefficient, it should be clear that \widehat{p} has degree $j-1$. Furthermore, we can now write

$$[\log p]' = \frac{p'}{p} = \frac{\widehat{p}'}{\widehat{p}} + \frac{\widetilde{p}'}{\widetilde{p}} = \frac{\widehat{p}}{\widehat{p}}[\log \widehat{p}]' + \frac{\widetilde{p}}{\widetilde{p}}[\log \widetilde{p}]'.$$

Evaluating at the point t , we get that

$$[\log p]'(t) = \lambda[\log \widehat{p}]'(t) + (1-\lambda)[\log \widetilde{p}]'(t)$$

for some $\lambda \in (0, 1)$, so since $[\log p]'(t) = [\log \widehat{p}]'(t)$ we must have

$$[\log p]'(t) = [\log \widehat{p}]'(t) = [\log \widetilde{p}]'(t).$$

As is mentioned above, this implies

$$\widehat{\mathcal{H}}_p^n(t) = \widehat{\mathcal{H}}_{\widehat{p}}^n(t) = \widehat{\mathcal{H}}_{\widetilde{p}}^n(t)$$

and so by (4.3)

$$\maxroot\left(\widehat{W}_\zeta^n p\right) = \maxroot\left(\widehat{W}_\zeta^n \widetilde{p}\right) = \maxroot\left(\widehat{W}_\zeta^n \widehat{p}\right).$$

The fact that $a < \mu < b$ shows that $\widetilde{p} \in \mathcal{P}_j^+$, and so it remains to show that $\kappa > 0$ and that $\rho \in (b, t)$. We first multiply out the equation $p = \widetilde{p} + \widehat{p}$ and equate coefficients to get

$$\kappa = 2\mu - a - b \quad \text{and} \quad \rho = \frac{\mu^2 - ab}{2\mu - a - b}. \quad (4.5)$$

Now to see that $\kappa > 0$, recall that (4.4) expresses $t - \mu$ as the harmonic average of two positive numbers. The inequality between the arithmetic and harmonic means therefore implies

$$t - \mu < \frac{1}{2}(2t - a - b)$$

which, after rearranging, gives that $2\mu > a + b$.

To see that $\rho \in (b, t)$, we can solve for t in (4.4) to get

$$t = \frac{\mu(a + b) - 2ab}{2\mu - a - b}$$

and then use the formula for ρ in (4.5) to compute

$$\rho - b = \frac{(\mu - b)^2}{2\mu - a - b} > 0 \quad \text{and} \quad t - \rho = \frac{(b - t)(t - a)}{2\mu - a - b} > 0$$

as needed. □

To move from the “linearized” pinch to the “quadratic” pinch that we need, we will use the following observation:

Lemma 4.6. Fix $\alpha > 0$, $j \geq 2$, and let $p \in \mathcal{P}_j^+$. Then

$$\text{maxroot}(\widehat{W}_\alpha^n p) = t \iff \text{maxroot}(W_{2\alpha t}^n p) = t^2$$

Proof. By definition, we have

$$\begin{aligned} \text{maxroot}(\widehat{W}_\alpha^n p) = t &\implies W_\alpha^n p(t) = 0 \\ &\iff [\mathbb{S}q](t)[\mathbb{S}\mathbb{V}^n p](t) - \alpha^2[\mathbb{S}p]'(t)[\mathbb{S}\mathbb{V}^n p]'(t) = 0 \\ &\iff p(t^2)[\mathbb{V}^n p](t^2) - 4t^2\alpha^2[\mathbb{V}^n p]'(t^2)p'(t^2) = 0 \end{aligned}$$

and so t^2 is some root of $\text{maxroot}(W_{2\alpha t}^n p)$; we wish to show it is the largest one. Assume (for contradiction) that there exists some $\omega > t^2$ for which $\widehat{W}_{2\alpha t}^n p(\omega) = 0$. This would imply that

$$\widehat{\mathcal{H}}_p^n(\omega) = \widehat{\mathcal{H}}_p^n(t^2) = \frac{1}{4\alpha^2 t^2}.$$

However this is impossible, since $\widehat{\mathcal{H}}_p^n(x)$ is strictly decreasing whenever $x > \text{maxroot}(p)$ and we have assumed that $\omega > t^2 > \text{maxroot}(p)$. The reverse direction (once we have fixed the value of t) is essentially the same. □

Lemma 4.4 then follows easily:

Proof of Lemma 4.4. Apply Lemma 4.5 with $\zeta = 2\alpha t$ to get polynomials \tilde{p} and \hat{p} . Using Lemma 4.6, the properties provided by Lemma 4.5 then translate directly into the properties needed. □

4.2. The Lemmas

We now give proofs of the two main lemmas stated at the beginning of the section (the lemmas are restated here for convenience).

Lemma 4.1. Let $2 \leq j < d$. Assume that, for some fixed $q \in \mathcal{P}_{\leq d-1}^+$ and $\alpha > 0$, we have

$$(i) \quad \phi_\alpha^{n,d-1}(p, q) > 0 \text{ for all } p \in \mathcal{P}_{j-1}^+$$

$$(ii) \quad \phi_\alpha^{n,d}(p, x^{d-1}) > 0 \text{ for all } p \in \mathcal{P}_j^+$$

Then $\phi_\alpha^{n,d}(p, q) > 0$ for all $p \in \mathcal{P}_j^+$.

Proof. Since both p and q have degree less than d , we can apply Lemma 2.2 to get

$$[p \boxplus_d^n q] = \frac{1}{d(d+n)} [\Delta[p] \boxplus_{d-1}^n q]$$

where $\Delta[p] = [p \boxplus_d^n x^{d-1}] \in \mathcal{P}_{j-1}^+$. In particular,

$$\Theta_\alpha^n([p \boxplus_d^n q]) = \Theta_\alpha^n([\Delta[p] \boxplus_{d-1}^n q])$$

where by (i) we have

$$\Theta_\alpha^n([\Delta[p] \boxplus_{d-1}^n q]) < \Theta_\alpha^n(\Delta[p]) + \Theta_\alpha^n(q) - (n + 2(d-1))\alpha$$

and by (ii) we have

$$\Theta_\alpha^n(\Delta[p]) < \Theta_\alpha^n(p) + \Theta_\alpha^n(x^{d-1}) - (n + 2d)\alpha.$$

where $\Theta_\alpha^n(x^{d-1}) = (n + 2(d-1))\alpha$ by Lemma 4.3. Combining these gives

$$\Theta_\alpha^n([p \boxplus_d^n q]) < \Theta_\alpha^n(p) + \Theta_\alpha^n(q) - (n + 2d)\alpha$$

as required. □

Lemma 4.2. Let $2 \leq j \leq d$. Assume that, for a fixed $q \in \mathcal{P}_{\leq d}^+$ and $\alpha > 0$, we have

$$(i) \quad \phi_\alpha^{n,d}(p, q) > 0 \text{ for all polynomials } p \in \mathcal{P}_{j-1}^+$$

$$(ii) \quad \phi_\alpha^{n,d}(p, q) > 0 \text{ for all basic polynomials } p \in \mathcal{P}_j^+$$

Then $\phi_\alpha^{n,d}(p, q) > 0$ for all polynomials $p \in \mathcal{P}_j^+$.

Proof. Fix α and q and assume (for contradiction) that there exists a polynomial $p \in \mathcal{P}_j^+$ for which $\phi_\alpha^{n,d}(p, q) \leq 0$. Since p has a finite number of roots, we can find a constant R for which all of the roots of p lie in the interval $[0, R]$, and then consider the collection of all polynomials in \mathcal{P}_j^+ whose roots lie in the interval $[0, R]$. This collection is compact, and so $\phi_\alpha^{n,d}(\cdot, q)$ (a continuous function in the roots of p) achieves its minimum. Let p_0 be a polynomial achieving this minimum; in particular, note that $\phi_\alpha^{n,d}(p_0, q) \leq \phi_\alpha^{n,d}(p, q) \leq 0$, which means (by hypothesis (ii)) p_0 must be nonbasic.

Since p_0 is nonbasic, it has a decomposition $p_0 = \tilde{p}_0 + \hat{p}_0$ with the properties of Lemma 4.4. In particular, note that

- a. implies (by the linearity of \boxplus_d^n) that $[p_0 \boxplus_d^n q] = [\tilde{p}_0 \boxplus_d^n q] + [\hat{p}_0 \boxplus_d^n q]$
- b. implies that \tilde{p}_0 has all of its roots in $[0, R]$, so the choice of p_0 as a minimizer ensures that $\phi_\alpha^{n,d}(p_0, q) \leq \phi_\alpha^{n,d}(\tilde{p}_0, q)$.
- c. implies (by plugging in to the definition of $\phi_\alpha^{n,d}(\cdot, \cdot)$)

$$\phi_\alpha^{n,d}(p_0, q) + \Theta_\alpha^n(p_0 \boxplus_d^n q) = \phi_\alpha^{n,d}(\tilde{p}_0, q) + \Theta_\alpha^n(\tilde{p}_0 \boxplus_d^n q) = \phi_\alpha^{n,d}(\hat{p}_0, q) + \Theta_\alpha^n(\hat{p}_0 \boxplus_d^n q).$$

Hence Properties b. and c. combine to give $\Theta_\alpha^n(\tilde{p}_0 \boxplus_d^n q) \leq \Theta_\alpha^n(p_0 \boxplus_d^n q)$ and so

$$\text{maxroot}(W_\alpha^n[\tilde{p}_0 \boxplus_d^n q]) \leq \text{maxroot}(W_\alpha^n[p_0 \boxplus_d^n q]). \tag{4.6}$$

Now let $\beta = \text{maxroot}(W_\alpha^n[p_0 \boxplus_d^n q])$ and assume (for the moment) that

$$\beta^2 \geq \text{maxroot}([p_0 \boxplus_d^n q] \times [\tilde{p}_0 \boxplus_d^n q] \times [\hat{p}_0 \boxplus_d^n q]). \tag{4.7}$$

This would imply that the decomposition

$$[p_0 \boxplus_d^n q] = [\tilde{p}_0 \boxplus_d^n q] + [\hat{p}_0 \boxplus_d^n q]$$

satisfies the requirements of Corollary 3.9, which would allow us to extend (4.6) to

$$\text{maxroot}(W_\alpha^n[\tilde{p}_0 \boxplus_d^n q]) \leq \text{maxroot}(W_\alpha^n[p_0 \boxplus_d^n q]) \leq \text{maxroot}(W_\alpha^n[\hat{p}_0 \boxplus_d^n q]).$$

Using Property c. again, this would then imply that $\phi_\alpha^{n,d}(\hat{p}, q) \leq \phi_\alpha^{n,d}(p_0, q) < 0$, a contradiction to the initial hypothesis (since $\hat{p} \in \mathcal{P}_{j-1}^+$).

Hence it suffices to prove (4.7). First we note that Corollary 3.6 easily gives

$$\beta^2 = \text{maxroot}(W_\alpha^n[p_0 \boxplus_d^n q])^2 \geq \text{maxroot}(W_0^n[p_0 \boxplus_d^n q])^2 = \text{maxroot}([p_0 \boxplus_d^n q])$$

and that (4.6) combined with the same Corollary gives

$$\beta^2 \geq \text{maxroot}(W_\alpha^n[\tilde{p}_0 \boxplus_d^n q])^2 \geq \text{maxroot}(W_0^n[\tilde{p}_0 \boxplus_d^n q])^2 = \text{maxroot}([\tilde{p}_0 \boxplus_d^n q]).$$

Finally the fact that $\beta^2 \geq \text{maxroot}([\hat{p}_0 \boxplus_d^n q])$ comes directly from Property d. in Theorem 4.4. □

5. Base Cases

Our proof of Theorem 1.3 is inductive and utilizes both Lemma 4.1 and Lemma 4.2. Each of these lemmas will require its own “base case”.

1. $p \in \mathcal{P}_d^+$ and $q(x) = x^{d-1}$ (Corollary 5.3), and
2. $p, q \in \mathcal{P}_d^+$ are both basic polynomials (Lemma 5.8).

Neither of these lemmas is particularly simple (despite being the “base cases” of an induction). We prove Corollary 5.3 using a separate induction, first considering the case when p is a basic polynomial (Lemma 5.2) and then using Lemma 4.2 to extend this to all polynomials. While the proof of Lemma 5.2 is mostly calculus, the functions that one needs to consider become complicated enough that we were forced to appeal to the aid of a computer in order to calculate them. Section 5.1 is dedicated to this case.

Lemma 5.8, on the other hand, will require an entire investigation of its own. Corollary 5.6 will relate the rectangular additive convolution of two basic polynomials a well-studied class of orthogonal polynomials and we will utilize a number of well-known properties of these polynomials to prove the necessary inequalities. To simplify the presentation, the proof of Lemma 5.8 given in Section 5.2 will be contingent upon a bound on the Cauchy transform of Gegenbauer polynomials that will be proved in Section 6.

Assuming these two base cases, the proof of Theorem 1.3 is then straightforward:

Theorem 1.3. For $n \geq 0, \alpha > 0, d \geq 1$ and $p, q \in \mathcal{P}_{\leq d}^+$, we have $\phi_\alpha^{n,d}(p, q) > 0$.

Proof. We proceed by induction on d . The base case (when $d = 1$) is covered by Lemma 5.8, since degree 1 polynomials are (by definition) basic.

Now assume that, for some $D \geq 2$ the theorem holds for all $d < D$ and consider $p, q \in \mathcal{P}_{\leq D}^+$. If $\deg(p) < D$ or $\deg(q) < D$, we can appeal to the inductive hypothesis using Lemma 4.1 and Corollary 5.3. On the other hand, if $\deg(p) = \deg(q) = D$, we can appeal to the inductive hypothesis using Lemma 4.2 and Lemma 5.8. \square

In order to make the presentation of the two lemmas more readable, we first show that the parameter α can be effectively scaled out of statements pertaining to $\phi_\alpha^{n,d}(p, q)$. Note that the transformation $p \rightarrow p_\alpha$ in Lemma 5.1 preserves basic polynomials, so statements that are restricted to basic polynomials can be scaled out as well.

Lemma 5.1. For a fixed $\alpha > 0$ and polynomials p, q with nonnegative roots, let $p_\alpha(x) = p(\alpha^2 x)$ and $q_\alpha(x) = q(\alpha^2 x)$. Then

$$\phi_\alpha^{n,d}(p, q) = \alpha \phi_1^{n,d}(p_\alpha, q_\alpha)$$

Proof. By (4.1), it suffices to show

$$\maxroot(W_\alpha^n p) = \alpha \cdot \maxroot(W_1^n p_\alpha)$$

for arbitrary p . Computing, we have

$$\begin{aligned} W_1^n p_\alpha(x) &= x^{2n} p_\alpha(x^2)^2 - 1[x^{2n} p_\alpha(x^2)]'[r(x^2)]' \\ &= x^{2n} p(\alpha^2 x^2)^2 - (2nx^{2n-1} p(\alpha^2 x^2) + 2x^{2n+1} \alpha^2 p'(\alpha^2 x^2)) (2\alpha^2 x p'(\alpha^2 x^2)) \\ &= x^{2n} (p(\alpha^2 x^2)^2 - 4n\alpha^2 p(\alpha^2 x^2) p'(\alpha^2 x^2) - 4x^2 \alpha^4 p'(\alpha^2 x^2)^2) \end{aligned}$$

so

$$\begin{aligned} \maxroot(W_1^n p_\alpha) &= \maxroot(p(\alpha^2 x^2)^2 - 4n\alpha^2 p(\alpha^2 x^2) p'(\alpha^2 x^2) - 4x^2 \alpha^4 p'(\alpha^2 x^2)^2) \\ &= \frac{1}{\alpha} \cdot \maxroot(p(x^2)^2 - 4n\alpha^2 p(x^2) p'(x^2) - 4x^2 \alpha^2 p'(x^2)^2) \\ &= \frac{1}{\alpha} \cdot \maxroot(W_\alpha^n p). \end{aligned} \quad \square$$

5.1. The case of $q(x) = x^{d-1}$

As mentioned earlier, we start by proving the case when p is a basic polynomial.

Lemma 5.2. *Let $d \geq 2$, $n \geq 0$, $\alpha > 0$ and let $p \in \mathcal{P}_d^+$ be a basic polynomial. Then $\phi_\alpha^{n,d}(p, x^{d-1}) > 0$.*

Proof. By Lemma 5.1, it suffices to prove the lemma when $\alpha = 1$. To simplify notation slightly, we will sometimes write $m = n + d$. Note that for $p(x) = (x - \lambda)^d$ and $q(x) = x^{d-1}$, Lemma 4.3 implies that

$$\Theta_1^n(p) = \sqrt{m^2 + \lambda} + d \quad \text{and} \quad \Theta_1^n(q) = n - 2.$$

Letting

$$\Delta[p] = [p \boxplus_d^n x^{d-1}] = (x - \lambda)^{d-2} \left(x - \frac{n+1}{m} \lambda \right)$$

it then suffices to show that

$$\sqrt{n^2 + \maxroot(W_1^n \Delta[p])^2} < \sqrt{m^2 + \lambda} + d - 2$$

or, squaring both sides and rearranging,

$$\maxroot(W_1^n \Delta[p])^2 < \left(2(d-1) - m + \sqrt{m^2 + \lambda} \right) \left(m - 2 + \sqrt{m^2 + \lambda} \right) := \mu_\lambda \quad (5.1)$$

Furthermore, we can calculate

$$W_1^n \Delta[p](x) = (x^2 - \lambda)^{2(d-3)} x^{2n} s_\lambda(x^2)$$

where $s_\lambda \in \mathcal{P}_4^+$ has coefficients that depend on n, d, λ . Hence

$$\maxroot(W_1^n \Delta[p](x))^2 = \maxroot(s_\lambda),$$

and so (5.1) is equivalent to having $\maxroot(s_\lambda) < \mu_\lambda$. Our approach to proving this will be to show that $s_\lambda(\mu_\lambda + \epsilon) > 0$ for all $\epsilon \geq 0$. Consider the Taylor series of the function

$$f(\lambda) = s_\lambda(\mu_\lambda).$$

around $\lambda = 0$. With the help of a computer, one can calculate that $f(0) = f'(0) = 0$ and that

$$f^{(2)}(0) = \frac{32(d-1)^2(m-1)(n+1)(m+d-2)}{m^3} > 0$$

and

$$f^{(3)}(\lambda) = \frac{24(n+1)(d-2)g(\lambda)}{m^2(m^2 + \lambda)^{\frac{5}{2}}}$$

where

$$g(\lambda) = 2(d-1)(m^4 - m^3) + m^2 \left((2d-1)^2 + 3\lambda d - 4\lambda \right) + \frac{5}{4}(d-1)\lambda^2 + (d-1)m\lambda.$$

So for $d \geq 2$, we have $(2d - 1)^2 + 3\lambda d - 4\lambda \geq 0$, and for $m \geq d$, all other terms in $f^{(3)}(\lambda)$ are trivially nonnegative. Hence $f(\lambda) > 0$ for all $\lambda > 0$.

Now note that, because the leading coefficient of s_λ is positive, the fact that $s_\lambda(\mu_\lambda) > 0$ implies that the number of roots of s_λ that are larger than μ_λ is even. On the other hand, when $\lambda = 0$,

$$s_0(x) = x^3(x - 4(m - 1)(d - 1))$$

and so has three roots at $x = 0$ and one at $\mu_0 = 4(m - 1)(d - 1) > 0$. As s_λ is real rooted for all $\lambda \geq 0$ and the roots of s_λ are continuous functions of its coefficients (and thus of λ) we can conclude that for small λ all but one of the roots of s_λ must be near 0; in other words, the function

$$g(\lambda) = \text{maxroot}(s_\lambda) - \mu_\lambda$$

is positive for sufficiently small λ . Hence it suffices to show that $g(\lambda) \neq 0$ for any $\lambda > 0$.

Assume (for contradiction) that there exists $\lambda_0 > 0$ for which $g(\lambda_0) = 0$. In other words, $\text{maxroot}(s_{\lambda_0}) = \mu_{\lambda_0}$ which (in particular) means that $s_{\lambda_0}(\mu_{\lambda_0}) = 0$. But this is a contradiction, since we have shown $f(\lambda) = s_\lambda(\mu_\lambda)$ to be strictly positive for $\lambda > 0$. Hence $g(\lambda)$ must remain strictly positive for all $\lambda > 0$, finishing the proof. \square

Using Lemma 4.2, we can then extend Lemma 5.2 to all polynomials.

Corollary 5.3. *Let $2 \leq j \leq d$, $n \geq 0$, $\alpha > 0$ and let $p \in \mathcal{P}_j^+$ be an (arbitrary) polynomial. Then $\phi_\alpha^{n,d}(p, x^{d-1}) > 0$.*

Proof. Let $S(j, d)$ be the statement

$$\phi_\alpha^{n,d}(p, x^{d-1}) > 0 \text{ for all } n \geq 0, \alpha > 0 \text{ and all } p \in \mathcal{P}_j^+.$$

Our goal is then to prove $S(j, d)$ for all $2 \leq j \leq d$, and we will proceed by induction on $j + d$. The base case is when $j = d = 2$, which we will consider below. Now assume $S(j, d)$ to hold whenever $j + d < K$ and consider a polynomial p with degree $j \leq d$ for which $j + d = K$. We split into two cases:

1. $j < d$:

Since both p and x^{d-1} have degree less than d , we can apply Lemma 2.2 to get

$$[p \boxplus_d^n x^{d-1}] = \frac{1}{d(d+n)} [p \boxplus_{d-1}^n [x^{d-1} \boxplus_d^n x^{d-1}]] = c_{d,n} [p \boxplus_{d-1}^n x^{d-2}]$$

for some constant $c_{d,n}$. That is, $S(j, d-1) \Rightarrow S(j, d)$.

2. $j = d$:

A combination of Lemma 4.2 and Lemma 5.2 gives that

$$S(d-1, d) \Rightarrow S(d, d)$$

Hence in both cases we have reduced the statement $S(j, d)$ to one that we know (by the inductive hypothesis) is true, so $S(j, d)$ is true as well.

To finish the proof, we then need to consider only the base case: $j = d = 2$. For $t \in [0, a]$, set

$$p_t(x) = (x - a)^2 - t^2 = (x - a - t)(x - a + t)$$

and note that

$$[p_t \boxplus_2^n x] = x - a \frac{n + 1}{n + 2}$$

is independent of t . Hence we have

$$\phi_\alpha^{n,1}(p_t, x) = \Theta_\alpha^n(p_t) - \Theta_\alpha^n(p_0) + \phi_\alpha^{n,1}(p_0, x)$$

where $\phi_\alpha^{n,1}(p_0, x) > 0$ by Lemma 5.2, and so it suffices to show that

$$\maxroot(W_\alpha^n p_t) \geq \maxroot(W_\alpha^n p_0). \tag{5.2}$$

By Corollary 3.4 and Lemma 3.7, this is equivalent to showing that

$$[\log p_t]'(x) \leq [\log p_0]'(x).$$

for all $x > \maxroot(p_0 p_t) = a + t$. However, it is easy to check that

$$\frac{\partial}{\partial t} [\log p_t]'(x) = \frac{4t(a - x)}{(x - a - t)^2(x - a + t)^2},$$

which is negative for $x > a + t$. □

5.2. The case of basic polynomials

We start by finding a generating function for the rectangular convolution of two basic polynomials. The derivation will use the following well known generalization of the *binomial theorem* (see, for example, [24]).

Theorem 5.4. *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 5.5. *For all $\lambda, \mu > 0$ and $n \geq 0$, the polynomials*

$$q_{n,d}^{\lambda,\mu}(y) = \binom{n + d}{d} [(x - \lambda)^d \boxplus_d^n (x - \mu)^d](y).$$

satisfy the formal power series identity

$$\sum_d q_{n,d}^{\lambda,\mu}(y) t^d = \frac{1}{((1 + \mu t)(1 + \lambda t) - yt)^{n+1}}.$$

Proof. Using (2.2), we can write $q_{n,d}^{\lambda,\mu}$ explicitly. Letting

$$c(s, i, j, \ell) = \frac{(s-i)!(s-j)!}{s!(s-\ell)!}$$

we have

$$\begin{aligned} q_{n,d}^{\lambda,\mu}(y) &= \binom{n+d}{d} \sum_{\ell} y^{d-\ell} \left(\sum_{i+j=\ell} c(d, i, j, \ell) c(n+d, i, j, \ell) \binom{d}{i} \binom{d}{j} (-\mu)^i (-\lambda)^j \right) \\ &= \sum_{i+j+\ell=d} \binom{n+\ell}{\ell} \binom{n+j+\ell}{j} \binom{n+i+\ell}{i} (-\mu)^i (-\lambda)^j y^{\ell}. \end{aligned}$$

Hence we have the formal power series identity

$$\begin{aligned} \sum_d q_{n,d}^{\lambda,\mu}(y) t^d &= \sum_{i,j,\ell} \binom{n+\ell}{\ell} \binom{n+j+\ell}{j} \binom{n+i+\ell}{i} (-\mu)^i (-\lambda)^j y^{\ell} t^{i+j+\ell} \\ &= \sum_{\ell} (1+\mu t)^{-(n+\ell+1)} (1+\lambda t)^{-(n+\ell+1)} \binom{n+\ell}{\ell} y^{\ell} t^{\ell} \quad (\text{Theorem 5.4 on } i, j) \\ &= ((1+\mu t)(1+\lambda t))^{-(n+1)} \sum_{\ell} \binom{n+\ell}{\ell} \left(\frac{yt}{(1+\mu t)(1+\lambda t)} \right)^{\ell} \\ &= ((1+\mu t)(1+\lambda t) - yt)^{-(n+1)} \quad (\text{Theorem 5.4 on } \ell). \end{aligned}$$

□

This provides an easy link between the rectangular convolution of basic polynomials and the Gegenbauer polynomials studied in Section 6:

Corollary 5.6. *For all $\lambda, \mu > 0, n \geq 0, d \geq 1$ we have*

$$\binom{n+d}{d} [(x-\lambda)^d \boxplus_d^n (x-\mu)^d](y) = (\lambda\mu)^{d/2} C_d^{n+1} \left(\frac{y - (\lambda + \mu)}{2\sqrt{\lambda\mu}} \right).$$

Proof. Compare Lemma 5.5 to (6.1). □

Bounding the H -transform of the convolution of two basic polynomials can therefore be reduced to bounding the Cauchy transform of Gegenbauer polynomials. We prove the necessary bound in Section 6, but reproduce the theorem here for continuity.

Theorem 6.11. Consider the bivariate polynomial

$$f(x, y) = d(x^2 - 1)y^2 + 2nxy - (2n + d)$$

and assume that for a given $s > \gamma_{n/d}$ and $t > 0$ that $f(s, t) \geq 0$. Then

$$\mathcal{G}_{C_d^{n+1}}(s) = \frac{[C_d^{n+1}]'(s)}{d[C_d^{n+1}](s)} \leq t.$$

The proof of Lemma 5.8 will require a number of identities that are not, by themselves, important to understanding the overall proof. In order to keep the continuity of ideas in the proof, we have separated these identities out into a separate lemma:

Lemma 5.7. For $\lambda, \mu, n, d \in \mathbb{R}$, let

$$t := \sqrt{(n+d)^2 + \lambda} + \sqrt{(n+d)^2 + \mu} \quad \text{and} \quad t_* := \sqrt{t^2 - 2tn}$$

and also let

$$T := \frac{t_*^2 - (\lambda + \mu)}{2\sqrt{\lambda\mu}} \quad \text{and} \quad R := \frac{\sqrt{\lambda\mu}}{dt} \quad \text{and} \quad \gamma_{n/d} := \sqrt{1 - \frac{n^2}{(n+d)^2}}.$$

Then the following identities hold:

1. $d(T^2 - 1)R^2 + 2nTR - (2n + d) = 0.$
2. $(dR(T^2 - 1) + nT)^2 = (n + d)^2(T^2 - \gamma_{n/d}^2)$

Proof. For real numbers a, b , it is easy to check that $\sqrt{a} + \sqrt{b}$ is a root of the polynomial $x^4 - 2(a + b)x^2 + (a - b)^2$ by simple substitution. Hence

$$t^4 - 2(\lambda + \mu + 2(n + d)^2)t^2 + (\lambda - \mu)^2 = 0,$$

or (rearranging slightly),

$$(t^2 - (\lambda + \mu))^2 = 4(n + d)^2t^2 + 4\lambda\mu = 4t^2((n + d)^2 + d^2R^2) \tag{5.3}$$

By definition of T , however, we have

$$(t^2 - (\lambda + \mu))^2 = \left(2\sqrt{\lambda\mu}T + 2tn\right)^2 = (2dtRT + 2tn)^2 = 4t^2(d^2R^2T^2 + 2dnRT + n^2) \tag{5.4}$$

Equating (5.3) and (5.4) and then rearranging gives

$$d^2R^2(T^2 - 1) + 2ndRT = ((n + d)^2 - n^2) = 2nd + d^2, \tag{5.5}$$

which clearly implies 1. Now note that if we multiply out 2., we get

$$\begin{aligned} d^2R^2(T^2 - 1)^2 + 2dR(T^2 - 1)nT + n^2T^2 &= (n + d)^2 \left(T^2 - \left(1 - \frac{n^2}{(n + d)^2} \right) \right) \\ &= (n + d)^2(T^2 - 1) + n^2T^2 \quad \square \end{aligned}$$

Finally, we are able to prove the lemma:

Lemma 5.8. Let $d \geq 1$, $n \geq 0$, $\alpha > 0$ and let $p, q \in \mathcal{P}_d^+$ be basic polynomials. Then $\phi_\alpha^{n,d}(p, q) > 0.$

Proof. Again by Lemma 5.1 it suffices to consider the case $\alpha = 1$. We set

$$p(x) = (x - \lambda)^d \quad \text{and} \quad q(x) = (x - \mu)^d \quad \text{and} \quad r(x) = [p \boxplus_d^n q](x).$$

and note that Corollary 5.6 and Theorem 6.11 show that

$$\frac{\text{maxroot}(r) - (\lambda + \mu)}{2\sqrt{\lambda\mu}} \leq \gamma_{n/d} = \sqrt{1 - \frac{n^2}{(n+d)^2}}. \quad (5.6)$$

Also set $w_r = \text{maxroot}(W_1^n r)$; substituting this and Lemma 4.3 into the definition of $\phi_1^{n,d}(\cdot, \cdot)$, we get that it suffices to show

$$\sqrt{n^2 + w_r^2} \leq \sqrt{(n+d)^2 + \lambda} + \sqrt{(n+d)^2 + \mu} - n. \quad (5.7)$$

Next define the quantities

$$t := \sqrt{(n+d)^2 + \lambda} + \sqrt{(n+d)^2 + \mu} \quad \text{and} \quad t_* := \sqrt{t^2 - 2tn},$$

noting that the string of inferences

$$\lambda, \mu > 0 \Rightarrow t \geq 2(n+d) \Rightarrow t(t-2n) \geq 2dt \geq 0$$

show that t^* is well defined. Then to prove (5.7), it suffices to show (after squaring both sides, and substituting) the inequality $w_r \leq t_*$.

Finally, define

$$W := \frac{w_r^2 - \lambda - \mu}{2\sqrt{\lambda\mu}} \quad \text{and} \quad T := \frac{t_*^2 - \lambda - \mu}{2\sqrt{\lambda\mu}}$$

and notice that the identity in Lemma 5.7.2 implies that $T \geq \gamma_{n/d}$. Then if $W \leq \gamma_{n/d}$, we are done, so we are left to consider the case when $W \geq \gamma_{n/d}$.

For $W, T \geq \gamma_{n/d}$, we then have by (5.6) that

$$w_r \leq \sqrt{\text{maxroot}(r)} \quad \text{and} \quad t_* \leq \sqrt{\text{maxroot}(r)}$$

so Lemma 3.5 implies that the inequality $w_r \leq t_*$ holds if and only if

$$\mathcal{H}_r^n(t_*) \leq \mathcal{H}_r^n(w_r) = 1. \quad (5.8)$$

By definition, we have

$$\mathcal{H}_r^n(x) = 2dx\mathcal{G}_r(x^2) \left(\frac{2n}{x} + 2dx\mathcal{G}_r(x^2) \right) = 4d \left(n\mathcal{G}_r(x^2) + dx^2\mathcal{G}_r(x^2)^2 \right)$$

and so we can rewrite (5.8) as

$$dt_*^2\mathcal{G}_r(t_*^2) + n\mathcal{G}_r(t_*^2) \leq \frac{1}{4d}.$$

We now bound $\mathcal{G}_r(t_*^2)$ by noting that the identity in Lemma 5.7.1 allows us to apply Theorem 6.11 to the ordered pair

$$(s, t) = \left(\mathcal{G}_{C_d^{(n+1)}}(T), \frac{\sqrt{\lambda\mu}}{dt} \right).$$

By Corollary 5.6, we then get the inequality

$$2\sqrt{\lambda\mu}\mathcal{G}_r(t_*^2) = \mathcal{G}_{C_d^{(n+1)}}(T) \leq \frac{\sqrt{\lambda\mu}}{dt} \implies \mathcal{G}_r(t_*^2) \leq \frac{1}{2dt}.$$

Hence

$$dt_*^2 \mathcal{G}_r(t_*^2)^2 + n\mathcal{G}_r(t_*^2) \leq \frac{t_*^2}{4dt^2} + \frac{n}{2dt} = \frac{t_*^2 + 2nt}{4dt^2} = \frac{1}{4d}$$

as required. □

6. Gegenbauer polynomials

The Gegenbauer (or ultraspherical) polynomials $C_n^{(\alpha)}(x)$ are a collection of polynomials which are orthogonal with respect to the weight function $w(x) = (1 - x^2)^{\alpha-1/2}$ on the interval $[-1, 1]$. They can be computed explicitly using a generating function

$$\sum_n C_n^{(\alpha)}(x)t^n = \frac{1}{(1 - 2xt + t^2)^\alpha} \tag{6.1}$$

or by the three-term recurrence given in Lemma 6.1. They are a special case of the more general *Jacobi polynomials*:

$$C_d^{(\alpha)}(x) = \frac{\Gamma(\alpha + 1/2)\Gamma(d + 2\alpha)}{\Gamma(2\alpha)\Gamma(d + \alpha + 1/2)} P_d^{(\alpha-1/2, \alpha-1/2)}(x)$$

and are themselves a generalization of two other well-studied families of orthogonal polynomials:

1. Chebyshev polynomials of the second kind ($\alpha = 1$), and
2. Legendre polynomials ($\alpha = 1/2$).

We will use the following identities, which can be derived directly from (6.1) or taken as a specialization of known identities for Jacobi polynomials (see, for example, [21]). Note that the $\Gamma(\cdot)$ appearing in the second identity is the usual *Gamma function*

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt.$$

Lemma 6.1. *The following hold for all real numbers $\alpha > 0$ and all integers $d > 0$.*

1. *Recurrence relation, with convexity coefficients:*

$$xC_d^{(\alpha)}(x) = \frac{d + 1}{2d + 2\alpha} C_{d+1}^{(\alpha)}(x) + \frac{d + 2\alpha - 1}{2d + 2\alpha} C_{d-1}^{(\alpha)}(x)$$

2. *Value(s) at 1:*

$$C_d^{(\alpha)}(1) = \binom{d+2\alpha-1}{d} = \frac{\Gamma(d+2\alpha)}{\Gamma(2\alpha)\Gamma(d+1)} \quad \text{and} \quad \mathcal{G}_{C_d^{(\alpha)}}(1) = \frac{d+2\alpha}{2\alpha+1}$$

3. *Differential equation:*

$$(1-x^2)DC_d^{(\alpha)}(x) = -dxC_d^{(\alpha)}(x) + (d+2\alpha-1)C_{d-1}^{(\alpha)}(x)$$

Our goal will be to derive bounds on the Cauchy transform of $C_d^{(\alpha)}$ at values greater than its roots. Fortunately, much is known about the largest roots of Gegenbauer polynomials; in particular, the following bound from [5] provides a useful starting point:

Lemma 6.2 (Elbert, Laforgia). *For fixed $\alpha > 0$, the roots of $C_d^{(\alpha)}$ lie in the interval $[-\gamma_{\alpha/d}, \gamma_{\alpha/d}]$ where*

$$\gamma_{\theta} := \frac{\sqrt{2\theta+1}}{\theta+1}$$

Remark 6.3. Note that γ_{θ} is decreasing in θ , and so decreasing the degree d or increasing the parameter α will cause the interval to shrink. This coincides with the well known fact that all of the positive roots of $C_d^{(\alpha)}$ are increasing as d grows and decreasing as α grows [21].

The bound we will derive differs from Lemma 6.2 (and all other known bounds, as far as we can tell) in that it requires us to compare the largest roots of $C_d^{(\alpha)}$ as d and α grow together in a linear way. Specifically, we will be interested in understanding the largest root of $C_d^{(1+d\theta)}$ for a generic constant θ . The intuition above tells us that the growth in d and the growth in $\alpha = 1 + d\theta$ should push the roots in opposite directions, and so it should not be surprising that the computations become somewhat delicate.

It should be noted that the asymptotic behavior of these polynomials has been studied, and it is well known that the density of the roots of these polynomials converges to a fixed distribution that will depend on the parameter θ . The distribution can be computed using a result of Kuijlaars and Van Assche regarding the asymptotic root distributions of Jacobi polynomials [9]. The following is a reformulated version of the theorem specific to our polynomials:

Theorem 6.4 (Kuijlaars, Van Assche). *For all $\theta > 0$, the density of the roots of $C_d^{(1+d\theta)}$ converges (as $d \rightarrow \infty$) to the function*

$$\mu_{\theta}(y) = \frac{1+\theta}{\pi} \frac{\sqrt{\gamma_{\theta}^2 - y^2}}{1-y^2} \mathbf{1}_{[-\gamma_{\theta}, \gamma_{\theta}]}(y).$$

An asymptotic bound on the Cauchy transform can then be derived from Theorem 6.4:

Corollary 6.5. *For all $\theta > 0$, and all $x > \gamma_{\theta}$, we have*

$$\lim_{d \rightarrow \infty} \mathcal{G}_{C_d^{(1+d\theta)}}(x) = \frac{-\theta x + (1+\theta)\sqrt{x^2 - \gamma_{\theta}^2}}{(x^2 - 1)}. \quad (6.2)$$

We give a brief sketch of the computation in an Appendix for the benefit of the reader. Note that for the specific case of $\theta = 0$, the formula in Theorem 6.4 simplifies greatly to a well-known bound on the Cauchy transform of Chebyshev polynomials:

$$\lim_{d \rightarrow \infty} \mathcal{G}_{C_d^{(1)}}(x) = \frac{1}{\sqrt{x^2 - 1}}$$

for all $x > 1$.

The rest of this section will be devoted to showing that the bound in Corollary 6.5 holds for the individual polynomials $C_d^{(1+\theta d)}$ as well. In particular, we will show that the sequence $\{\mathcal{G}_{C_d^{(1+\theta d)}}(x)\}_d$ is increasing in d for all $\theta > 0$ and all $x > \gamma_\theta^d$ where we define

$$\gamma_\theta^d := \text{maxroot} \left(C_d^{(1+\theta d)}(x) \right).$$

One part of this will be to show that the sequence $\{\gamma_\theta^d\}$ is an increasing function of d (for fixed θ). For the moment, however, we will find it convenient to consider the following “two-step” maximum root:

$$\Gamma_\theta^d := \max \left\{ \gamma_\theta^d, \gamma_\theta^{(d+1)} \right\}.$$

6.1. Nonasymptotic bounds

To simplify notation slightly, we will fix θ and normalize the polynomials of interest (in a manner that is common when deriving such inequalities — see [2], for example). For j, k nonnegative integers, we will set

$$p_{j,k}(x) = \frac{C_j^{(1+\theta k)}(x)}{C_j^{(1+\theta k)}(1)}.$$

Definition 6.6. We will say that a polynomial p is β -orthogonal-fit if

1. $p(x) < 0$ in $(\beta, 1)$, and
2. $p(x) > 0$ in $(1, \infty)$.

It is easy to check that β -orthogonal-fitness is closed under nonnegative linear combinations (a fact that will be used in Lemma 6.8) and that Remark 6.3 implies that any polynomial which is $\Gamma_{\theta_1}^d$ -orthogonal-fit is also $\Gamma_{\theta_2}^d$ -orthogonal-fit whenever $\theta_2 < \theta_1$.

The following lemma reduces the monotonicity statement we are interested in to one regarding certain polynomials being β -orthogonal-fit.

Lemma 6.7. *The following statements are equivalent:*

1. For any fixed $x > \Gamma_\theta^d$, we have

$$\mathcal{G}_{C_d^{(1+\theta d)}}(x) \leq \mathcal{G}_{C_{d+1}^{(1+\theta(d+1))}}(x).$$

2. The polynomial

$$\Delta_d(x) = p_{d+1,d+1}(x)p_{d-1,d}(x) - p_{d,d}(x)p_{d,d+1}(x) \quad (6.3)$$

is Γ_θ^d -orthogonal-fit.

Proof. We start by rewriting the first statement as

$$\frac{1}{d} \frac{DC_d^{(1+\theta d)}(x)}{C_d^{(1+\theta d)}(x)} < \frac{1}{d+1} \frac{DC_{d+1}^{(1+\theta(d+1))}(x)}{C_{d+1}^{(1+\theta(d+1))}(x)}.$$

and notice that the third identity in Lemma 6.1 implies that

$$DC_d^{(1+\theta d)}(x) = -dx C_d^{(1+\theta d)}(x) + (d-1+2(1+d\theta))C_{d-1}^{(1+\theta d)}(x)$$

and

$$DC_{d+1}^{(1+\theta(d+1))}(x) = -(d+1)x C_{d+1}^{(1+\theta(d+1))}(x) + \left(d+2(1+(d+1)\theta)\right)C_d^{(1+\theta(d+1))}(x).$$

Plugging these in and canceling the like terms, we get the equivalent statement

$$\frac{(d+2d\theta+1)}{d(1-x^2)} \frac{C_{d-1}^{(1+\theta d)}(x)}{C_d^{(1+\theta d)}(x)} < \frac{(d+2d\theta+2\theta+2)}{(d+1)(1-x^2)} \frac{C_d^{(1+\theta(d+1))}(x)}{C_{d+1}^{(1+\theta(d+1))}(x)}.$$

Plugging in the second identity in Lemma 6.1 and simplifying, we see that this is equivalent to having $\Delta_d(x) < 0$ on the interval $(\Gamma_\theta^d, 1)$ and $\Delta_d(x) > 0$ on the interval $(1, \infty)$, which is the definition of being Γ_θ^d -orthogonal-fit. \square

Before proving anything with the polynomials $p_{j,k}$, it will be worthwhile to note the translation of the first identity in Lemma 6.1 into these polynomials:

$$xp_{j,k} = \lambda_{j,k}p_{j+1,k} + (1-\lambda_{j,k})p_{j-1,k} \quad (6.4)$$

where (for our fixed θ)

$$0 \leq \lambda_{j,k} = \frac{j+2k\theta+2}{2j+2k\theta+2} \leq 1.$$

We will use the following fact (that comes from direct calculation):

$$\lambda_{j,k} - \lambda_{j-1,k-1} = \frac{(j-k)\theta - 1}{2(1+j+k\theta)(j+(k-1)\theta)}. \quad (6.5)$$

Lemma 6.8. *The polynomial*

$$\Delta_{j,d}(x) = p_{j+1,d+1}(x)p_{j-1,d}(x) - p_{j,d}(x)p_{j,d+1}(x) \quad (6.6)$$

is Γ_θ^d -orthogonal-fit for all $1 \leq j \leq d$.

Proof. The fact that $x = 1$ is a root follows from the definition of $p_{j,k}$, and so it would suffice to show that $\Delta_{j,d}$ has at most one root in the desired interval, a fact that we will prove by induction (on j). The base case can be computed explicitly:

$$\Delta_{1,d} = \frac{x^2 - 1}{3 + 2(d + 1)\theta}$$

which is clearly Γ_θ^d -orthogonal-fit. For the inductive step, let $s = \lambda_{j,d+1}$ and $t = \lambda_{j-1,d}$. By (6.4), we have

$$xp_{j,d+1} = sp_{j+1,d+1} + (1 - s)p_{j-1,d+1} \quad \text{and} \quad xp_{j-1,d} = tp_{j,d} + (1 - t)p_{j-2,d}$$

and so

$$tp_{j,d+1}p_{j,d} + (1 - t)p_{j,d+1}p_{j-2,d} = sp_{j-1,d}p_{j+1,d+1} + (1 - s)p_{j-1,d}p_{j-1,d+1}.$$

Plugging in

$$p_{j+1,d+1}p_{j-1,d} = \Delta_{j,d} + p_{j,d}p_{j,d+1} \quad \text{and} \quad p_{j,d+1}p_{j-2,d} = \Delta_{j-1,d} + p_{j-1,d}p_{j-1,d+1}$$

then gives

$$s\Delta_{j,d} = (1 - t)\Delta_{j-1,d} + (t - s)(p_{j,d+1}p_{j,d} - p_{j-1,d}p_{j-1,d+1})$$

The induction hypothesis gives that $\Delta_{j-1,d}$ is Γ_θ^d -orthogonal-fit and (6.5) shows that $t - s > 0$. Since θ -orthogonal-fitness is closed under nonnegative combinations, it would then suffice to show that the polynomial $p_{j,d+1}p_{j,d} - p_{j-1,d+1}p_{j-1,d}$ is Γ_θ^d -orthogonal-fit for any $j \leq d$. Again, $x = 1$ is obviously a root, and so the theorem would follow by showing that there is at most one real root in the desired interval.

However, this follows easily from interlacing properties: it is well known that consecutive orthogonal polynomial have interlacing roots, and one can show that a polynomial p interlaces a polynomial q if and only if there exist nonnegative constants λ_i such that

$$\frac{p}{q} = \sum_{i=1}^d \frac{\lambda_i}{x - r_i} \quad \text{and}$$

where the r_i are the roots of q (see, for example, [23]). In particular, this ratio is nonnegative and monotone decreasing at any $x > \max\text{root}(q)$. Hence $p_{j-1,d}$ interlaces $p_{j,d}$ and $p_{j-1,d+1}$ interlaces $p_{j,d+1}$ and the product

$$\frac{p_{j-1,d} p_{j-1,d+1}}{p_{j,d} p_{j,d+1}}$$

is nonnegative and monotone decreasing for $x > \max\text{root}(p_{j,d})$ (recall that the monotonicity mentioned in Remark 6.3 implies that $p_{d,k}$ has the largest root of these polynomials). Hence

$$(p_{j,d+1}p_{j,d} - p_{j-1,d}p_{j-1,d+1}) = p_{j,d+1}p_{j,d} \left(1 - \frac{p_{j-1,d} p_{j-1,d+1}}{p_{j,d} p_{j,d+1}} \right)$$

can have at most one solution in the interval $(\max\text{root}(p_{j,d}), \infty)$. But for $j \leq d$, we have

$$\max\text{root}(p_{j,d}) = \max\text{root}\left(C_j^{(1+\theta d)}(x)\right) \leq \max\text{root}\left(C_d^{(1+\theta d)}(x)\right) \leq \Gamma_\theta^d$$

implying that $p_{j,d+1}p_{j,d} - p_{j-1,d}p_{j-1,d+1}$ has a single root in $(\Gamma_\theta^d, \infty)$ and proving the theorem. \square

Note that, even though we were forced to consider the possibility that $\gamma_\theta^d > \gamma_\theta^{d+1}$ in the proof of Lemma 6.8, one direct implication of the lemma is that such a scenario is impossible.

Corollary 6.9. *For all $\theta > 0$, $n \geq 0$, and $d \geq 1$, the sequence*

$$\left\{ \maxroot \left(C_d^{(d\theta+1)} \right) \right\}_d$$

is monotone increasing, and

$$\lim_{d \rightarrow \infty} \maxroot \left(C_d^{(d\theta+1)} \right) = \gamma_\theta$$

where $\gamma_\theta = \frac{\sqrt{2\theta+1}}{\theta+1}$ (as in Lemma 6.2).

Proof. By Theorem 6.4, we know that

$$\lim_{d \rightarrow \infty} \maxroot \left(C_d^{(1+d\theta)}(x) \right) \leq \frac{\sqrt{2\theta+1}}{\theta+1} = \gamma_\theta \quad (6.7)$$

and the combination of Lemma 6.7 and Lemma 6.8 show that this convergence is monotone increasing. On the other hand, Lemma 6.2 shows that the asymptotic root distribution is dense in $[-\gamma_\theta, \gamma_\theta]$, and so the inequality in (6.7) must be an equality. \square

For our purposes, we will need a similar statement about the Cauchy transform:

Corollary 6.10. *For $n \geq 0$ and $d \geq 1$, we have*

1. $\maxroot \left(C_d^{(n+1)} \right) \leq \gamma_{n/d}$
2. *for all $x > \gamma_{n/d}$, we have*

$$\mathcal{G}_{C_d^{(n+1)}}(x) \leq \frac{-nx + \sqrt{(n+d)^2(x^2-1) + n^2}}{d(x^2-1)}$$

where $\gamma_{n/d} = \sqrt{1 - \frac{n^2}{(n+d)^2}}$.

Proof. The first claim follows directly from Corollary 6.9 with $\theta = n/d$. For the second claim, Corollary 6.5 shows that

$$\lim_{d \rightarrow \infty} \mathcal{G}_{C_d^{(1+\theta d)}}(x) = \frac{-\theta x + (1+\theta)\sqrt{x^2 - \gamma_\theta^2}}{(x^2-1)}.$$

and the combination of Lemma 6.7 and Lemma 6.8 show that this convergence is monotone increasing. Setting $\theta = n/d$ and simplifying then gives the corollary. \square

Note that while the form of Corollary 6.10.2 is the more popular one in the literature, it has a downside when appearing in inequalities (the false appearance of a sign change at $x = 1$). It is easy to check by cross multiplication that an equivalent way to write this inequality is

$$\mathcal{G}_{C_d^{(1+n)}}(x) \leq \frac{2n + d}{nx + \sqrt{(n + d)^2(x^2 - 1) + n^2}} \tag{6.8}$$

and this will be the form we use in our proof of Theorem 6.10.

Theorem 6.11. *Consider the bivariate polynomial*

$$f(x, y) = d(x^2 - 1)y^2 + 2nxy - (2n + d)$$

and assume that for a given $s > \gamma_{n/d}$ and $t > 0$ that $f(s, t) \geq 0$. Then

$$\mathcal{G}_{C_d^{n+1}}(s) \leq t.$$

Proof. Fix $u > \gamma_{n/d}$ and let $G = \mathcal{G}_{C_d^{(n+1)}}(u)$. By Corollary 6.10 and (6.8), we have

$$\frac{1}{G} \geq \frac{nu + \sqrt{(n + d)^2(u^2 - 1) + n^2}}{2n + d} \implies \frac{2n + d}{G} - nu \geq \sqrt{(n + d)^2(u^2 - 1) + n^2}.$$

Since both sides are positive, we can square them, to get

$$\frac{(2n + d)^2}{G^2} - \frac{2nt(2n + d)}{G} + n^2u^2 \geq (n + d)^2(u^2 - 1) + n^2.$$

Using the fact that $G > 0$, we can then simplify to get

$$d(u^2 - 1)G^2 + 2nuG - (2n + d) \leq 0.$$

Hence for any $u > \gamma_{n/d}$, we have that

$$f\left(u, \mathcal{G}_{C_d^{(n+1)}}(u)\right) \leq 0. \tag{6.9}$$

Now let $s > \gamma_{n/d}$ and $t > 0$ satisfy $f(s, t) \geq 0$. It is well known that the range of the Cauchy transform is the positive reals, so there exists some r for which $t = \mathcal{G}_{C_d^{(n+1)}}(r)$. It then suffices to show that $r \leq s$, since (due to the fact that the Cauchy transform is decreasing) that would imply

$$\mathcal{G}_{C_d^{(n+1)}}(s) \leq \mathcal{G}_{C_d^{(n+1)}}(r) = t.$$

as required.

To see that $r \leq s$, we consider two cases. If $r \leq \gamma_{n/d}$, then $r \leq s$ trivially. Otherwise, we have $r > \gamma_{n/d}$, so by (6.9), we must have $f(r, t) \leq 0 \leq f(s, t)$. However it is easy to check that

$$\frac{\partial}{\partial x} f(x, y) = 2dxy^2 + 2ny > 0 \tag{6.10}$$

whenever $x, y \geq 0$, implying $r \leq s$ in this case as well. □

7. Illustrative Examples

In this section we hope to give some intuition as to how one can view the W_α^n operator and (in particular) the role that the parameter α plays. We then give computational examples that show the relative accuracy of Theorem 1.3. All plots and computations in this section were done using Mathematica 12.

7.1. Singular values of rectangular matrices

Before discussing polynomials directly, it will be informative to recall the connection with the singular values of matrices that was mentioned Section 1.3.3. That is, let A and B be $(d+n) \times d$ matrices with

$$p_A(x) = \det [x\mathbb{I} - AA^*] \quad \text{and} \quad p_B(x) = \det [x\mathbb{I} - BB^*].$$

so that the roots of $\mathbb{S}p$ are

$$\{\pm\sigma_1(A), \dots, \pm\sigma_d(A)\}$$

(and similar for B and q). One (important) consequence of Theorem 2.3 is that there exists a $(d+n) \times d$ matrix C for which

$$[p_A \boxplus_d^n p_B](x) = \det [x\mathbb{I} - CC^*].$$

Thus if we define $[A]$ to be the class of matrices which have the same dimensions and same singular values as A , then the rectangular additive convolution can be viewed as a binary operation on these classes. Coupled with the observation made in Section 1.3.3 that we have the explicit formula

$$[p_A \boxplus_d^n p_B](x) = \int_{Q,R} \det [x\mathbb{I} - (A + QBR)(A + QBR)^*] dQdR \quad (7.1)$$

when dQ and dR are Haar-uniform random unitary matrices of the appropriate size, one could feel justified in writing this binary relation as

$$[A] \tilde{+} [B] = [C].$$

An obvious question, then, is to what extent $[A] \tilde{+} [B]$ and $[A + B]$ have similar behavior. In general, they cannot be the same because the singular values of $A + B$ will depend on the singular vectors of A and B , whereas $[A] \tilde{+} [B]$ is independent of the singular vectors. However, if we were to interpret $\tilde{+}$ as some sort of “unitarily invariant” way of “adding” two matrices together, then Theorem 1.3 would be a statement about the behavior of the largest singular value under this operation.

The first relevant observation in this direction is that Theorem 1.3 simplifies greatly when $\alpha = 0$, in that

$$\Theta_0^n(p) = \text{maxroot}(\mathbb{S}p).$$

Hence in this case, Theorem 1.3 reduces to

$$\text{maxroot}(\mathbb{S}[p \boxplus_d^n q]) \leq \text{maxroot}(\mathbb{S}p) + \text{maxroot}(\mathbb{S}q)$$

which, when written in the matrix context, becomes the inequality

$$\sigma_{\max}([A] \tilde{+} [B]) \leq \sigma_{\max}([A]) + \sigma_{\max}([B]). \tag{7.2}$$

That is, $\tilde{+}$ satisfies a triangle inequality similar to normal matrix addition (and in fact this triangle inequality can be derived from the normal one using (7.1)).

The shortcoming of (7.2) is that it neglects much of the information we have concerning the singular values of the original matrices. If we are to accept the interpretation that $\tilde{+}$ is some sort of a “unitarily invariant” version of addition, then we should suspect that the true answer will depend on all of the singular values of A and B . This is the case, and the purpose of W_α^n is to try to obtain better bounds by accessing this extra information.

7.2. The role of W_α^n

For the purpose of illustration, let us consider the matrix classes $[A]$ and $[B]$ with:

$$\sigma_d([A]) = \sigma_{d-1}([A]) = \dots = \sigma_2([A]) = 1 \quad \text{and} \quad \sigma_1([A]) = 2.$$

versus

$$\sigma_d([B]) = 1 \quad \text{and} \quad \sigma_{d-1}([B]) = \sigma_{d-2}([B]) = \dots = \sigma_1([B]) = 2$$

It should be clear that the triangle inequality (7.2) would treat all of the sums involving A and B similarly:

$$\sigma_{\max}([A] \tilde{+} [A]) \leq 4 \quad \text{and} \quad \sigma_{\max}([A] \tilde{+} [B]) \leq 4. \quad \text{and} \quad \sigma_{\max}([B] \tilde{+} [B]) \leq 4.$$

However, the interpretation of $\tilde{+}$ as a “unitarily invariant” version of addition suggests that $\sigma_{\max}([A] \tilde{+} [A])$ should be significantly larger than $\sigma_{\max}([B] \tilde{+} [B])$, since many of the rotations Q, R in (7.1) will result in

$$\sigma_{\max}(A + QAR) = 4 \quad \text{and} \quad \sigma_{\max}(B + QBR) = 3.$$

The goal then is to improve this bound by taking into account the positions of the other singular values (of both matrices). However, it is important for applications that we incorporate this information in a way that can be easily iterated, as we will often find ourselves wanting to apply knowledge we have gained about $\sigma_{\max}([A] \tilde{+} [B])$ to gain knowledge about $\sigma_{\max}([A] \tilde{+} [B] \tilde{+} [C])$ (for example). The upshot of Theorem 1.3 is that we can accomplish this using the W_α^n transformation.

The intuition behind the action of W_α^n that we find the most illuminating is inspired by a model from [20]. If one were to view the roots of $p(x)$ as particles, then the application of W_α^n can be seen as a gust of wind (with total strength parametrized by α) pushing the particles forward. In theory, one would expect any such wind to divide its force evenly between the particles (pushing them the same amount). However the operator W_α^n has the property that the positive roots of p and those of $W_\alpha^n p$ will always *interlace*¹. That is, if

$$s_1 \geq s_2 \geq \dots \geq s_d \quad \text{and} \quad s'_1 \geq s'_2 \geq \dots \geq s'_d$$

¹We will not prove this here, but it is a direct consequence of Theorem 2.3 — see [4].

are the roots of p and $W_\alpha^n p$ (respectively), then one will always have the inequality

$$s_{i-1} \geq s'_i \geq s_i.$$

As a result, the force of the wind treats the particles like “billiard balls” with polynomials like p_B (that have many roots close to the largest root) transferring much of the energy onto the largest root. In comparison, p_A will see a lot of its energy transferred onto its second-largest root, but as long as α is not too big, there will not be much energy transferred on to the largest root.

Hence the distance that the operator W_α^n pushes the root of a polynomial, when viewed as a function of α , encodes different levels of information about the locations of the other roots. Values of α that are too large will result in most of the energy being transferred onto the largest root (independent of the initial configuration). Similarly, small values of α will result in very little of the energy being transferred on to the largest root (also independent of the initial configuration). Intuitively, the optimal α will be one that allows as much initial force as possible while still avoiding too much transfer onto the largest particle.

7.3. Example computations

Let us consider the examples from the previous section in the case where A and B are 6×9 matrices (so $d = 6, n = 3$). Hence

$$p_A(x) = (x - 1)^5(x - 4) \quad \text{and} \quad p_B(x) = (x - 1)(x - 4)^5.$$

and the resulting convolutions are (where we have replaced \boxplus_6^3 with \boxplus to aid readability):

$$\begin{aligned} [p_A \boxplus p_A](x) &= x^6 - 18x^5 + 120x^4 - 380x^3 + 600x^2 - 442x + \frac{2431}{21} \\ [p_A \boxplus p_B](x) &= x^6 - 30x^5 + 350x^4 - \frac{6050}{3}x^3 + 5975x^2 - 8450x + \frac{90080}{21} \\ [p_B \boxplus p_B](x) &= x^6 - 42x^5 + \frac{2060}{3}x^4 - 5520x^3 + 22600x^2 - \frac{130624}{3}x + \frac{622336}{21}. \end{aligned}$$

Figure 7.1 shows two functions of α — the true value of $\max\text{root}(W_\alpha^3 q)$ (in yellow) plotted against the bound given by Theorem 1.3 (in blue) — for each of the convolutions above. In particular, the blue curve in each case will have value 4 when $\alpha = 0$ (as dictated by the triangle inequality), while the value we are trying to bound will be the value of the yellow curve at $\alpha = 0$. Unfortunately, we were forced to cut off the values at $\alpha = 0$ in the graphs due to precision issues, so we have listed the relevant values in Table 7.1.

8. Further Research

One possible direction of research comes from an observation in [3] concerning a degeneration in the rectangular convolution in the case $n = 0$ ($\lambda = 1$ in [3]) that allows one to compute singular value distributions of square matrices using the additive convolution. In particular, if the asymptotic singular laws of two independent (left and right unitarily invariant) matrices converge

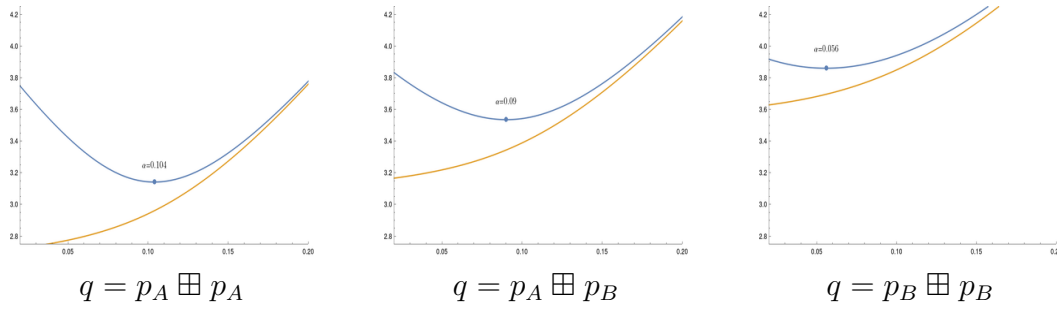


Figure 7.1: Optimal values of α for three convolutions involving p_A and p_B . The blue curve shows the bound provided by Theorem 1.3 as a function of α . The yellow curve shows $\maxroot(W_\alpha^3 q)$ for the convolution polynomial q .

Convolution $q(x)$	Best α_0 in Thm. 1.3	Resulting bound	$\maxroot(q(x^2))$	(7.2) bound
$p_A \boxplus p_A$	0.104	3.143	2.702	4
$p_A \boxplus p_B$	0.090	3.537	3.144	4
$p_B \boxplus p_B$	0.056	3.861	3.606	4

Table 7.1: The best α when Theorem 1.3 is applied to each convolution, the resulting bound, and the true maximum (the yellow curve in Figure 7.1 at $\alpha = 0$).

to μ_1 and μ_2 , then the asymptotic singular law of their sum is the unique probability measure on $[0, \infty)$ which, when symmetrized, is equal to the free convolution of the symmetrizations of μ_1 and μ_2 . In terms of polynomial convolutions, this would mean that

$$\mathbb{S}[p \boxplus_d^0 q] = [\mathbb{S}p \boxplus_{2d} \mathbb{S}q]$$

which is not true (in general). It is true that they share the same bounds; by Theorem 1.1,

$$\maxroot(U_\alpha[\mathbb{S}p \boxplus_{2d} \mathbb{S}q]) \leq \maxroot(U_\alpha \mathbb{S}p) + \maxroot(U_\alpha \mathbb{S}q) - 2\alpha d$$

and by Theorem 1.2,

$$\maxroot(U_\alpha \mathbb{S}[p \boxplus_d^0 q]) \leq \maxroot(U_\alpha \mathbb{S}p) + \maxroot(U_\alpha \mathbb{S}q) - 2\alpha d.$$

One might then ask whether there is an inequality between the two, and we conjecture that, in fact, there is:

Conjecture 8.1. For all p, q in $\mathcal{P}_{\leq d}^+$, we have

$$\maxroot(U_\alpha \mathbb{S}[p \boxplus_d^0 q]) \leq \maxroot(U_\alpha[\mathbb{S}p \boxplus_{2d} \mathbb{S}q]).$$

It is not hard to show that Conjecture 8.1 is true for basic polynomials in \mathcal{P}_d^+ . Recall from Corollary 5.6 that for basic polynomials $p(x) = (x - \lambda)^d$ and $q(x) = (x - \mu)^d$ that

$$\mathbb{S}[p \boxplus_d^0 q](y) = (\lambda\mu)^{d/2} C_d^1 \left(\frac{y^2 - (\lambda + \mu)}{2\sqrt{\lambda\mu}} \right).$$

Using a similar method, one can show that for these polynomials

$$[\mathbb{S}p \boxplus_{2d} \mathbb{S}q](y) = \frac{(16\lambda\mu)^{d/2}}{\binom{2d}{d}} C_d^{1/2} \left(\frac{y^2 - (\lambda + \mu)}{2\sqrt{\lambda\mu}} \right).$$

As was noted in Remark 6.3, the positive roots of $C_d^\alpha(x)$ are decreasing with α , which will result in the roots of $[\mathbb{S}p \boxplus_{2d} \mathbb{S}q]$ majorizing the roots of $\mathbb{S}[p \boxplus_d^0 q](y)$. Since the function $1/(x - t)$ is convex for $x > t$, the Cauchy transform is Schur convex on that range, and this implies Conjecture 8.1 (for this particular p and q). In particular, it may be possible to prove Conjecture 8.1 using an inductive method like the one developed in Section 4.

Later versions of [14] use a similar argument on basic polynomials to prove the “base case” for Theorem 1.2 from Theorem 1.1 and, if true in general, Conjecture 8.1 would imply Theorem 1.2 directly from Theorem 1.1. It would be interesting to know whether Theorem 1.3 can be proved as a corollary of Theorem 1.1, as this would imply new identities in free probability. However, this seems unlikely; as was noted in Section 1.1, the only reason Theorem 1.2 can be stated using the U_α operator is due to a degeneration at $n = 0$ that does not hold for any other n and there is no reason to believe that U_α and Θ_α^n can be compared in an asymptotically tight way for general n .

Lastly, it would be interesting to determine whether Theorem 1.3 is a special case of some submodularity inequality similar to Theorem 1.1 and Theorem 1.4.

References

- [1] O. Arizmendi, D. Perales, *Cumulants for finite free convolution*, Journal of Combinatorial Theory, Series A, 155 (2018), 244–266.
- [2] R. Askey, G. Gasper, *Positive Jacobi polynomial sums. II*, Amer. J. of Math. (1976), 98 (3), 709–737.
- [3] F. Benaych-Georges, *Rectangular random matrices, related convolution*, Probab. Theory Relat. Field (2009) 144:471-515. arXiv:math/0507336
- [4] J. Borcea, P. Brändén, *The Lee–Yang and Pólya–Schur programs, II: theory of stable polynomials and applications*, Comm. on Pure and Appl. Math., 62(12):1595–1631, 2009.
- [5] A. Elbert, A. Laforgia, *On the zeros of associated polynomials of classical orthogonal polynomials*, Differential and Integral Equations 6 (1993), 1137–1143.
- [6] A. Gribinski, A. W. Marcus, *Biregular, bipartite Ramanujan graphs of arbitrary size and degree (in polynomial time)*, preprint. arXiv:2108.02534
- [7] A. Gribinski, A. W. Marcus, *A theory of singular values for finite free probability*, in preparation.

- [8] A. Knutson, T. Tao, *The honeycomb model of $GL_n(\mathbb{C})$ tensor products I: Proof of the saturation conjecture*. Jour. of the Amer. Math. Soc. 12.4 (1999): 1055-1090.
- [9] A. B. J. Kuijlaars, W. Van Assche, *The asymptotic zero distribution of orthogonal polynomials with varying recurrence coefficients*. J. of Approx. Theory 99.1 (1999): 167-197.
- [10] J. Leake, N. Ryder, *On the further structure of the finite free convolutions*, preprint. arXiv:1811.06382
- [11] E. H. Lieb, A. D. Sokal, *A general Lee–Yang theorem for one-component and multicomponent ferromagnets*. Comm. Math. Phys. 80 (1981), no. 2, 153–179.
- [12] A. W. Marcus, *Discrete unitary invariance*, preprint. arXiv:1607.06679
- [13] A. W. Marcus, *Polynomial convolutions and (finite) free probability*, preprint. arXiv:2108.07054
- [14] A. W. Marcus, D. A. Spielman, N. Srivastava, *Finite free convolutions of polynomials*, preprint. arXiv:1504.00350
- [15] A. W. Marcus, D. A. Spielman, N. Srivastava, *Interlacing families I: bipartite Ramanujan graphs of all degrees*, Ann. of Math. 182-1 (2015), 307-325. arXiv:1304.4132
- [16] A. W. Marcus, D. A. Spielman, N. Srivastava, *Interlacing families III: improved bounds for restricted invertibility*, to appear (Isr. Jour. of Math.) arXiv:1712.07766
- [17] A. W. Marcus, D. A. Spielman, N. Srivastava, *Interlacing families IV: bipartite Ramanujan graphs of all sizes*, FOCS (2015). arXiv:1505.08010
- [18] J. A. Mingo, R. Speicher. *Free probability and random matrices*. Vol. 35. New York, NY, U.S.A.: Springer, 2017.
- [19] A. Naor and P. Youssef, *Restricted invertibility revisited*, preprint. arXiv:1601.00948.
- [20] N. Srivastava. *Spectral Sparsification and Restricted Invertibility* Ph.D. Thesis, Yale University. <https://math.berkeley.edu/nikhil/dissertation.pdf>
- [21] G. Szegő. *Orthogonal polynomials*. AMS, Providence, RI MR 51 (1975): 8724.
- [22] D. V. Voiculescu, *Limit laws for random matrices and free products*, D. Invent. math. (1991) 104: 201.
- [23] D. Wagner, *Multivariate stable polynomials: theory and applications*. Bull. of the Amer. Math. Soc. 48.1 (2011): 53-84.
- [24] H. Wilf, *generatingfunctionology*. AK Peters/CRC Press (2005).
- [25] Xie, J., Xu, Z. (2020). Subset selection for matrices with fixed blocks. To appear, Israel J. Math. arXiv:1903.06350

Appendix

In this appendix, we give a brief sketch of the computation leading from Theorem 6.4 to Corollary 6.5. We have seen such a calculation referred to as “standard” in various places in the literature, but wanted to give some indication as to how the proof goes for those who, like the authors, might be less familiar with such computations.

Lemma 8.2. For $\theta > 0$ and $n > 0$, let μ_θ^n be a sequence of compact distributions for which

$$\mu_\theta^n(y) \xrightarrow{n \rightarrow \infty} \mu_\theta(y) = \frac{1 + \theta}{\pi} \frac{\sqrt{\gamma_\theta^2 - y^2}}{1 - y^2} \mathbf{1}_{[-\gamma_\theta, \gamma_\theta]}(y)$$

where $\gamma_\theta = \frac{\sqrt{2\theta+1}}{\theta+1}$. Then for all $x > \gamma_\theta$, we have

$$\mathcal{G}_{\mu_\theta^n}(y) \xrightarrow{n \rightarrow \infty} \mathcal{G}_{\mu_\theta}(x) = \frac{-\theta x + (1 + \theta)\sqrt{x^2 - \gamma_\theta^2}}{(x^2 - 1)}.$$

Proof. Since all of the distributions are compact, we can interchange the limit with the integral defining the Cauchy transform. That is,

$$\lim_{n \rightarrow \infty} \mathcal{G}_{\mu_\theta^n}(x) = \int \lim_{n \rightarrow \infty} \frac{\mu_\theta^n(z)}{x - z} dz = \frac{1 + \theta}{\pi} \int_{-\gamma_\theta}^{\gamma_\theta} \frac{\sqrt{\gamma_\theta^2 - z^2}}{(x - z)(1 - z^2)} dz.$$

We then make a change of variable from z to u , using the Euler substitution:

$$z = \frac{-2\gamma_\theta u}{1 + u^2} \quad \text{so that} \quad dz = 2\gamma_\theta \frac{u^2 - 1}{(u^2 + 1)^2} du,$$

yielding

$$\lim_{n \rightarrow \infty} \mathcal{G}_{\mu_\theta^n}(x) = \frac{1 + \theta}{\pi} \int_{-1}^1 \frac{2\gamma_\theta^2(1 - u^2)^2}{(xu^2 + 2\gamma_\theta u + x)((1 + u^2)^2 - 4\gamma_\theta^2 u^2)} du. \quad (8.1)$$

We can then rewrite the integrand using partial fractions:

$$\frac{2\gamma_\theta^2(1 - u^2)^2}{(xu^2 + 2\gamma_\theta u + x)((1 + u^2)^2 - 4\gamma_\theta^2 u^2)} = 2 \frac{x^2 - \gamma_\theta^2}{(x^2 - 1)x} g\left(\frac{1}{x}\right) + \frac{\gamma_\theta^2 - 1}{(x + 1)} g(-1) + \frac{\gamma_\theta^2 - 1}{(x - 1)} g(1)$$

where we have defined

$$g(s) := \frac{1}{u^2 + 2su\gamma_\theta + 1} = \frac{1}{(u - s\gamma_\theta)^2 + 1 - s^2\gamma_\theta^2}.$$

For $|s\gamma_\theta| < 1$, these integrals can be computed explicitly using the trigonometric substitution $u = \tan(\theta)\sqrt{1 - \gamma_\theta^2 s^2}$. This gives

$$\int_{-1}^1 g(s) du = \frac{1}{\sqrt{1 - \gamma_\theta^2 s^2}} \left(\arctan\left(\sqrt{\frac{1 - \gamma_\theta s}{1 + \gamma_\theta s}}\right) + \arctan\left(\sqrt{\frac{1 + \gamma_\theta s}{1 - \gamma_\theta s}}\right) \right) = \frac{\pi}{2\sqrt{1 - \gamma_\theta^2 s^2}}$$

since $\arctan(z) + \arctan(1/z) = \pi/2$ for all z . The result follows by plugging these into (8.1) and simplifying. \square