Polynomial convolutions and (finite) free probability

Adam W. Marcus
Princeton University

April 30, 2018

Abstract

We introduce a finite version of free probability and show the link between recent results using polynomial convolutions and the traditional theory of free probability. One tool for accomplishing this is a seemingly new transformation that allows one to reduce computations in our new theory to computations using classically independent random variables. We then explore the idea of finite freeness and its implications. Lastly, we show applications of the new theory by deriving the finite versions of some well-known free distributions and then proving their associated limit laws directly. In the process, we gain a number of insights into the behavior of convolutions in traditional free probability that seem to get lost when the operators being convolved are no longer finite.

1 Introduction

After its introduction in 1986 in a series of papers by Dan Voiculescu, free probability has seen an incredible growth in both its theory and its applications. This included a theory of *free cumulants*, first introduced by Nica and Speicher, which gave a unified framework for understanding classical and free independence through the lens of combinatorics [26]. It has been used as a tool in a variety of areas, including random matrix theory, combinatorics, representations of symmetric groups, large deviations, and quantum information theory. In most cases, the relationships mentioned above only exist in an asymptotic sense, primarily due to the fact that no nontrivial free objects exist in finite dimensions. However, recent work of the author with Daniel Spielman and Nikhil Srivastava [17, 18, 21] suggests that the behavior of finite structures closely resembles the asymptotic "free" behavior, despite not technically being "free". The goal of this paper is to introduce a theory that we call "finite free probability" as a way to extend the fundamental concepts and insights of free probability to finite objects.

1.1 A brief introduction to free probability

We begin with an informal introduction to free probability (in particular, as it relates to classical probability). Let (M_1, μ_1) and (M_2, μ_2) be probability spaces and let $p(x, y)$ a bivariate polynomial. In classical probability, one realizes μ_1 and μ_2 as *random variables* X_1 and X_2 with the goal of investigating the distribution μ_Y of objects of the form $Y = p(X_1, X_2)$ living in the tensor product $M_1 \otimes M_2$. Each of these probability spaces are equipped with a *test function* $\mathbb{E}\{\}$ that allows one to measure various functions of the X_i. Typically, μ_Y cannot be calculated without knowledge of the joint probability distribution μ_{X_1, X_2}. However, there is one "special" joint distribution for

which one can calculate μ_Y only knowing μ_1 and μ_2 for any p — the case when X_1 and X_2 are *independent*.

In noncommutative probability, one realizes μ_1 and μ_2 as the spectrum of *random operators* A_1 and A_2. The goal is still to investigate the distribution μ_Y of objects of the form $Y = p(A_1, A_2)$, however this time such objects live in the free product of M_1 and M_2. Free independence (or "freeness") is the free product analogue of classical independence on tensor products. In particular, it is the "special" joint distribution that allows μ_Y to be calculated completely in terms of μ_1 and μ_2. This time the spaces are equipped with a test function $\phi[]$ which has tracial properties (since the distribution lives on the spectrum of operators) that is used to measure various functions of the A_i.

Formally, we say that two random operators A, B are freely independent if for all n and all polynomials $p_1, \ldots, p_{2n}$, we have that

$$\phi\left[p_1(A)p_2(B)\ldots p_{2n-1}(A)p_{2n}(B)\right] = 0$$

whenever $\phi\left[p_{2i-1}(A)\right] = \phi\left[p_{2i}(B)\right] = 0$ for all $1 \le i \le n$. In practice, the main difference between free and classical independence is that free independence respects the noncommutativity of random operators. For example, when X and Y are independent, we have

$$\mathbb{E}\left\{X^2 Y^2\right\} = \mathbb{E}\{XYXY\} = \mathbb{E}\left\{X^2\right\}\mathbb{E}\left\{Y^2\right\}$$

whereas when A and B are freely independent, we have

$$\phi\left[A^2 B^2\right] = \phi\left[A^2\right]\phi\left[B^2\right]$$

while

$$\phi\left[ABAB\right] = \phi\left[A^2\right]\phi\left[B\right]^2 + \phi\left[A\right]^2\phi\left[B^2\right] - \phi\left[A\right]^2\phi\left[B\right]^2$$

On the other hand, there are a number of connections between classical and free independence, with varying degrees of understanding. Random matrix theory, in particular, captures a number of these connections, as many times classical independence of individual entries leads to free independence of spectral distributions. For example, we have the following theorem of Voiculescu:

Theorem 1.1. *Let μ_A and μ_B be discrete probability distributions and let A_m and B_m be $m \times m$ real symmetric matrices with eigenvalue distribution μ_A and μ_B (respectively). Let R_m and R'_m be i.i.d. random matrices drawn uniformly (via the Haar measure) from $O(m)$ (the orthogonal group). Then the operators*

$$A = \lim_{m\to\infty} R_m A_m R_m^T \qquad and \qquad B = \lim_{m\to\infty} R'_m B_m R'^T_m$$

are freely independent.

Theorem 1.1 reveals one of the hurdles in applying free probability theory to finite structures: finite dimensional matrices are freely independent if and only if one of them is a multiple of the identity. As a result, applications of the theory must be done in an asymptotic sense.

1.2 Convolutions

The computational aspects of free probability center around computing polynomial functions of random operators. Unlike in the classical case, however, even the basic operations of addition and multiplication are nontrivial to compute, even when the operators involved are freely independent.

2

1.2.1 Additive Convolution

Let A and B be operators with spectral distributions μ_A and μ_B. The *free additive convolution* of μ_A and μ_B (written $\mu_A \boxplus \mu_B$) is the distribution of the operator $A + B$ in the case that A and B are freely independent. To compute such a thing, we begin by computing the *Cauchy transform* of μ_A

$$\mathcal{G}_{\mu_A}(x) = \sum_i \frac{1}{x^{i+1}} \phi\left[A^i\right] = \phi\left[(xI - A)^{-1}\right] = \int \frac{\mu_A(t)}{x - t}\, \mathrm{d}t \tag{1}$$

and then the so-called *R-transform* as a function of the inverse of $\mathcal{G}_{\mu_A}(x)$

$$\mathcal{R}_{\mu_A}(x) = \mathcal{G}_{\mu_A}^{-1}(x) - \frac{1}{x} = \mathcal{G}_{\mu_A}^{-1}(x) - \mathcal{G}_{\mu_0}^{-1}(x)$$

where 0 is the zero matrix. Here, "inverse" means the compositional inverse of the power series of $\mathcal{G}_{\mu_A}(x)$ expanded around $x = \infty$. When A and B are freely independent, one then has

$$\mathcal{R}_{\mu_A \boxplus \mu_B}(x) = \mathcal{R}_{\mu_A}(x) + \mathcal{R}_{\mu_B}(x). \tag{2}$$

Alternatively, one could simply define the free additive convolution $\mu_A \boxplus \mu_B$ as being the spectral distribution satisfying Equation (2). In the case that all of the operators have compact support, this is known to uniquely define the distribution $\mu_A \boxplus \mu_B$ [32].

We remark that Equation (1) shows the relationship between the Cauchy transform and the (non-exponential) moment generating function

$$\frac{1}{x}\mathcal{G}_\mu\left(\frac{1}{x}\right) = \sum_i M_i(\mu_A)x^i = \sum_i \phi\left[A^i\right] x^i$$

where

$$M_i = \int x^i \mu_A(x)\, \mathrm{d}x$$

is the ith moment. In this respect, it is similar to the way in which one would form the classical additive convolution by forming the moment generating function $\mathbb{E}\{e^{xY}\}$ and then adding some transformation of it (in this case, $\ln \mathbb{E}\{e^{xY}\}$).

1.2.2 Multiplicative Convolution

For the multiplicative case, one uses a variant of the Cauchy transform above

$$\widetilde{\mathcal{M}}_{\mu_A}(x) = x\mathcal{G}_{\mu_A}(x) - 1$$

and then form (what we call) the *modified S-transform* by inverting a power series (again taken at $x = \infty$)

$$\widetilde{\mathcal{S}}_{\mu_A}(x) = \frac{x}{1 + x}\widetilde{\mathcal{M}}_{\mu_A}^{(-1)}(x) = \frac{\widetilde{\mathcal{M}}_{\mu_A}^{(-1)}(x)}{\widetilde{\mathcal{M}}_{\mu_I}^{(-1)}(x)}$$

where I is the identity matrix. When A and B are freely independent, the moments of $\mu_A \boxtimes \mu_B$ are then defined by

$$\widetilde{\mathcal{S}}_{\mu_A \boxtimes \mu_B}(x) = \widetilde{\mathcal{S}}_{\mu_A}(x)\, \widetilde{\mathcal{S}}_{\mu_B}(x) \tag{3}$$

Remark 1.2. We call this the "modified" S-transform because it does not use the same definition that is typically seen in free probability [32]. The relation to the usual definition is simply

$$\widetilde{\mathcal{S}}_{\mu_A}(x) = \frac{1}{\mathcal{S}_{\mu_A}(x)}.$$

This of course does not change the relationship in Equation (3) in any way, and so for the purposes of traditional free probability it seems like a silly alteration. However, in the finite setting, it does seem to be the more appropriate choice (as is discussed in [18]).

1.3 Previous Work

The idea of "finite free probability" is not new. As best the author could tell, it was first proposed by Edelman and Rao in 2005 [11]. Later, the same authors suggested an implementation of such a theory, although with a focus on the computational (i.e. with Matlab) aspects of free probability [28]. Their implementation uses polynomials (just as ours will); however, their polynomials are substantially different from the ones used in this paper. Rather than encoding distributions in the roots of a polynomial, they encode various transforms as the solutions of bivariate polynomials and then give operations on these polynomials that capture the behavior of additive and multiplicative convolution.

The use of polynomials in [28] is far from coincidental. Polynomials have been closely connected to random matrix theory from its initial beginnings. It was recognized early on that the spectral distributions of random matrices matched the asymptotic root distribution of various orthogonal polynomials, and many of the known results regarding spectral distributions have been proved using this correspondence [8]. Polynomials have also been used in similar ways in a strictly free probabilistic setting [2], and are one of the major tools in the investigation of *universality* for random matrices [12].

Lastly, we mention the work of Pereira connecting so-called *trace vectors* to hyperbolic polynomials [27], which can be seen as a special case of the finite freeness developed in Section 5. See Section 7.1 for more details.

1.4 New Results

The contribution of this paper is to show a direct link between polynomials and free probability that the author believes is fundamentally new. Previous work used polynomials only for their asymptotic properties, and usually only orthogonal polynomials. In this work, we show that polynomials actually exhibit a close link to free probability on a finite scale. Furthermore, this will provide a framework for understanding recent work in topics such as restricted invertibility [17] and Ramanujan graphs [21]. By linking these methods, we hope to provide a systematic way to use free probability as a tool in combinatorics and graph theory (and vice versa).

1.5 Organization

We begin in Section 2 by listing the collection of ideas and definitions that we will use. In Section 3, we will introduce a transformation on finite multisets that we call the U transform, which will be useful in the computations of Section 4. In Section 4, we show how the symmetric additive and symmetric multiplicative polynomial convolutions introduced in [18] relate to convolutions from free

4

probability. In Section 5, we then examine the properties of finite freeness. Lastly, in Section 6, we show some applications of the theory by deriving the finite versions of various laws in free probability and then deriving their associated limit theorems directly. We also show how one the theory can be used to shine a more intuitive light on results such as Bourgain and Tzafriri's restricted invertibility theorem [5, 17]. We finish with some concluding remarks, suggestions for further research directions, and acknowledgements.

2   Preliminaries

Throughout the paper, we will use ρ_A to denote the spectral radius of an operator A and $\operatorname{tr}[]$ to denote the normalized trace (so that $\operatorname{tr}[I] = 1$). In the case of matrices, we will use $\operatorname{Tr}[]$ to denote the usual trace (so that $\operatorname{Tr}[I] = \dim(I)$).

The first part of this section will review the polynomial convolutions from [18] that will be the focus of this paper. The remainder of the section will be used to introduce some of the tools that will be used in the analysis.

2.1   Polynomial convolutions

Finite convolutions of polynomials were introduced in [18]. We will be concerned with the real, symmetric cases. Let A and B be $m \times m$ real symmetric matrices with

$$p(x) = \det[xI - A] \qquad \text{and} \qquad q(x) = \det[xI - B].$$

Definition 2.1. The *symmetric additive convolution* of p and q is defined as

$$[p \boxplus_m q](x) = \mathbb{E}_Q\big\{\det\big[xI - A - QBQ^T\big]\big\}$$

where the expectation is taken over orthonormal matrices Q distributed uniformly (via the Haar measure). For A and B positive definite, the *symmetric multiplicative convolution* of p and q is defined as

$$[p \boxtimes_m q](x) = \mathbb{E}_Q\big\{\det\big[xI - AQBQ^T\big]\big\}.$$

Furthermore, they note that when p and q each have all real roots, then both

$$[p \boxplus_m q](x) \qquad \text{and} \qquad [p \boxtimes_m q](x)$$

have all real roots, due to the (much more general) theory developed in [4]. The following linear formulas were proved as well:

Theorem 2.2. *If*

$$p(x) = \sum_i x^{m-i}(-1)^i p_i \qquad \text{and} \qquad q(x) = \sum_i x^{m-i}(-1)^i q_i$$

then

$$[p \boxplus_m q](x) = \sum_{i+j \leq m} x^{m-i-j}(-1)^{i+j}\frac{(m-i)!(m-j)!}{(m-i-j)!m!}p_i q_j \tag{4}$$

and

$$[p \boxtimes_m q](x) = \sum_{i=0}^{m} x^{m-i}(-1)^i \frac{p_i q_i}{\binom{m}{i}} \tag{5}$$

2.2   Laplace Transform

For a function f, the *Laplace transform* is defined as

$$\mathcal{L}\{f\}(s) = \int_0^\infty e^{-xs} f(x)\, \mathrm{d}x.$$

The Laplace transform is quite useful combinatorially due to its ability to turn exponential generating functions into ordinary generating functions (and vice versa), as

$$\mathcal{L}\left\{x^k\right\}(s) = \frac{k!}{s^{k+1}}. \tag{6}$$

Those uninterested in the details of convergence should feel free to simply treat the transform as a linear operator on power series (at least as far as this paper goes, very little would be lost in doing so). In fact, many of the technical details in Section 4 (the computations regarding the domain of integration where the boundary ends up disappearing from the solution) are for the most part showing that we can safely treat the Laplace transform in exactly this way.

2.3   Legendre Transform

Let f be a function that is convex on an interval $X \subseteq \mathbb{R}$. The *Legendre transform* is defined to be the function

$$f^*(s) = \sup_{x \in X} \{xs - f(x)\} \tag{7}$$

where the domain of f^* is the space

$$X^* = \left\{x^* \in \mathbb{R} : \sup_{x \in X}\{xx^* - f(x)\} < \infty\right\}.$$

Note that the convention is to have f^* be a function in the variable p, but we will use p for other purposes, and so we will use the variable s. In the case that f is differentiable, one has the following relation:

Lemma 2.3. *Let f be strictly convex on X and differentiable at a point $z \in X$. Then $f'(z) \in X^*$ and*

$$f^*(f'(z)) = zf'(z) - f(z).$$

Proof. Let $z \in X$. Since f is differentiable at z and strictly convex, it satisfies the inequality

$$f(x) \geq f(z) + (x - z)f'(z)$$

for all $x \in X$, with equality if and only if $x = z$. Rearranging, we have

$$\sup_{x \in X}\{xf'(z) - f(x)\} = zf'(z) - f(z) < \infty$$

which, by definition, means $f'(z) \in X^*$ and $f^*(f'(z)) = zf'(z) - f(z)$. $\qquad\square$

Many of the useful properties of the Legendre transform follow directly from Lemma 2.3. In particular, it implies that f^* is strictly convex whenever f is twice differentiable.

Corollary 2.4. *Let f satisfy the conditions of Lemma 2.3. Then*

1. $f'(f^{*\prime}(x)) = x$

2. $f''(f^{*\prime}(x)) = 1/f^{*\prime\prime}(x)$

Proof. Take derivatives of the formula in Lemma 2.3. □

The Legendre transform appears in many areas under many different names. It is often referred to as the *convex conjugate* in the analysis literature and as the *Fenchel transform* (or *Legendre–Fenchel transform*) in optimization. In particular, the term "convex conjugate" is typically applied when one applies Equation (7) to more general Banach spaces, and the term "Fenchel dual" is typically applied when one applies Equation (7) in primal–dual algorithms of convex programming. In this paper, we use only the most basic facts from theory, and as result, we will maintain the nomenclature *Legendre transform* to highlight this fact.

2.4 L^p norms

The main tool we will use is the theory of L^p spaces from Banach space theory. Given a measure space (X, μ), For $0 < p < \infty$, the L^p-norm of a function f is defined to be

$$\|f\|_p = \left(\int |f|^p \, d\mu \right)^{\frac{1}{p}}$$

and for $p = \infty$, we have

$$\|f\|_\infty = \lim_{p \to \infty} \|f\|_p = \inf \{a \geq 0 : \mu(\{x : |f(x)| > a\}) = 0\}$$

We will only use one simple result from the theory of L^p spaces:

Lemma 2.5. *If μ is absolutely continuous with respect to the Lebesgue measure and f is continuous, then*

$$\|f\|_\infty = \sup_{x \in X} \{|f(x)|\}$$

The following simple observation relates the Legendre transform to L^p spaces.

Corollary 2.6. *Let X be a subset of the real line and μ a measure that is absolutely continuous with respect to the Lebesgue measure. Then for any continuous function $f : X \to \mathbb{R}$, we have*

$$f^*(s) = \ln \left\| e^{xs - f(x)} \right\|_\infty$$

for all $s \in X^*$ (where both the Legendre transform and L^∞ norm are taken over the space X).

Proof. Since f is real, f^* is real as well. Hence we can write $f^*(s)$ as

$$\begin{aligned}
f^*(s) &= \ln(\exp(f^*(s))) \\
&= \ln \left(\exp \left(\sup_{x \in X} \{xs - f(x)\} \right) \right) \\
&= \ln \left(\sup_{x \in X} \{\exp(xs - f(x))\} \right).
\end{aligned}$$

7

Since f is real-valued, $\exp(xs - f(x))$ is strictly nonnegative, and so

$$\ln \left(\sup_{x \in X} \{ |\exp(xs - f(x))| \} \right) = \ln \left\| e^{xs - f(x)} \right\|_\infty$$

as required. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

For what follows, we will always have μ be the Lebesgue measure and the domain of integration will always be the real line. In the case that we want to integrate on a smaller domain, we will explicitly set the function to 0 as in Equation (8). We will also use the convention that all norms should be taken with respect to the variable x. That is, when we write

$$\| f(x, s) \|_p$$

we mean

$$\| f(\cdot, s) \|_p$$

and so the former should always be seen as a function of s.

2.5 Fuglede–Kadison determinants

Rather than give the general definition of the Fuglede–Kadison determinant in operator theoretic terms, we will simply define it in the generality for which we will need it. For a finite dimensional $m \times m$ positive definite matrix A, we define the *normalized determinant* to be

$$\Delta^+ (A) = \begin{cases} \det [A]^{1/m} & \text{if } A \text{ is positive definite} \\ 0 & \text{otherwise} \end{cases} \qquad (8)$$

Fuglede and Kadison showed that for a sequence of positive definite matrices $A_1, A_2, \ldots$, which converges in a suitable way to a limiting operator a, the normalized determinant converges to a well-defined limit and the limiting operator has many of the usual properties that one expects in a determinant (for example, multiplicativity) [13].

The Fuglede–Kadison determinant is typically defined in much greater generality, however we will only need the case where it acts as a normalized version of the determinant function (the goal being to trade generality (and perhaps rigor) for readability by those not familiar with C^*-algebras). The only nontrivial property that we will use is that $\Delta^+ ()$ is well-defined and is a limit of its values on a suitable sequence of matrices. It should be noted that the Fuglede–Kadison determinant (in the greater generality) already plays a role in the theory of free probability in the form of *Brown measures* (see [9] for a survey of such connections). It is unclear whether there is any relationship between the methods in this paper and such results.

2.6 Mixed discriminants

Let $X_1, \ldots, X_m$ be $m \times m$ matrices. We define the *mixed discriminant* to be

$$\mathrm{D}(X_1, \ldots, X_m) = \frac{\partial^m}{\partial t_1 \ldots \partial t_m} \det \left[\sum_{i=1}^m t_i X_i \right]$$

8

Remark 2.7. Note that our definition of the mixed discriminant differs by a factor of $m!$ from many other treatments. The literature is far from standard in this respect, and our reason for taking this normalization is that it will simplify a number of the formulas we will use.

The properties of mixed discriminants are well studied (see, for example [3]), and appear in numerous contexts. Some of the more well-known properties are the following:

Lemma 2.8. *Let* $A, B, X_2, \dots, X_m$ *be* $m \times m$ *matrices and let* c *be a scalar. Then*

1. $\mathrm{D}\,()$ *is invariant under permutation of its arguments*

2. $\mathrm{D}\,(A + cB, X_2, \dots, X_m) = \mathrm{D}\,(A, X_2, \dots, X_m) + c\mathrm{D}\,(B, X_2, \dots, X_m)$

3. $\mathrm{D}\,(A, \dots, A) = m!\det[A]$

4. $\mathrm{D}\,(AX_1, \dots, AX_m) = \det[A]\,\mathrm{D}\,(X_1, \dots, X_m)$

One corollary of these properties is that the determinant can be decomposed in a manner similar to the binomial theorem:

$$\det[xI + A] = \sum_i x^{m-i} \binom{m}{i} \mathrm{D}\,(A[i], I[m - i]). \tag{9}$$

While the mixed discriminant clearly has a close relationship to the determinant, it is also closely related to the *permanent*. In particular, when $X_1, \dots, X_m$ are *diagonal* matrices, one gets the formula

$$\mathrm{D}\,(X_1, \dots, X_m) = \mathrm{Perm}(Q) \tag{10}$$

where the columns of Q are the diagonals of the X_i.

The only nonstandard property of the mixed discriminant that we will use was proven in [6] and can be seen as a type of distributivity:

Lemma 2.9. *Let* A *and* $X_1, \dots, X_m$ *be* $m \times m$ *matrices. Then*

$$\mathrm{Tr}\,[A]\,\mathrm{D}\,(X_1, \dots, X_m) = \sum_{i=1}^m \mathrm{D}\,(X_1, \dots, X_{i-1}, AX_i, X_{i+1}, \dots, X_m)$$

While we will be using the mixed discriminant in the context of real symmetric matrices, it is worth mentioning that all of the properties mentioned here hold for arbitrary square matrices (except Equation (10), obviously, which requires the matrices to be diagonal).

3 U Transform

Given a multiset S, we will write $|S|$ for the number of elements in the multiset (with multiplicity). To be pedantic, $|\{1, 2, i, i\}| = 4$. To abuse notation, we will often treat multisets as though they were random variables. In any such case, the random variable should be considered to be uniformly distributed on the elements of the multiset. For example,

$$\frac{1}{|S|} \sum_{s_i \in S} f(s_i) \qquad \text{and} \qquad \mathbb{E}\{f(S)\}$$

will be used interchangably.

Lemma 3.1. *Let S be a finite multiset of complex numbers with $|S| = m$. Then there exists a unique multiset T of complex numbers with $|T| = m$ such that*

$$\prod_{s_i \in S} (x - s_i) = \mathbb{E}\{(x - T)^m\}.$$

for all x.

Proof. Clearly both sides of the equation are monic polynomials, so it suffices to prove that the equality holds for each of the other m coefficients. Each such equality can be seen as a constraint on $\mathbb{E}\{T^k\}$ for $1 \leq k \leq m$. Using Newton's identities, this is equivalent to having constraints on the first m elementary symmetric functions of the elements of T. However, this is equivalent to having m solutions to a polynomial of degree m, which is true (and unique) over the complex numbers. $\square$

Given a multiset S, we will refer to the multiset T which satisfies the constraints of Lemma 3.1 as the *U transform* of S. The property of having $|T| = |S|$ is an important one, as the next lemma will show.

Lemma 3.2. *Let S be a multiset of real numbers and T its U transform. Then*

$$\mathbb{E}\{f(T)\} \in \mathbb{R}.$$

for any function f that is analytic on the support of T.

Proof. Since all elements of S are real, the coefficients of

$$\prod_{s_i \in S}(x - s_i) = \sum_{k=0}^{m} \binom{m}{k} x^{m-k}(-1)^k \, \mathbb{E}\{T^k\}$$

are all real. Now consider the polynomial

$$q(x) = \prod_{t_i \in T}(x - t_i) = \sum_{i=0}^{m} q_i x^{m-i}.$$

By the Newton identities, the coefficients q_i are expressable as functions of the first m moments (which we have just seen are real) and are therefore real. Now let A_T be a real symmetric matrix with eigenvalues the elements of T. Then the Cayley–Hamilton theorem asserts that (as a matrix equation) $q(A_T) = 0$. Hence

$$0 = \text{tr}\left[A_T q(A_T)\right] = \sum_i q_i \text{tr}\left[A_T^{m-i+1}\right]$$

which is an expression for $\text{tr}\left[A_T^{m+1}\right]$ as a linear combination of $\text{tr}\left[A_T^k\right]$ for $k \leq m$ with real coefficients (and so itself is real). By proceeding inductively, we have $\text{tr}\left[A_T^k\right] = \mathbb{E}\{T^k\} \in \mathbb{R}$ for all k and so the same will be true for any analytic function. $\square$

Note that if $|T|$ were larger than $|S|$, there could be no such guarantees. To see this, let $k > 0$. We can use the same logic as Lemma 3.1 to find a multiset W with $|W| = (m + k)$ which has the same first m moments as μ, but then has a complex $(m + 1)^{th}$ moment.

10

Despite our willingness to treat multisets as distributions, one should be careful when going in the reverse direction. We will say that a distribution μ is *m-realizable* if there exists a multiset S_μ with $|S_\mu| = m$ such that the uniform distribution on S_μ gives the same probabilities as μ. The multiset S_μ will then be referred to as its *m-realization*. In particular, the U transform should always be defined in terms of a realization of a distribution and not the distribution itself. It is easy to see that an m-realizable distribution will also be km-realizable for any positive integer k. The realizations, however, will be (except in very special cases) quite different, as the next example shows.

Example 3.3. Let μ be the two-point distribution taking values $\{-1, 1\}$ each with probability $1/2$. Then $S_2 = \{-1, 1\}$ is its 2-realization, which has U transform $T_2 = \{-i, i\}$ which can be seen by computing

$$\mathbb{E}\{(x - S_2)^2\} = \frac{1}{2}(x^2 - 2xi - 1) + \frac{1}{2}(x^2 + 2xi - 1) = x^2 - 1 = (x+1)(x-1)$$

However, the 4-realization of μ ($S_4 = \{1, 1, -1, -1\}$) has U transform

$$T_4 = \left\{ \sqrt{\frac{2\sqrt{2} - 1}{3}}, \sqrt{\frac{2\sqrt{2} - 1}{3}}, i\sqrt{\frac{2\sqrt{2} + 1}{3}}, -i\sqrt{\frac{2\sqrt{2} + 1}{3}} \right\}$$

which can be checked by computing

$$\mathbb{E}\{T_4^4\} = 1 \qquad \text{and} \qquad \mathbb{E}\{T_4^2\} = -\frac{1}{3}$$

so that

$$\mathbb{E}\{(x - T_4)^4\} = x^4 - 2x^2 + 1 = (x - 1)^2(x + 1)^2.$$

The utility of the U transform will lie in its ability to turn polynomial convolutions (and therefore, as we shall see, finite free independence) into classical independence. This is illustrated in the following lemma:

Lemma 3.4. *Let p and q be degree m polynomials with S and T the U transforms of their roots. If S and T are independent, then*

$$[p \boxplus_m q](x) = \mathbb{E}\{(x - S - T)^m\} \qquad \text{and} \qquad [p \boxtimes_m q](x) = \mathbb{E}\{(x - ST)^m\}$$

Proof. Let $p(x) = \sum_i x^{d-i}(-1)^i p_i$. As observed previously, we have

$$p(x) = \mathbb{E}\{(x - S)^m\} = \sum_i x^{m-i} \binom{m}{i} \mathbb{E}\{(-S)^i\}$$

and so $p_i = \binom{m}{i} \mathbb{E}\{S^i\}$ (and similarly for q and T). By Equation (4), we have

$$
\begin{aligned}
[p \boxplus_m q](x) &= \sum_{i=0}^{m} \sum_{j=0}^{m-i} x^{m-i-j} (-1)^{i+j} \frac{(m-i)!(m-j)!}{m!(m-i-j)!} p_i q_j \\
&= \sum_{i=0}^{m} \sum_{j=0}^{m-i} x^{m-i-j} (-1)^{i+j} \frac{(m-i)!(m-j)!}{m!(m-i-j)!} \binom{m}{i}\binom{m}{j} \mathbb{E}\{S^i\} \, \mathbb{E}\{T^j\} \\
&= \sum_{i=0}^{m} \sum_{j=0}^{m-i} x^{m-i-j} (-1)^{i+j} \binom{m}{i,j} \mathbb{E}\{S^i T^j\} \qquad\qquad \text{(independence)} \\
&= \mathbb{E}\{(x - S - T)^m\} \, .
\end{aligned}
$$

The multiplicative case is similar, using Equation (5):

$$
\begin{aligned}
[p \boxtimes_m q](x) &= \sum_{i=0}^{m} x^{m-i} (-1)^i \frac{p_i q_i}{\binom{m}{i}} \\
&= \sum_{i=0}^{m} x^{m-i} (-1)^i \binom{m}{i}\binom{m}{i} \frac{\mathbb{E}\{S^i\} \, \mathbb{E}\{T^i\}}{\binom{m}{i}} \\
&= \sum_{i=0}^{m} x^{m-i} (-1)^i \binom{m}{i} \mathbb{E}\{S^i T^i\} \qquad\qquad \text{(independence)} \\
&= \mathbb{E}\{(x - ST)^m\} \, .
\end{aligned}
$$

as required. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

An easy corollary of Lemma 3.4 is scale and translation invariance of the U transform.

Corollary 3.5. *Let T be the U transform of a multiset S. Then the U transforms of*

$$
\{s + k : s \in S\} \qquad and \qquad \{ks : s \in S\}
$$

are

$$
\{t + k : t \in T\} \qquad and \qquad \{kt : t \in T\}
$$

respectively.

Proof. Use Lemma 3.1 with $q = (x - k)^m$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

4 The link between polynomial convolutions and free probability

The goal of this section is to show the relationship between the polynomials convolutions defined in Section 2.1 and the convolutions of free probability defined in Section 1.2. Our approach will be to introduce sequences of transforms (each one an infinite power series), indexed by positive integers, which converge to Voiculescu's R-transform and S-transform when applied to the spectral distributions of Hermitian bounded linear operators. Furthermore, in the case that we have

finite operators (say, $m \times m$ Hermitian matrices), the corresponding power series will linearize the convolutions of polynomials.

Before getting into any details, it is worth considering what to expect out of our transforms. Firstly, we note that, while these transforms will exist for all Hermitian operators and all integers m, we will only try to give meaning to those that that are associated with $m \times m$ matrices (and the appropriate value of m). Since the spectral distribution of an $m \times m$ matrix is completely determined by the first m moments, the same will be true for any function of these moments, including the transforms that we will define. It will nonetheless be computationally beneficial to be able to operate on generic power series which, in our context, will simply have extraneous information (that we can choose to ignore if and when it is useful to us).

Interestingly, the additive and multiplicative convolutions will treat its "storing of moments" in quite different ways, and therefore "extraneous information" will take different forms. In the additive case, the storage proceeds like normal generating functions and so we should expect to get power series for which we are only interested in certain coefficients. In the multiplicative case, the storage will take the form of poles of a rational function (or evaluations of a polynomial, if you prefer) and so we will only be interested in evaluations at certain points. The additive case, in particular, will require some operations that do not obviously preserve such information (as this is not true for general operations). Hence we begin by proving that the operations that will appear can be performed without destroying information.

Let $f(x) = \sum_i a_i x^i$ be a formal power series. For a nonnegative integer k, we will write $f \mod [x^k]$ to denote the polynomial

$$\sum_{i=0}^{k-1} a_i x^i$$

and we will write

$$f \equiv g \mod [x^k]$$

if $f - g \mod [x^k]$ is the 0 polynomial. Many properties can be derived easily from the definition. For example, if

$$a(x) \equiv b(x) \mod [x^k] \qquad \text{and} \qquad c(x) \equiv d(x) \mod [x^k]$$

then we have

$$a(x)c(x) \equiv b(x)d(x) \mod [x^k] \qquad \text{and} \qquad a(x) + c(x) \equiv b(x) + d(x) \mod [x^k]$$

Combining these implies (under the same assumptions), that

$$h(a(x)) \equiv h(b(x)) \mod [x^k]$$

for any power series h. This leads to the following observation, which we state as a corollary:

Corollary 4.1. *Let a, b, h be power series with h invertible (that is, there exists power series g such that $h(g(x)) = x$ for all x. Then*

$$a(x) \equiv b(x) \mod [x^k] \iff h(a(x)) \equiv h(b(x)) \mod [x^k]$$

4.1 Additive convolution

We begin with the additive case.

Definition 4.2. Let A be a Hermitian bounded linear operator with spectral distribution μ_A. For an integer m, we define the *m-finite K-transform of μ_A* to be the power series

$$\mathcal{K}_{\mu_A}^m (s) = -\frac{\partial}{\partial s} \ln \left\| e^{-xs} \Delta^+ (xI - A) \right\|_m. \tag{11}$$

In the case that A is an $m \times m$ matrix, the m-finite K-transform will be the analogue of the inverse Cauchy transform from Voiculescu's theory. As noted in Section 2.4, it will be an infinite power series in the variable s (the x appearing in the formula is integrated out by the norm). It should be clear from the definition that it is invariant under unitary transformations of A and exists for all bounded linear operators A and all integers m. We then define the *m-finite R-transform* by

$$\mathcal{R}_{\mu_A}^m (s) = \mathcal{K}_{\mu_A}^m (s) - \mathcal{K}_{\mu_0}^m (s) \tag{12}$$

where μ_0 is the constant 0 distribution. It is not hard to calculate (and we will do most of it in Lemma 4.4) that

$$\mathcal{K}_{\mu_0}^m (s) = \left(1 + \frac{1}{m} \right) \frac{1}{s}$$

which one can view as the discrete version of the familiar $1/s$ term that is subtracted from the inverse Cauchy transform to obtain the R-transform in Voiculescu's theory.

4.1.1 Relation to free probability

We first show that the our definition is, in fact, a generalization of Voiculescu's R-transform.

Theorem 4.3. *Let A be a Hermitian bounded linear operator with spectral distribution μ_A. Then*

$$\lim_{m \to \infty} \mathcal{K}_{\mu_A}^m (s) = \mathcal{G}_{\mu_A}^{-1} (s)$$

at all points $s \in (\rho_A, \infty)$.

Proof. We first note that the function

$$f(x) = - \ln \Delta^+ (xI - A)$$

exists and is continuous on (ρ_A, ∞). Furthermore, we have

$$f''(x) = \text{tr} \left[(xI - A)^2 \right] > 0$$

and so f is strictly convex. By Corollary 2.4, we therefore have $f'^{-1}(s) = f^{*\prime}(s)$ where

$$f'(x) = -\text{tr} \left[(xI - A)^{-1} \right] = -\mathcal{G}_{\mu_A} (x)$$

and $\mathcal{G}_{\mu_A} (x)$ is the Cauchy transform. Hence

$$f'^{-1}(s) = (-\mathcal{G}_{\mu_A} (x))^{-1} (s) = \mathcal{G}_{\mu_A}^{-1} (-s).$$

Plugging this into Corollary 2.6 gives

$$\mathcal{G}_{\mu_A}^{-1} (s) = f^{*\prime}(-s) = -\frac{\partial}{\partial s} \ln \left\| e^{-xs} \Delta^+ (xI - A) \right\|_\infty$$

The result then follows from Lemma 2.5. $\qquad \square$

14

4.1.2 Relation to polynomial convolutions

We now show the connection between certain finite R-transforms and finite free additive convolutions of polynomials. Our first job will be to find the value of the m-finite R-transform of a distribution on an $m \times m$ real symmetric matrix. The computation will employ the Laplace transform from Section 2.2 as well as the U transform introduced in Section 3.

Lemma 4.4. *If A is an $m \times m$ real symmetric matrix, then*

$$\frac{\|e^{-xs}\Delta^+(xI - A)\|_m^m}{\|e^{-xs}\Delta^+(xI - 0)\|_m^m} \equiv \mathbb{E}\{e^{-msT_A}\} \mod [s^{m+1}]$$

where T_A is the U transform of $\lambda(A)$.

Proof. For an $m \times m$ real symmetric matrix A, we have the simplification

$$\Delta^+(xI - A)^m = \det[xI - A] = \mathbb{E}\{(x - T_A)^m\}.$$

where T_A is the U transform of $\lambda(A)$. Hence

$$\left\|e^{-xs}\Delta^+(xI - A)\right\|_m^m = \int_{\rho_A}^\infty e^{-mxs}\mathbb{E}\{(x - T_A)^m\}\, \mathrm{d}x \qquad\qquad (y := x - \rho_A)$$

$$= e^{-m\rho_A s}\int_0^\infty e^{-mys}\mathbb{E}\{(y + \rho_A - T_A)^m\}\, \mathrm{d}x$$

$$= \mathbb{E}\{e^{-m\rho_A s}\mathcal{L}\{(x + \rho_A - T_A)^m\}(ms)\}$$

where

$$\mathcal{L}\{(x + \rho_A - T_A)^m\}(ms) = \sum_{i=0}^m \binom{m}{i}(\rho_A - T_A)^{m-i}\mathcal{L}\{x^i\}(ms)$$

$$= \sum_{i=0}^m \binom{m}{i}(\rho_A - T_A)^{m-i}\frac{i!}{(ms)^{i+1}}$$

$$= m!\sum_{i=0}^m (\rho_A - T_A)^{m-i}\frac{(ms)^{-i-1}}{(m-i)!}.$$

When A is the 0 matrix, we have $\rho_A = T_A = 0$, so

$$\left\|e^{-xs}\Delta^+(xI - A)\right\|_m^m = \frac{m!}{(ms)^{m+1}}.$$

Hence we can write

$$\frac{\|e^{-xs}\Delta^+(xI - A)\|_m^m}{\|e^{-xs}\Delta^+(xI - 0)\|_m^m} = \frac{(ms)^{m+1}}{m!}\mathbb{E}\{e^{-m\rho_A s}\mathcal{L}\{(x + \rho_A - T_A)^m\}(ms)\}$$

$$= e^{-m\rho_A s}\mathbb{E}\left\{\sum_{i=0}^m (\rho_A - T_A)^{m-i}\frac{(ms)^{m-i}}{(m-i)!}\right\}.$$

where

$$\sum_{i=0}^{m}(\rho_A - T_A)^{m-i}\frac{(ms)^{m-i}}{(m-i)!} \equiv \mathbb{E}\left\{e^{(\rho_A-T_A)ms}\right\} \quad \mathrm{mod}\ [s^{m+1}]$$

Hence

$$\frac{\left\|e^{-xs}\Delta^+(xI-A)\right\|_m^m}{\left\|e^{-xs}\Delta^+(xI-0)\right\|_m^m} \equiv e^{-m\rho_A s}\,\mathbb{E}\left\{e^{(\rho_A-T_A)ms}\right\} \quad \mathrm{mod}\ [s^{m+1}]$$

$$\equiv \mathbb{E}\left\{e^{-msT_A}\right\} \quad \mathrm{mod}\ [s^{m+1}]$$

as claimed. □

Corollary 4.5. *If A is an $m \times m$ real symmetric matrix, then*

$$\mathcal{R}_{\mu_A}^m(s) \equiv -\frac{1}{m}\frac{\partial}{\partial s}\ln\mathbb{E}\left\{e^{-msT_A}\right\} \quad \mathrm{mod}\ [s^m]$$

where T_A is the U transform of $\lambda(A)$.

Proof. We start by unpacking Equation (12):

$$\mathcal{R}_{\mu_A}^m(s) = \mathcal{K}_{\mu_A}^m(s) - \mathcal{K}_{\mu_0}^m(s)$$

$$= -\frac{\partial}{\partial s}\ln\left\|e^{-xs}\Delta^+(xI-A)\right\|_m + \frac{\partial}{\partial s}\ln\left\|e^{-xs}\Delta^+(xI-0)\right\|_m$$

$$= -\frac{1}{m}\frac{\partial}{\partial s}\ln\left(\frac{\left\|e^{-xs}\Delta^+(xI-A)\right\|_m^m}{\left\|e^{-xs}\Delta^+(xI-0)\right\|_m^m}\right).$$

By Lemma 4.4, we have

$$\frac{\left\|e^{-xs}\Delta^+(xI-A)\right\|_m^m}{\left\|e^{-xs}\Delta^+(xI-0)\right\|_m^m} \equiv \mathbb{E}\left\{e^{-msT_A}\right\} \quad \mathrm{mod}\ [s^{m+1}]$$

where T_A is the U transform of $\lambda(A)$. Since ln is invertible, we can apply Corollary 4.1 to get

$$\ln\left(\frac{(ms)^{m+1}}{m!}\left\|e^{-xs}\Delta^+(xI-A)\right\|_m^m\right) \equiv \ln\mathbb{E}\{(s-T_A)^m\} \quad \mathrm{mod}\ [s^{m+1}].$$

This implies the claim, since having two power series match on the first $m+1$ coefficients implies their derivatives match on the first m coefficients. □

Lemma 4.6. *Let A and B be $m \times m$ real symmetric matrices. Then the following are equivalent:*

1. $\mathcal{R}_{\mu_A}^m(s) + \mathcal{R}_{\mu_B}^m(s) \equiv \mathcal{R}_{\mu_{A+B}}^m(s) \quad \mathrm{mod}\ [s^m]$

2. $\det[xI-A]\boxplus_m\det[xI-B] = \det[xI-A-B]$

Proof. Let T_A, T_B, T_{A+B} be the U transforms of $\lambda(A), \lambda(B), \lambda(A+B)$ respectively and where T_A and T_B are treated as independent random variables. By Lemma 3.4, we have that *2.* holds if and only if

$$\mathbb{E}\{(x-T_A-T_B)^m\} = \mathbb{E}\{(x-T_{A+B})^m\}$$

16

which holds if and only if $T_A + T_B$ and T_{A+B} have the same first m moments. This in turn is equivalent to the statement

$$\mathbb{E}\left\{e^{-ms(T_A+T_B)}\right\} \equiv \mathbb{E}\left\{e^{-msT_{A+B}}\right\} \quad \text{mod } [s^{m+1}]$$

which is true if and only if

$$\mathbb{E}\left\{e^{-ms(T_A)}\right\}\mathbb{E}\left\{e^{-ms(T_B)}\right\} \equiv \mathbb{E}\left\{e^{-msT_{A+B}}\right\} \quad \text{mod } [s^{m+1}]$$

since T_A and T_B were chosen to be independent. Since ln is invertible, we can apply to Corollary 4.1 to see that this is equivalent to the statement

$$f_A(s) + f_B(s) \equiv f_{A+B}(s) \quad \text{mod } [s^{m+1}]$$

where $f_X(s) = -\frac{1}{m}\ln \mathbb{E}\left\{e^{-ms(T_B)}\right\}$. So by Corollary 4.5, it remains to show that the two statements

1. $f_A(s) + f_B(s) = f_{A+B}(s) \mod [s^{m+1}]$

2. $\frac{\partial}{\partial s}f_A(s) + \frac{\partial}{\partial s}f_B(s) \equiv \frac{\partial}{\partial s}f_{A+B}(s) \mod [s^m]$

are equivalent. The forward implication is obvious (and was done in the proof of Corollary 4.5). The only issue with the reverse implication, however, is the constant term. Hence we merely need to show that $f_A(0) + f_B(0) = f_{A+B}(0)$, and we would be done. However this follows easily by noting that (by definition) $f_X(0) = 0$ for all X. $\qquad\square$

Note that Lemma 4.6 relied heavily on the fact that A and B were $m \times m$ matrices. This came in the form of the assertion that the first m moments completely characterize the distribution (which of course is not true for more general distributions).

4.2 Multiplicative convolution

We follow the same path as in the additive case. Note that this time we require the spectral distribution to be positive almost surely, just as in Voiculescu's theory.

Definition 4.7. Let A be a positive definite bounded linear operator with spectral distribution μ_A. For an integer m, we define the *m-finite N-transform of μ_A* to be the power series

$$\ln \widetilde{\mathcal{N}}^m_{\mu_A}(s) = -\frac{\partial}{\partial s}\ln \left\| e^{-xs}\Delta^+\left(I - e^{-x}A\right)\right\|_m. \tag{13}$$

Like before, the m-finite N-transform will be the analogue of the inverse M-transform from Voiculescu's theory for the case when A is an $m \times m$ matrix. Again, the invariance under unitary transformations of A should be clear from the definition. We then define the *m-finite S-transform* by

$$\ln \widetilde{\mathcal{S}}^m_{\mu_A}(s) = \ln \widetilde{\mathcal{N}}^m_{\mu_A}(s) - \ln \widetilde{\mathcal{N}}^m_{\mu_I}(s) \tag{14}$$

where I is the identity operator. Note that we have dropped the "modified" adjective since no other such thing exists in this context. This time calculating $\widetilde{\mathcal{N}}^m_{\mu_I}(s)$ is a bit more involved; using the calculation in Lemma 4.10, we get

$$\ln \widetilde{\mathcal{N}}^m_{\mu_I}(s) = -\frac{1}{m}\frac{\partial}{\partial s}\ln \left(\frac{\Gamma(ms)m!}{\Gamma(ms+m+1)}\right) = \psi(ms+m+1) - \psi(ms)$$

17

where $\psi(x) = \frac{\partial}{\partial s} \ln \Gamma(x)$ is the *digamma function*. Once again one can view this as the discrete version of the familiar $(s+1)/s$ term that is multiplied with the inverse M-transform to get the S-transform in Voiculescu's theory. To see this, we can use the standard asymptotic approximation for the digamma function $\psi(x) = \ln(x) + O(1/x)$, giving

$$\lim_{m \to \infty} \widetilde{\mathcal{N}}_{\mu_I}^m (s) = \lim_{m \to \infty} \frac{ms + m + 1}{ms} = \frac{s+1}{s}$$

as expected.

4.2.1   Relation to free probability

We begin by showing that the our definition is, again, a generalization of the modified S-transform (see Section 1.1 for the difference between this and Voiculescu's version).

Theorem 4.8. *Let A be a positive definite bounded linear operator with spectral distribution μ_A. Then*

$$\lim_{m \to \infty} \widetilde{\mathcal{N}}_{\mu_A}^m (s) = \widetilde{\mathcal{M}}_{\mu_A}^{(-1)} (s)$$

at all points $s \in (\ln \rho_A, \infty)$.

Proof. This time we consider the function

$$f(x) = -\ln \Delta^+ \left(I - e^{-x} A \right).$$

$f(x)$ exists and is continuous on the interval $(\ln \rho_A, \infty)$ and we have

$$f'(x) = -\text{tr} \left[A e^{-x} (I - A e^{-x})^{-1} \right] = \text{tr} \left[I - e^x (e^x I - A)^{-1} \right] = 1 - e^x \mathcal{G}_{\mu_A} (e^x) = -\widetilde{\mathcal{M}}_{\mu_A} (e^x)$$

and

$$f''(x) = \text{tr} \left[A e^{-x} (I - A e^{-x})^{-1} \right] + \text{tr} \left[A^2 e^{-2x} (I - A e^{-x})^{-2} \right].$$

In particular, $f''(x) > 0$ for $x > \ln \rho_A$ and so f is strictly convex on this region. By Corollary 2.4, we therefore have

$$f^{*'}(s) = f'^{-1}(s) = \left(-\widetilde{\mathcal{M}}_{\mu_A} (e^x) \right)^{-1} (s) = \ln \widetilde{\mathcal{M}}_{\mu_A}^{(-1)} (-s).$$

Plugging this into Corollary 2.6 gives

$$\ln \widetilde{\mathcal{M}}_{\mu_A}^{(-1)} (s) = f^{*'}(-s) = -\frac{\partial}{\partial s} \ln \left\| e^{-xs} \Delta^+ \left(I - e^{-x} A \right) \right\|_\infty$$

and so the lemma then follows from Lemma 2.5.   □

Note that when A is a real symmetric matrix with eigenvalue r_i, the function $\Delta^+ \left(I - e^{-x} A \right)$ has zeroes exactly when $x = \ln r_i$. This seems to be encoding the fact that the addition of logs is the log of multiplication without actually operating on the log of the operator itself.

4.2.2 Relation to polynomial convolutions

As before, we start by computing the quantities of interest in the case that A is an $m \times m$ real symmetric matrix.

Lemma 4.9. *Let A be an $m \times m$ real positive definite matrix with T_A the U transform of $\lambda(A)$. Then*

$$\frac{\left\|e^{-xs}\Delta^+\left(I - e^{-x}A\right)\right\|_m^m}{\left\|e^{-xs}\Delta^+\left(I - e^{-x}I\right)\right\|_m^m} = \rho_A^{-ms} f_A(s)$$

where $f_A(x)$ is the unique degree m polynomial which satisfies

$$f_A\left(-\frac{k}{m}\right) = \mathbb{E}\left\{\left(\frac{T_A}{\rho_A}\right)^k\right\}$$

for all integers $0 \le k \le m$.

Proof. We expand

$$\left\|e^{-xs}\Delta^+\left(I - e^{-x}A\right)\right\|_m^m = \int_{\ln\rho_A}^\infty e^{-xms}\,\mathbb{E}\left\{(1 - e^{-x}T_A)^m\right\}\,\mathrm{d}x \qquad (y := x - \ln\rho_A)$$

$$= \rho_A^{-ms}\int_0^\infty e^{-yms}\,\mathbb{E}\left\{(1 - e^{-y}T_A/\rho_A)^m\right\}\,\mathrm{d}y$$

$$= \rho_A^{-ms}\mathcal{L}\left\{\mathbb{E}\left\{(1 - e^{-y}T_A/\rho_A)^m\right\}\right\}(ms)$$

where

$$\mathcal{L}\left\{(1 - e^{-x}T_A/\rho_A)^m\right\}(ms) = \sum_{i=0}^m \binom{m}{i}(-T_A/\rho_A)^i\,\mathcal{L}\left\{e^{-ix}\right\}(ms) = \sum_{i=0}^m \binom{m}{i}\frac{(-T_A/\rho_A)^i}{ms + i}.$$

so that

$$\left\|e^{-xs}\Delta^+\left(I - e^{-x}A\right)\right\|_m^m = \rho_A^{-ms}\sum_{i=0}^m \binom{m}{i}\frac{\mathbb{E}\left\{(-T_A/\rho_A)^i\right\}}{ms + i} \tag{15}$$

Now when $A = I$ (the identity), we have $\log\rho_A = 0$ and $T_A = 1$, so that

$$\left\|e^{-xs}\Delta^+\left(I - e^{-x}I\right)\right\|_m^m = \sum_{i=0}^m \binom{m}{i}\frac{(-1)^i}{ms + i}.$$

To find a closed form solution to this sum, we can set

$$f(x) = \sum_{i=0}^m \frac{(-1)^i\binom{m}{i}x^{ms+i}}{ms + i}$$

so that

$$f'(x) = \sum_{i=0}^m (-1)^i\binom{m}{i}x^{ms+i-1} = x^{ms-1}(1 - x)^m.$$

Hence

$$f(1) - f(0) = \int_0^1 x^{ms-1}(1 - x)^m\,\mathrm{d}x = \beta(ms, m + 1)$$

19

where β is the well known *Beta function*. Assuming $s > 0$, we have $f(0) = 0$ and so

$$f(1) = \int_0^1 x^{ms-1}(1-x)^m \, \mathrm{d}x = \frac{\Gamma(ms)\Gamma(m+1)}{\Gamma(ms+m+1)} = m! \frac{\Gamma(ms)}{\Gamma(ms+m+1)}.$$

Hence we have

$$\frac{\|e^{-xs}\Delta^+ (I - e^{-x}A)\|_m^m}{\|e^{-xs}\Delta^+ (I - e^{-x}I)\|_m^m} = \frac{\rho_A^{-ms}}{m!} \frac{\Gamma(ms+m+1)}{\Gamma(ms)} \sum_{i=0}^m \binom{m}{i} \frac{\mathbb{E}\{(-T_A/\rho_A)^i\}}{ms+i}.$$

Since

$$\frac{\Gamma(ms+m+1)}{\Gamma(ms)} = \prod_{i=0}^m (ms+i)$$

the poles that appear in Equation (15) will be eliminated. In particular, we can write

$$f_A(s) := \rho_A^{ms} \frac{\|e^{-xs}\Delta^+ (I - e^{-x}A)\|_m^m}{\|e^{-xs}\Delta^+ (I - e^{-x}I)\|_m^m}.$$

where $f_A(s)$ is a degree m polynomial. One can check then check that

$$f_A\left(-\frac{k}{m}\right) = \frac{1}{m!} \binom{m}{k} \mathbb{E}\{(-T_A/\rho_A)^k\} \prod_{i \neq k}(i-k) = \mathbb{E}\left\{\left(\frac{T_A}{\rho_A}\right)^k\right\}$$

for integers $0 \leq k \leq m$, which therefore uniquely determines it. $\qquad \square$

Note that if one started with a collection of $m+1$ points

$$\left(-\frac{k}{m}, \mathbb{E}\left\{\left(\frac{T_A}{\rho_A}\right)^k\right\}\right),$$

and used the method of Lagrange interpolation to build a degree m polynomial going through those points, the formula would produce exactly $f_A(s)$. Of course it would have to produce $f_A(s)$ in some form (since it is uniquely determined), but it in situations where there were not enough points to completely determine the polynomial, this observation might be useful.

There is a marked difference between Lemma 4.4 (which produces a polynomial characterized by its coefficients) and Lemma 4.9 (which produces a polynomial characterized by evaluations). As a result, the connection between the finite S-transform and the finite free multiplicative convolution is also quite different.

Lemma 4.10. *Let A and B^{-1} be $m \times m$ real positive definite matrices. Then the following are equivalent:*

1. $\int_\alpha^\beta \log \widetilde{\mathcal{S}}_{\mu_A}^m (s) + \log \widetilde{\mathcal{S}}_{\mu_B}^m (s) \, \mathrm{d}s = \int_\alpha^\beta \log \widetilde{\mathcal{S}}_{\mu_{AB}}^m (s) \, \mathrm{d}s$ *for all* $\alpha, \beta \in \{-k/m\}_{k=0}^m$

2. $\det [xI - A] \boxtimes_m \det [xI - B] = \det [xI - AB]$

Proof. Similar to before, let T_A, T_B, T_{AB} be the U transforms of $\lambda(A), \lambda(B), \lambda(AB)$ respectively. Recall that the definition of the finite S-transform implies that the antiderivative of $\ln \widetilde{\mathcal{S}}_{\mu_A}^m(s)$ is

$$g_A(s) := -\log \left\| e^{-xs} \Delta^+ \left(I - e^{-x} A\right) \right\|_m + \log \left\| e^{-xs} \Delta^+ \left(I - e^{-x} I\right) \right\|_m$$

and that Lemma 4.9 implies

$$g_A\left(\frac{-k}{m}\right) = -\frac{1}{m} \log \mathbb{E}\left\{ T_A^k \right\}$$

for $k = 0, \dots, m$. Hence *1.* is equivalent to having

$$g_A(\beta) - g_A(\alpha) + g_B(\beta) - g_B(\alpha) = g_{AB}(\beta) - g_{AB}(\alpha) \tag{16}$$

for all $\alpha, \beta \in \{-k/m\}_{k=0}^m$.

On the other hand Lemma 3.4 shows that *2.* is equivalent to

$$\mathbb{E}\{(x - T_A T_B)^m\} = \mathbb{E}\{(x - T_{AB})^m\}.$$

which, after equating coefficients, is equivalent to having

$$\mathbb{E}\left\{ T_A^k \right\} \mathbb{E}\left\{ T_B^k \right\} = \mathbb{E}\left\{ T_{AB}^k \right\}$$

for all $0 \le k \le m$. Since A, B and AB have nonnegative eigenvalues, one can check that each $\mathbb{E}\{T_X^k\} \ge 0$ for all k, and so *2.* is in fact equivalent to having

$$g_A(\alpha) + g_B(\alpha) = g_{AB}(\alpha) \tag{17}$$

for all $\alpha \in \{-k/m\}_{k=0}^m$.

Clearly Equation (17) implies Equation (16) and so it suffices to prove the converse. However it is easy to check that $g_X(0) = 0$ for any X, so plugging in $\beta = 0$ into Equation (16) gives Equation (17). $\square$

The proofs of Lemma 4.6 and Lemma 4.10 show how the the information concerning moments is being held in a fundamentally different way in the multiplicative case than the additive case. That is, the information regarding additive convolutions seems to live naturally in the coefficients of a power series, whereas the information regarding multiplicative convolutions seems more natural living in the poles of some rational function (or, when normalized, the values of a function at various points). Even in the classical case, the method of forming an additive convolution using the exponential moment generating function is quite popular, but a search through the literature does not provide much in the way of multiplicative analogues. This suggests that the Hadamard product of polynomials (which is, up to scaling, what the multiplicative convolution is doing) is a fundamentally different operation than other more standard power series operations.

It is worth mentioning that, even though the multiplicative convolution seems not to be holding its information directly in the coefficients of the finite S-transform, the coefficients do seem to converge to the S-transform in a manner similar to the finite R-transform (experimentally). Recalling that the coefficients of a power series are simply the values of its derivatives at 0, one could imagine estimating these values by taking small finite differences. But Lemma 4.10 tells us that certain finite differences (ones with endpoints on $-k/m$ for $0 \le k \le m$) are completely determined by the multiplicative convolution. This seems to prevent the coefficients of the finite multiplicative convolutions from varying too much and in turn seems to result in convergence of the coefficients.

21

5   Finite freeness

One of the appealing attributes of free probability is the fact that one can "instantiate" freeness. That is, given two distributions μ_A and μ_B, one can find operators A and B with those spectral distributions such that $\mu_A \boxplus \mu_B = \mu_{A+B}$ [32]. Of course finding such instances is hard (hence the need for convolutions). That said, the mere knowledge that an additive convolution *could* be achieved by simple addition is quite useful, as there is much known about the relationship between the eigenvalues of a sum and the eigenvalues of the summands. For example, one can deduce a trivial bound on the spectral radius of the sum given the addends $\rho_{A+B} \leq \rho_A + \rho_B$, an inequality that is not directly obvious from the definition of the additive convolution.

In this section we will attempt to define a concept of finite freeness. Note that this is the completely opposite ordering of how one would typically develop free probability. Rather than defining freeness and then investigating how operations act with respect to that definition, we defined a collection of operations and will try to characterize the property that leads to them. This is because "the property" is no longer a universal trait and will depend on the operation in question. This is both a blessing and a curse: on the one hand it means that we will not have the convenience of just calling things "free" and then doing arbitrary things to them. On the other hand, it will allow us to apply our theory to operators that satisfy much a weaker constraint than "freeness". In particular, we will be able to "instantiate" finite freeness using (computable) matrices, something that was not possible in free probability.

We should note that we will intentionally blur the lines between polynomials and (classes of) real symmetric matrices in this section. Both will be associated to a multiset, either via its roots (in the case of a polynomial) or its eigenvalues (in the case of matrices). In this regard, the convolutions defined in Section 2.1 can be extended to multisets by operating on the monic polynomials with the elements of that multiset as roots. Equivalently, it can be extended to a (rotation invariant) operation on (classes of) real symmetric matrices by operating on the characteristic polynomials of those matrices. Because all of the convolutions preserve real rootedness [18], it is plausible that they could coincide with classical matrix operations, and this will fuel our definition of finite freeness.

Definition 5.1. We will say two $m \times m$ real symmetric matrices A and B are *in finite free position* (or simply *m-free*) if

$$\det[xI - yA - zB] = \det[xI - yA] \boxplus_m \det[xI - zB]$$

for all $y, z \in \mathbb{R}$.

Remark 5.2. Because of the heightened role that additive convolution plays in probability theory (for example, in many of the limit theorems in Section 6), we will take a noticably "additive centric" view. This is done to avoid having to resort to conventions like "finitely multiplicatively free," but whether such terms are inevitable is another question. So, at least in regards to this paper, the term "free" will be used in association with an additive property.

Also worth noting is that, at least with respect to the additive convolution, our definition is stronger than it needs to be (requiring the equality to work on all real numbers y and z rather than just $y = z = 1$). It turns out that the increased structure will make more sense when applied to the multiplicative convolution, but it would be interesting to explore the effects of weakening these conditions.

To show that finite freeness exists, we will use some results in the theory of hyperbolic polynomials. Given a vector $\vec{e}$, a homogeneous polynomial p is said to be *hyperbolic with respect to $\vec{e}$* if $p(\vec{e}) \neq 0$ and the univariate polynomial $\hat{p}(t) = p(\vec{x}t\vec{e})$ has only real roots for all points x. The canonical example of a hyperbolic polynomial is the determinant acting on the space of real symmetric matrices: the determinant is hyperbolic with respect to the identity matrix since $\det [tI - X]$ is real rooted for all symmetric matrices X (the roots then being simply the eigenvalues). An extremely useful characterization of hyperbolic polynomials [15, 14]:

Theorem 5.3. *A polynomial p on $\mathbb{R}^3$ is hyperbolic of degree m with respect to the vector $\vec{e} = (1, 0, 0)$ and satisfies $p(\vec{e}) = 1$ if and only if there exist $m \times m$ real symmetric matrices B, C such that*

$$p(x, y, z) = \det [xI + yB + zC].$$

A corollary to Theorem 5.3 is a rather strong statement concerning the existence of matrices in finite free position.

Lemma 5.4. *For any real symmetric matrices A, B there exists a rotation matrix R such that A and $R^T B R$ are in finite free position.*

Proof. Consider the polynomial

$$p(x, y, z) = \det [xI - yA] \boxplus_m \det [xI - zB]$$

We first show that p is hyperbolic with respect to the vector $\vec{e} = (1, 0, 0)$.

Clearly $p(1, 0, 0) \neq 0$ and so it remains to show that $\hat{p}(x_0 - t, y_0, z_0)$ is real rooted for all fixed values x_0, y_0, z_0. Plugging in, we get

$$\begin{aligned} \hat{p}(x_0 - t, y_0, z_0) &= \det [(x_0 - t)I - y_0 A] \boxplus_m \det [(x_0 - t)I - z_0 B] \\ &= (-1)^m \det \left[tI + \hat{A} \right] \boxplus_m (-1)^m \det \left[tI + \hat{B} \right] \end{aligned} \qquad (18)$$

where

$$\hat{A} = y_0 A - x_0 I \qquad \text{and} \qquad \hat{B} = z_0 B - x_0 I$$

are both Hermitian. However, the fact that 18 is always real rooted when A and B are Hermitian was shown in [18] (as noted in Section 2.1).

Hence p meets the requirements for Theorem 5.3, which means it can be written in the form

$$p(x, y, z) = \det [xI + yU + zV]$$

for some matrices U and V. Since $p(x, 1, 0) = \det [xI - A]$, it must be that $\lambda(U) = \lambda(-A)$, and similarly, $\lambda(V) = \lambda(-B)$. In particular there exists rotations P and Q such that

$$P^T U P = -A \qquad \text{and} \qquad Q^T V Q = -B$$

Let $R = Q^T P$. Then we have

$$\begin{aligned} p(x, y, z) &= \det [xI + yU + zV] \\ &= \det \left[xI + yU - QBQ^T \right] \det \left[P^T P \right] \\ &= \det \left[xI - yA - P^T QBQ^T P \right] \\ &= \det \left[xI - yA - R^T B R \right] \end{aligned}$$

which is exactly the statement that A and $R^T B R$ are in finite free position. $\square$

Equation (4) then provides a collection of identities that characterize matrices in finite free position. The decomposition of determinants into mixed discriminants shown by Equation (9) allow us to state these identities explicitly:

Lemma 5.5. *For two $m \times m$ real symmetric matrices A and B, the following are equivalent:*

1. *A and B are in finite free position*

2. *$m! \, \mathrm{D}\left(A[j], B[i], I[m-j-i]\right) = \mathrm{D}\left(A[j], I[m-j]\right) \mathrm{D}\left(B[i], I[m-i]\right)$ for all i, j.*

Proof. Simply expand the relation

$$\det\left[xI - yA - zB\right] = \det\left[xI - yA\right] \boxplus_m \det\left[xI - zB\right]$$

using Equation (4) and equate coefficients. $\square$

In general, the rotation guaranteed by Lemma 5.4 may not be unique. Since determinants are invariant under rotations, we have

$$p(x, y, z) = \det\left[xI - yA - zB\right] = \det\left[R^T(xI - yA - zB)R\right] = \det\left[xI - yR^T AR - zR^T BR\right]$$

and so any rotation R that leaves A unchanged provides a new rotation of B that is in finite free position with A. The extreme case of this is when $A = I$ (the identity), which is in finite free position with every matrix. Since I is freely independent (in the Voiculescu sense) from all operators, this should be of little surprise. However, this fact has useful consequences:

Corollary 5.6. *If A and B are in finite free position, then $sA + tI$ and $uB + vI$ are in finite free position for any $s, t, u, v \in \mathbb{R}$.*

Proof. Follows directly from the definition. $\square$

Despite the lack of uniqueness, it would be useful to characterize any property of finite free position that can be guaranteed (either by all or by some rotation). The next lemma is a step in that direction:

Lemma 5.7. *Let A and B be $m \times m$ real symmetric matrices with A diagonal. Then there exists a rotation matrix R such that A and $R^T BR$ are in finite free position **and** the diagonal of $R^T BR$ is constant.*

Proof. By Corollary 5.6, the statement is true for all pairs (A, B) if and only if is true for all pairs $(A, B - \mathrm{tr}\,[B]\,I)$, so without loss of generality, we can assume $\mathrm{tr}\,[B] = 0$. Let Q be the rotation guaranteed by Lemma 5.4. For any matrix X, we have the relation $\mathrm{D}\left(X, I[m-1]\right) = m! \mathrm{tr}\,[X]$, so Lemma 5.5 implies

$$m! \, \mathrm{D}\left(A[j], Q^T BQ, I[m-j-1]\right) = \mathrm{D}\left(A[j], I[m-j]\right) \mathrm{D}\left(Q^T BQ, I[m-1]\right) = 0$$

for all j. In particular, taking various linear combinations of these gives

$$\mathrm{D}\left((xI - A)[m-1], Q^T BQ\right) = 0$$

24

for all values of x. Letting $p(x) = \det[xI - A]$, this gives

$$p(x) \sum_i \frac{b_i}{x - a_i} = 0$$

where b_i is the ith diagonal entry of $R^T BR$ (and same for a_i). Since the set of values for which $p(x) = 0$ is finite, continuity implies that

$$\sum_i \frac{b_i}{x - a_i} = 0$$

for all x.

Note that for each distinct value of $u \in \{a_i\}$, we can choose x sufficiently close to u such that the terms with $1/(x - u)$ dominate the other terms. Thus for the sum to be 0, it must be that the set of b_i for which $a_i = u$ sum to zero (in the case that there is a unique a_i, then the corresponding b_i must equal 0).

Recall from the previous discussion that, given A and B in finite free position, we have that A and $S^T BS$ are in finite free position for any rotation S such that $S^T AS = A$. This is exactly the set of rotations that act independently on the eigenspaces of A, which (because A Is diagonal) are the set of rotations that act on the submatrices of A for which the diagonal entries are the same. Thus we can pick a rotation S that averages the elements of the diagonal of B corresponding to a single eigenspace of A. Thus the rotation $R = QS$ places A and $R^T BR$ in finite free position and causes $R^T BR$ to have a 0 diagonal. $\qquad\square$

Lemma 5.7 can be quite a bit more useful than simple "instantiation". This is due to a collection of so-called *majorization* inequalities that relate the diagonal entries of a matrix to its eigenvalues. We refer the interested reader to [22].

We now observing the analogous statement to Lemma 5.5 for the multiplicative convolution:

Lemma 5.8. *For two $m \times m$ real symmetric matrices A and B, the following are equivalent:*

1. $\det[xI - AB] = \det[xI - A] \boxtimes_m \det[xI - B]$

2. $m! \, \mathrm{D}(AB[i], I[m - i]) = \mathrm{D}(A[i], I[m - i]) \, \mathrm{D}(B[i], I[m - i])$ *for all i.*

Proof. Simply expand the relation

$$\det[xI - AB] = \det[xI - A] \boxtimes_m \det[xI - B]$$

using Equation (5) and equate coefficients. $\qquad\square$

Hence we have the following relationship between finite freeness and the multiplicative convolution:

Corollary 5.9. *If B is invertible and A and B^{-1} are in finite free position, then*

$$\det[xI - AB] = \det[xI - A] \boxtimes_m \det[xI - B]$$

Proof. For A and B^{-1} in finite free position, we have by Lemma 5.5

$$m! \, \mathrm{D}\left(A[i], B^{-1}[m - i]\right) = \mathrm{D}(A[i], I[m - i]) \, \mathrm{D}\left(B^{-1}[m - i], I[i]\right)$$

for all i. Multiplying both sides by $\det[B]$ gives

$$m! \, \mathrm{D}(AB[i], I[m - i]) = \mathrm{D}(A[i], I[m - i]) \, \mathrm{D}(B[i], I[m - i])$$

which is precisely what is needed by Lemma 5.8. $\qquad\square$

5.1 Comparison to Voiculescu's freeness

To compare finite freeness with Voiculescu's version, it is instructive to see what is implied by the latter which is not implied by the former. For example, if operators A and B are freely independent (in the Voiculescu sense), then

1. $\mu_A \boxplus \mu_B = \mu_{A+B}$

2. $\mu_A \boxtimes \mu_B = \mu_{AB}$

3. $f(\mathcal{A})$ and $g(\mathcal{B})$ are freely independent for any functions f, g.

The first statement is true for finite freeness (by construction), but the second and third are not true in general. This should not come as a huge surprise, as the requirements for a matrix to instantiate additive convolution (Lemma 5.5) and multiplicative convolution (Lemma 5.8) are different. Furthermore, different instantiations of finite freeness can give different results when multiplied, as the next example shows:

Example 5.10. Consider

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \qquad \text{and} \qquad B = \begin{pmatrix} 2 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 2 \end{pmatrix},$$

both of which have characteristic polynomial $p_A(x) = p_B(x) = x^3 - 6x^2 + 11x - 6$. On one hand, we have

$$[p_A \boxplus_3 p_B](x) = x^3 - 12x^2 + 46x - 56 \qquad \text{and} \qquad [p_A \boxtimes_3 p_B](x) = x^3 - 12x^2 + \frac{121}{3}x - 36$$

while on the other hand, we have

$$\det[xI - A - B] = x^3 - 6x^2 + 11x - 6 \qquad \text{and} \qquad \det[xI - AB] = x^3 - 12x^2 + 41x - 36$$

Hence A and B are in free position, but do not instantiate the multiplicative convolution. We could also have taken

$$B^+ = \begin{pmatrix} 2 & \sqrt{2}/2 & 0 \\ \sqrt{2}/2 & 2 & \sqrt{2}/2 \\ 0 & \sqrt{2}/2 & 2 \end{pmatrix} \qquad \text{or} \qquad B^- = \begin{pmatrix} 2 & \sqrt{2}/2 & 0 \\ \sqrt{2}/2 & 2 & -\sqrt{2}/2 \\ 0 & -\sqrt{2}/2 & 2 \end{pmatrix}$$

which give

$$\det\left[xI - A - B^+\right] = \det\left[xI - A - B^-\right] = x^3 - 6x^2 + 11x - 6$$

and

$$\det\left[xI - AB^+\right] = \det\left[xI - AB^-\right] = x^3 - 12x^2 + 40x - 36$$

On the other hand, there are special cases where the additive and multiplicative convolutions do coincide. One such situation occurs when one of the matrices is a projection. The proofs are heavy in computation and are not particularly enlightening, so we have separated them out into Section 5.2. The result of these computations, however, leads to the following corollary:

Corollary 5.11. *If A and B are $m \times m$ real symmetric matrices in finite free position with $A^2 = A$ then*

$$\det[xI - AB] = \det[xI - A] \boxtimes_m \det[xI - B]$$

Proof. The fact that A and B are in finite free position and $A^2 = A$ are exactly the hypotheses in Lemma 5.13. Hence we have

$$m! \, \mathrm{D}\left(A[i], B[j], AB[k], I[m-i-j-k]\right) = \mathrm{D}\left(A[i+k], I[m-i-k]\right) \mathrm{D}\left(B[j+k], I[m-j-k]\right).$$

for all i, j, k with $i + j + k \leq m$. The case $i = j = 0$, in particular, gives

$$m! \, \mathrm{D}\left(AB[k], I[m-k]\right) = \mathrm{D}\left(A[k], I[m-k]\right) \mathrm{D}\left(B[k], I[m-k]\right)$$

for all k, which by Lemma 5.8 suffices to prove the lemma. $\qquad\square$

5.2 Computations

The goal of this section is to prove Lemma 5.13 below. We first prove a combinatorial identity that is an easy consequence of the extension to the binomial theorem:

$$(x+1)^{-k} = \sum_i \binom{-k}{i} x^i = \sum_i (-1)^i \binom{k+i-1}{i}.$$

Lemma 5.12. *Let j, k, n be nonnegative integers such that $j + k \leq n$. Then*

$$\sum_{t=0}^{k} (-1)^t \binom{k}{t} \binom{n-t}{n-j-k-t} = \binom{n-k}{n-j-k}$$

Proof. We expand

$$(x+1)^k = \sum_i x^i \binom{k}{i}$$

and

$$(x+1)^{-j-k-1} = \sum_i (-1)^i x^i \binom{j+k+i}{i}.$$

Hence the coefficient of x^{n-j-k} in the product is

$$\sum_t \binom{k}{t} (-1)^{n-j-k-t} \binom{n-t}{n-j-k-t}.$$

On the other hand, we can multiply first and then expand to get

$$(x+1)^{-j-1} = \sum_i (-1)^i x^i \binom{j+i-1}{i}$$

so that the coefficient of x^{n-j-k} is

$$(-1)^{n-j-k} \binom{n-k}{n-j-k}.$$

The lemma then follows by equating the two representations for the same coefficient. $\qquad\square$

We now have the tools to prove the lemma. For $m \times m$ matrices, we define the quantity

$$f(i, j, k) = D(A[i], B[j], AB[k], I[m - i - j - k])$$

Lemma 5.13. *Let A and B be $m \times m$ matrices that satisfy the identities*

$$m! \, f(i, j, 0) = f(i, 0, 0)f(0, j, 0) \tag{19}$$

for all i, j and such that $A^2 = A$. Then

$$m! \, f(i, j, k) = f(i + k, 0)f(0, j + k, 0) \tag{20}$$

Proof. Let $\mathrm{Tr}\,[A] = a$. Without loss of generality, we can assume A is diagonal (with a 1s and $(m-a)$ 0s on the diagonal). Hence we can use the formula in Equation (10) to calculate $D(A[i], I[m - i])$ by counting the number of permutations for which the first i elements are at most a. That is,

$$f(i, 0, 0) = (a)(a - 1) \ldots (a - i + 1)(m - i) \ldots (1) = \frac{a!(m - i)!}{(a - i)!}. \tag{21}$$

Hence Equation (20) is equivalent to

$$f(i, j, k) = f(0, j + k, 0)\frac{a!(m - i - k)!}{m!(a - i - k)!}. \tag{22}$$

Using Lemma 2.9, we have the relation

$$af(i, j, k) = if(i, j, k) + jf(i, j - 1, k + 1) + kf(i, j, k) + (m - i - j - k)f(i + 1, j, k),$$

which, after rearranging and shifting indices $(j \to j + 1, k \to k - 1)$ gives

$$(j + 1)f(i, j, k) = -(a - i - k + 1)f(i, j + 1, k - 1) + (m - i - j - k)f(i + 1, j + 1, k - 1). \tag{23}$$

If we define the quantity $g(i, j, k)$ as

$$f(i, j, k) = (-1)^i \frac{j!(m - i - j - k)!}{(a - i - k)!}g(i, j, k)$$

then substituting gives

$$g(i, j, k) = g(i, j + 1, k - 1) + g(i + 1, j + 1, k - 1). \tag{24}$$

We claim that Equation (24) implies

$$g(i, j, k) = \sum_{t=0}^{k} \binom{k}{t} g(i + t, j + k, 0). \tag{25}$$

This is trivially true for the case $k = 0$, so we can assume it to be true for $k = K - 1$ with $K > 1$ and consider the case $k = K$. Using the relation, we have

$$g(i, j, k) = g(i, j + 1, k - 1) + g(i + 1, j + 1, k - 1)$$

which by the inductive hypothesis means

$$g(i,j,K) = \sum_{t=0}^{k-1} \binom{k-1}{t} g(i+t, j+k, 0) + \sum_{t=0}^{k-1} \binom{k-1}{t} g(i+t+1, j+k, 0)$$

$$= \sum_{t=0}^{k-1} \binom{k-1}{t} g(i+t, j+k, 0) + \sum_{t=1}^{k} \binom{k-1}{t-1} g(i+t, j+k, 0)$$

$$= \sum_{t=0}^{k} g(i+t, j+k, 0)$$

where the last equality uses the identity $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$. Combining Equations (23) and (25) give

$$f(i,j,k) = \sum_{t=0}^{k} \binom{k}{t} (-1)^t \frac{j!(m-i-j-k)!(a-i-t)!}{(j+k)!(m-i-j-k-t)!(a-i-k)!} f(i+t, j+k, 0).$$

Now Equations (19) and (21) imply

$$f(i+t, j+k, 0) = \frac{1}{m!} f(i+t, 0, 0) f(0, j+k, 0) = \frac{a!(m-i-t)!}{m!(a-i-t)!} f(0, j+k, 0)$$

so that

$$f(i,j,k) = a! f(0, j+k, 0) \sum_{t=0}^{k} \binom{k}{t} (-1)^t \frac{j!(m-i-j-k)!(m-i-t)!}{m!(j+k)!(m-i-j-k-t)!(a-i-k)!}$$

$$= f(0, j+k, 0) \frac{a! j!(m-i-j-k)!}{m!(a-i-k)!} \sum_{t=0}^{k} (-1)^t \binom{k}{t} \binom{m-i-t}{m-i-j-k-t}.$$

Using Lemma 5.12 reduces this to

$$f(i,j,k) = f(0, j+k, 0) \frac{a! j!(m-i-j-k)!}{m!(a-i-k)!} \frac{(m-i-k)!}{(m-i-k-j)! j!} = f(0, j+k, 0) \frac{a!(m-i-k)!}{m!(a-i-k)!}$$

which equals (22) as required. $\qquad\qquad\square$

6 Applications

In this final section, we explore some of the consequences of the definition of the m-finite R-transform. There are number of "free" versions of classical distributions, for example the role of "free Gaussian" is played by the semicircle law and the role of "free Poisson" distribution is played by the Marchenko–Pastur law [16]. The goal then is to try to derive the finite free version of such distributions and then to show they behave (with respect to the free additive convolution) in the way they should. In particular, we will focus on additive limit theorems associated with these distributions, such as the central limit theorem and Poisson limit theorem. To do so, we will need to compute additive convolutions, and so we begin by proving some computational properties.

6.1 Properties of the symmetric additive convolution

Recall from the Section 2.1, that we have the following formula for the additive convolution of two degree m polynomials:

$$[p \boxplus_m q](x) = \sum_{i+j \leq m} x^{m-i-j}(-1)^{i+j} \frac{(m-i)!(m-j)!}{(m-i-j)!m!} p_i q_j. \qquad (26)$$

While Equation (26) was (in that case) only defined for polynomials of degree m, such a formula can be applied to any polynomials. In particular, we will consider the case when p and q have degree *at most m*.

Three useful observations can be made directly from the formula. The first is linearity: for any polynomials p, q, r and any constant α, we have

$$[p \boxplus_m (\alpha q + r)] = \alpha[p \boxplus_m q] + [p \boxplus_m r]$$

The second is the observation that when $\deg(f) = m$, we have $[f \boxplus_m x^m](y) = f(y)$. Third is the observation that $[p \boxplus_m q] = [q \boxplus_m p]$.

We now write the formula in Equation (4) in a slightly different form:

$$[p \boxplus_m q](x) = \frac{1}{m!} \sum_{i=0}^{m} p^{(i)}(x) q^{(m-i)}(0) \qquad (27)$$

where $p^{(i)}$ denotes the ith derivative of p. In this form, the following lemma is almost immediate:

Lemma 6.1. *Let $R = \sum_i a_i \frac{\partial^i}{(\partial x)^i}$ be a linear differential operator. Then*

$$[R\{p\} \boxplus_m q](x) = [p \boxplus_m R\{q\}](x) = R\{[p \boxplus_m q](x)\}$$

Proof. By linearity, it suffices to prove this for the operator $R = \frac{\partial^k}{(\partial x)^k}$.

$$\left[p^{(k)} \boxplus_m q\right](x) = \frac{1}{m!} \sum_{i=0}^{m} p^{(i+k)}(x) q^{(m-i)}(0) = \frac{\partial^k}{(\partial x)^k} \frac{1}{m!} \sum_{i=0}^{m} p^{(i)}(x) q^{(m-i)}(0) = \frac{\partial^k}{(\partial x)^k} [p \boxplus_m q](x).$$

By the same argument,

$$\frac{\partial^k}{(\partial x)^k} [p \boxplus_m q](x) = [q^{(k)} \boxplus_m p](x) = [p \boxplus_m q^{(k)}](x)$$

where the last equality is due to the commutativity that was observed earlier. $\qquad \square$

Lemma 6.1 gives an effective way to compute the symmetric additive convolution. As an example, we give the following corollary proving that, in the space of degree m polynomials, the symmetric additive convolution is invertible:

Corollary 6.2. *For any degree m polynomial p, there exists a degree m polynomial q such that $[p \boxplus_m q] = x^m$.*

Proof. Since p has degree m, we can write $p(x) = R\{x^m\}$ where R is a linear differential operator with a nonzero identity term. Viewing this as a power series, we can compute the formal (multiplicative) inverse R^{-1} and then compute $q = R^{-1}\{x^m\}$. As a result, we have

$$[p \boxplus_m q](x) = [R\{x^d\}, R^{-1}\{x^d\}](x) = R\{R^{-1}\{[x^m \boxplus_m x^m]\}\} = x^m$$

$\square$

Last but not least, it is not hard to show that $\boxplus_m$ is associative (and therefore forms an algebra on polynomials degree m).

Lemma 6.3. *Let p, q, r be polynomials. Then*

$$[p \boxplus_m [q \boxplus_m r]] = [[p \boxplus_m q] \boxplus_m r]$$

Proof. A quick calculation shows that both are equal to

$$\frac{1}{(m!)^2} \sum_{i=0}^{m} \sum_{j=0}^{m-i} f^{(i)}(x) g^{(j)}(0) h^{(m-i-j)}(0)$$

$\square$

The algebra formed by $\boxplus_m$ is isomorphic to the algebra of truncated polynomials $\mathbb{C}[x]/\langle x^{m+1} \rangle$ under multiplication. This can easily be seen by considering the homomorphism ϕ taking a polynomial p to the differential operator R such that $R\{x^m\} = p$. As noted in [18], the symmetric additive convolution of two real rooted polynomials is another real rooted polynomial. This forms a cone in the algebra of truncated polynomials that could be interesting in its own right.

6.2 Distributions and Limit Theorems

The approach we will take to finding the finite free versions of various distributions is to reverse engineer them from their R-transforms. That is, for a given R-transform, we would like to find the degree m polynomial whose m-finite R-transform matches the given R-transform (on the first m coefficients). The introduces a slight technicality; the reverse procedure is not unique, since any constant multiple of a polynomial has the same roots that the original polynomial did. Hence we will introduce the notation $p \approx q$ to denote that p and q have the same roots (or $p = cq$ for some $c \neq 0$, if you prefer). The next lemma comes directly from the definitions in Section 4.1:

Lemma 6.4. *Let $Q(x)$ be a polynomial. Then*

$$Q\left(\frac{D_x}{m}\right)\{x^m\} \approx \det[xI - A] \iff \frac{1}{m}\frac{Q'(x)}{Q(x)} \equiv -\mathcal{R}_{\mu_A}^m(x) \mod [x^{m+1}]$$

Proof. By Corollary 4.5, we have

$$\mathcal{R}_{\mu_A}^m(s) \equiv -\frac{1}{m}D_s \ln \mathbb{E}\{e^{-msT_A}\} \mod [s^m]$$

and so

$$\frac{1}{m}\frac{Q'(x)}{Q(x)} \equiv -\mathcal{R}_{\mu_A}^m(x) \mod [x^{m+1}]$$

31

if and only if

$$Q(x) \approx \mathbb{E}\{e^{-msT_A}\} \quad \mathrm{mod}\ [x^{m+1}]$$

where T_A is the U transform of $\lambda(A)$. We note that for any scalar t and any polynomial $p(x)$, we have by Taylor's theorem

$$p(x+t) = \sum_i \frac{t^i}{i!} \frac{\partial^i}{(\partial x)^i}\{p(x)\} = e^{tD_x}\{p(x)\}.$$

Hence

$$Q\left(\frac{D_x}{m}\right)\{x^m\} \approx \mathbb{E}\{e^{-T_A D_x}\}\{x^m\} = \mathbb{E}\{(x - T_A)^m\}$$

which is $\det[xI - A]$ by definition of T_A. $\qquad\square$

It is worth noting that, for our purposes, we do not need to worry about the restriction to m coefficients, as this comes automatically from taking derivatives of x^m. We also note that the constant of integration (that will come when we integrate) is precisely the cause of the ambiguity that necessitated the use of $\approx$ (and so can safely be ignored).

The rest of this section will be the derivation of polynomials corresponding to known distributions and then proofs of their associated limit laws. The proofs will use the polynomials directly (and the techniques of Section 6.1) and not the R-transforms — something that is not possible in traditional free probability. Of course (in these cases) we know the R-transform, so it is somewhat backwards to find the polynomials and use them when we could simply use the R-transforms. The goal, however, is to prove that such a proof is possible, as there are (many) instances where one can not compute the R-transform from a given distribution, and so we anticipate that the techniques displayed below will serve useful in such scenarios.

6.2.1   Constant

The constant random variable is characterized by having first cumulant nonzero and all other cumulants zero, so that

$$\mathcal{R}^m_{\mu_A}(s) = \mu$$

for μ a constant. Using Lemma 6.4, we can compute

$$\ln Q(x) = -\mu m \int \mathrm{d}x = -x\mu m + c$$

for some constant c. Hence

$$Q\left(\frac{D_x}{m}\right)\left\{\frac{x^m}{m!}\right\} \approx e^{-\mu D_x}\{x^m\} = (x - \mu)^m.$$

This is precisely what we would hope, as

$$(x - \mu)^m = \det[xI - \mu I]$$

which is exactly the random matrix version of adding a constant random variable.

The limit theorem associated with the constant distribution is the *law of large numbers*. The classical version of this theorem states that if X_i are independent random variables with $\mathbb{E}\{X_i\} = \mu$, then

$$\frac{1}{n}\sum_{i=0}^{n} X_i \xrightarrow{n\to\infty} \mu$$

almost surely. Here we give an m-finite version of this result:

Theorem 6.5 (Law of large numbers)**.** *Let $p_1, p_2, \ldots$ be a sequence of degree m real rooted polynomials with*

$$p_i = \prod_j (x - r_{i,j}) \qquad and \qquad \frac{1}{m}\sum_j r_{i,j} = \mu$$

for all i. Assume further that

$$\frac{1}{m}\sum_j r_{i,j}^2 < C$$

for some constant C. Then

$$\lim_{n\to\infty} [q_1 \boxplus_m \cdots \boxplus_m q_n] = (x - \mu)^m$$

where $q_i(x) = n^{-m}p_i(nx)$.

Proof. For fixed i, we can write $p_i(x) = x^m + a_1 x^{m-1} + \ldots$. Since a_1 is the sum of the roots, we have (by the hypotheses) $a_1 = m\mu$. So if P_i is the linear differential operator such that $P_i\{x^m\} = p_i(x)$, we have

$$P_i = 1 + \frac{a_1}{m}D_x + \cdots = 1 + \mu D_x + \ldots$$

Now let Q_i be the differential operator such that $Q_i\{x^m\} = q_i(x)$ Then one can check that

$$Q_i = 1 + \frac{a_1}{mn}D_x + O\left(n^{-2}\right) = 1 - \frac{\mu}{n}D_x + O\left(n^{-2}\right)$$

and so

$$[q_1 \boxplus_m \cdots \boxplus_m q_n] = \left(\prod_{i=1}^{n} Q_i\right)\{x^m\} = \left(1 - \frac{\mu}{n}D_x + O\left(n^{-2}\right)\right)^n \{x^m\}$$

which converges to

$$e^{-\mu D_x}\{x^m\} = (x - \mu)^m$$

as $n \to \infty$. $\qquad\square$

6.2.2 Gaussian

The Gaussian random variable is characterized by having first two cumulants nonzero and all other cumulants zero. Given that we know how the constant random variable behaves, it suffices to consider the case when only the second cumulant is nonzero, so that

$$\mathcal{R}_{\mu_A}^m(s) = s\theta$$

for θ a constant. The computation is similar to the constant random variable, as we get

$$\ln Q(x) = -\theta m \int x \, dx = -\frac{m\theta}{2}x^2 + c$$

so that

$$Q\left(\frac{D_x}{m}\right)\{x^m\} \approx e^{-\theta x^2/(2m)}\{x^m\} \approx H_m\left(\sqrt{\frac{m}{\theta}}x\right)$$

where $H_m(x)$ is a Hermite polynomial. This leads to the following definition: For real numbers μ and σ, let

$$\mathfrak{N}_m[\mu, \sigma^2](x) := \left(\frac{m-1}{\sigma^2}\right)^{-m/2} H_m\left((x-\mu)\sqrt{\frac{m-1}{\sigma^2}}\right) \tag{28}$$

Firstly, we remark that the constant in front is merely to make the polynomial monic (it is irrelevant, since all that matters are the roots). Secondly, we remark that it should come as no surprise that the corresponding polynomial should be the Hermite polynomial, since the root distribution must converge asymptotically to the semicircle law (the free probability version of the Gaussian), and it is well known that the roots of the Hermite polynomials do exactly that [23].

Lastly, we want to comment on the factor of $m - 1$ (rather than m from the derivation above). Recall that the coefficients of the m-finite R-transform *converge* to the true R-transform. Apart from the first coefficient, however, they do not match exactly. In particular, an operator X with mean and variance τ and σ^2 will have R-transform

$$\mathcal{R}_{\mu_X}(s) = \tau + \sigma^2 s + \dots$$

and m-finite R-transform

$$\mathcal{R}_{\mu_X}(s) = \tau + \left(1 - \frac{1}{m}\right)\sigma^2 s + \dots$$

Hence by working backwards from $\kappa_2 = 1$, we were mistakenly giving the distribution a variance of $\frac{m}{m+1}$.

Remark 6.6. Note that there are two distinct scalings of Hermite polynomials in the literature. The first (which is used in [31]), defines Hermite polynomials as the polynomials which are orthogonal to the weight function e^{x^2}. The second (which we will use), defines Hermite polynomials as the polynomials which are orthogonal to the Gaussian weight function $e^{x^2/2}$. This seemed to be the more appropriate choice, given the role that the Hermite polynomials are playing in this context (as the "Gaussian" of finite free probability).

The limit theorem associated with the Gaussian is the *central limit theorem*. The classical version of this theorem says that for X_i i.i.d. random variables $\mathbb{E}\{X_i\} = 0$ and $\text{Var}[Xi] = \sigma^2 < \infty$, we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i \xrightarrow{n\to\infty} N(0, \sigma^2).$$

in distribution (where $N(\mu, \sigma^2)$ is the Gaussian distribution). Here we give a proof in the m-finite case:

Theorem 6.7 (Central limit theorem)**.** *Let $p_1, p_2, \dots$ be a sequence of degree m real rooted polynomials with $p_i = \prod_j(x - r_{i,j})$ such that*

$$\sum_j r_{i,j} = 0 \qquad and \qquad \frac{1}{m}\sum_j r_{i,j}^2 = \sigma^2$$

for all i. Then

$$\lim_{n\to\infty} [q_1 \boxplus_m \cdots \boxplus_m q_n] = \mathfrak{N}_m[0, \sigma^2]$$

where $q_i(x) = n^{-m/2} p_i(\sqrt{n}x)$.

Proof. For fixed i, we can write $p_i(x) = x^m + a_1 x^{m-1} + a_2 x^{m-2} + \dots$. Since a_1 is the sum of the roots and a_2 is the sum of the pairwise products of roots, we have (by the hypotheses)

$$a_1 = 0 \qquad \text{and} \qquad a_1^2 - 2a_2 = m\sigma^2$$

so that $a_2 = -m\sigma^2/2$. So if P_i is the linear differential operator such that $P_i\{x^m\} = p_i(x)$, we have

$$P_i = 1 + \frac{a_1}{m} D_x + \frac{a_2}{m(m-1)} D_x^2 + \cdots = 1 - \frac{\sigma^2}{2(m-1)} D_x^2 + \dots$$

Now let Q_i be the differential operator such that $Q_i\{x^m\} = q_i(x)$ The one can check that

$$Q_i = 1 + \frac{a_1}{m\sqrt{n}} D_x + \frac{a_2}{nm(m-1)} D_x^2 + O\left(n^{-3/2}\right) = 1 - \frac{\sigma^2}{2n(m-1)} D_x^2 + O\left(n^{-3/2}\right)$$

and so

$$[q_1 \boxplus_m \cdots \boxplus_m q_n] = \left(\prod_{i=1}^{n} Q_i\right) \{x^m\} = \left(1 - \frac{\sigma^2}{2n(m-1)} D_x^2 + O\left(n^{-3/2}\right)\right)^n \{x^m\} \qquad (29)$$

which one can check converges to

$$e^{-\sigma^2 D_x^2/2(m-1)} \{x^m\} = \mathfrak{N}_m[0, \sigma^2]$$

as $n \to \infty$. $\qquad\qquad\square$

Typically one needs to assume bounds on the higher moments in order to guarantee that the higher order terms in Equation (29) converge to 0 asymptotically. For polynomials of a fixed degree, however, this is implied by the bound on the second moment and the finite support of the underlying distribution.

6.2.3 Poisson

The Poisson random variable with parameter λ is characterized by having all cumulants equal to λ, so that

$$\mathcal{R}_{\mu_A}^m (s) = \frac{\lambda}{1 - s}$$

This time, we have

$$\ln Q(x) = -\lambda m \int \frac{1}{1 - x} \, dx = \lambda m \ln(1 - x) + c$$

so that

$$Q\left(\frac{D_x}{m}\right) \{x^m\} \approx \left(1 - \frac{D_x}{m}\right)^{\lambda m} \{x^m\} \approx L_m^{((\lambda - 1)m)}(mx)$$

where $L_m^{(\alpha)}(x)$ is an associated Laguerre polynomial.

$$L_n^{(\alpha)}(x) = \sum_i \frac{(-x)^i}{i!} \binom{n+\alpha}{n-i}. \tag{30}$$

This leads to the following definition (we will assume λm is an integer for simplicity):

$$\mathfrak{P}_m[\lambda] := m!(-m)^{-m} L_m^{((\lambda-1)m)}(mx) \tag{31}$$

where again the constant is only there so the resulting polynomial is monic.

Note that when $\lambda < 1$, we should expect to have an atom of probability $(1-\lambda)$ at 0. While this is not obvious in the current definition, the following Laguerre polynomial identity

$$\frac{(-x)^i}{i!} L_n^{(i-n)}(x) = \frac{(-x)^n}{n!} L_i^{(n-i)}(x). \tag{32}$$

can be applied to get

$$\mathfrak{P}_m[\lambda](x) = (\lambda m)!(-m)^{-\lambda m} x^{m(1-\lambda)} L_{\lambda m}^{(m(1-\lambda))}(xm) \tag{33}$$

which (more obviously) has the atom at 0.

Again, it should come as no surprise that the resulting polynomial is a Laguerre polynomial, since the root distribution must converge asymptotically to the free Poisson law (or *Marchenko–Pastur*) law, and it is known that the Laguerre polynomials do exactly that. One does need to check that the correct scaling of the law occurs, which one can do simply by checking that the support of the law matches that of the free Poisson law (the free Poisson law with expectation λ has support $[(1-\sqrt{\lambda})^2, (1+\sqrt{\lambda})^2]$). This can be done easily using the following result of Dette and Studden [10].

Theorem 6.8. *For any $a > 0$, the root distribution of the polynomials*

$$L_n^{(an)}(bnx)$$

has asymptotic density function

$$\frac{b}{2\pi x}\sqrt{(r_2 - x)(x - r_1)}$$

on the interval $[r_1, r_2]$ where

$$r_1 = \frac{1}{b}(2 + a - 2\sqrt{1+a}) \qquad and \qquad r_2 = \frac{1}{b}(2 + a + 2\sqrt{1+a}).$$

For $\lambda > 1$, we can apply Theorem 6.8 to Equation (31) (so $a = \lambda - 1$ and $b = 1$) to get a root distribution supported on the interval

$$[(1-\sqrt{\lambda})^2, (1+\sqrt{\lambda})^2].$$

When $\lambda < 1$, we can instead apply Theorem 6.8 to what is left of Equation (33) after removing the atom at $x = 0$ with measure $1 - \lambda$. This time, we take $a = 1/\lambda$ and $b = 1/\lambda$ to get a root distribution (again) supported on the interval

$$[(1-\sqrt{\lambda})^2, (1+\sqrt{\lambda})^2].$$

The limit theorem associated with Poisson distribution is known as the *Poisson limit theorem*. This classical version of this theorem states that if X_i are independent Bernoulli random variables with $P(X_i = 1) = p$, then

$$\sum_i X_i \xrightarrow[np \to \lambda]{n \to \infty} \text{Pois}(\lambda)$$

in distribution. Here we give a proof in the m-finite case:

Theorem 6.9 (Poisson limit theorem). *For all λ, m such that λm is an integer, we have*

$$[\underbrace{p \boxplus_m \cdots \boxplus_m p}_{\lambda m \ times}] = \mathfrak{P}_m[\lambda](x)$$

where $p(x) = x^{m-1}(x - 1)$.

Proof. We start by noting another identity of Laguerre polynomials:

$$n! x^\alpha L_n^{(\alpha)}(x) = (D_x - 1)^n x^{n+\alpha}$$

which when combined with Equation (32) (and a change of variables) gives

$$(D_x - r)^n x^{n+\alpha} = (n + \alpha)!(-r)^{-\alpha} L_{n+\alpha}^{(-\alpha)}(rx) \tag{34}$$

for all constants r.

Now we write

$$p(x) = x^m - x^{m-1} = \left(1 - \frac{1}{m} D_x\right)\{x^m\}$$

so that

$$[\underbrace{p \boxplus_m \cdots \boxplus_m p}_{\lambda m \ times}] = \left(1 - \frac{1}{ks} D_x\right)^{\lambda m}\{x^m\} = (-m)^{-\lambda m}(D_x - m)^{\lambda m}\{x^m\}.$$

So using Equation (34) with $r = m$ and $n = \lambda m$ and $\alpha = (1 - \lambda)m$ gives

$$[\underbrace{p \boxplus_m \cdots \boxplus_m p}_{\lambda m \ times}] = m!(-m)^{-m} L_m^{((\lambda-1)m)}(mx)$$

as claimed. □

6.2.4   Compound Poisson

The compound Poisson random variable with parameter λ and second distribution μ is characterized by having ith cumulant $\kappa_i = M_i(\mu)$ where M_i is the ith moment of μ. We assume that μ is distributed uniformly over the roots of a degree m polynomial h. Hence

$$\mathcal{R}_{\mu_A}^m(s) = \lambda \int \frac{t}{1 - st} \mu(t) \, dt$$

Hence

$$\ln Q(x) = -\lambda m \iint \frac{t}{1 - xt} \mu(t) \, dx \, dt = c + \lambda m \int \ln(1 - xt) \mu(t) \, dt.$$

Now if $h(x) = \prod_i (x - r_i)$, then we have

$$\int \ln\left(1 - xt\right) \mu(t)\, dt = \sum_i \ln(1 - xr_i) = \ln \prod_i (1 - xr_i).$$

Hence we have

$$Q\left(\frac{D_x}{m}\right)\{x^m\} \approx \prod_i \left(1 - \frac{r_i}{m} D_x\right)^{\lambda m} \{x^m\} \approx L_m^{((\lambda-1)m)}(mr_1 x) \boxplus_m \cdots \boxplus_m L_m^{((\lambda-1)m)}(mr_m x)$$

where $L_m^{(\alpha)}(x)$ is the associated Laguerre polynomial from Equation (30).

6.3 Restricted Invertibility

To conclude the applications, we relate the well known concept of *restricted invertibility* first proved by Bourgain and Tzafriri in [5] to the theory developed here. An argument of this type was first introduced in [17], but many aspects of the proof become much more intuitive in the language of finite free probability. Theorem 3.1 of [17] is the following:

Theorem 6.10. *If $v_1, \ldots, v_n \in \mathbb{C}^m$ are vectors with*

$$\sum_{i=1}^n v_i v_i^* = I,$$

then for all $k < n$, there exists a set $S \subset [n]$ with $|S| = k$ such that

$$\lambda_k\left(\sum_{i \in S} v_i v_i^T\right) \geq \left(1 - \sqrt{\frac{k}{m}}\right)^2 \left(\frac{m}{n}\right)$$

Those familiar with random matrix theory might recognize the quantity

$$\left(1 - \sqrt{\frac{k}{n}}\right)^2 \left(\frac{m}{n}\right)$$

as being the lower bound on the spectrum of the Marchenko–Pastur distribution with parameters $\lambda = k/n$ and intensity $\alpha = \frac{m}{n}$. This should not be much of a surprise, as the typical way one would form such a distribution would be to consider the spectrum of

$$\sum_{i=1}^k u_i u_i^T$$

where the u_i are *random* vectors where each coordinate of u_i is an independent Gaussian random variable with variance m/n. The resulting matrix is known as a *Wishart matrix*, and the distribution of the eigenvalues of such matrices was calculated in [16]. Using the results of Voiculescu, the addition of these random vectors approaches free convolution (in the asymptotic limit), and so one can instead see this as the sum of k freely independent rank-1 matrices with trace m/n. That is, the suspected bound in the finite case seems to be governed by the calculation in free probability.

The finite free version of rank-1 matrices with trace m/n is simply

$$p(x) = x^m - \left(\frac{m}{n}\right) x^{m-1}$$

and a calculation similar to the one done in Theorem 6.9 gives

$$[\underbrace{p \boxplus_m \cdots \boxplus_m p}_{k \text{ times}}] = m!(-n)^{-m} L_m^{k-m}(nx)$$

where we pick $\lambda = k/m$. The remainder of the proof then lies in building an "interlacing family" as developed in [19] that allows one to translate bounds on the roots of this Laguerre polynomial to bounds on the roots of individual polynomial. One can easily check that such an interlacing family can be built by picking k vectors uniformly at random *with replacement*. That is, if u_i is a random vector that is uniformly distributed over $v_1, \ldots, v_n$, then

$$\mathbb{E}\left\{\det\left[xI - \sum_{i=0}^{k} u_i u_i^T\right]\right\} = m!(-n)^{-m} L_m^{(k-m)}(nx)$$

and the hierarchy of polynomials that one gets by picking each vector one at a time forms such an interlacing family. We refer the reader to [17] for details.

7 Conclusions and Acknowledgements

The purpose of this paper was to draw the connection between the recent work of the author with Daniel Spielman and Nikhil Srivastava [17, 18, 19, 20, 21] with free probability and more traditional random matrix theory. It shows how the asymptotic intuitions that have been established in each of these fields can be translated into finite results, both computationally and theoretically. It gives a unified framework for solving problems using the "method of interlacing polynomials" that the author hopes will inspire new and creative uses.

7.1 Further Research

There are a number of interesting directions. Ongoing work between the author with Dimitri Shlyakhtenko and Nikhil Srivastava extends the concepts of free entropy and free Fisher information to the finite setting. Finite free probability also seems to have an interesting combinatorial interpretation that draws from both the classical (all partitions) and the free (noncrossing partitions) interpretations. Extending the results here to asymmetric matrices and to other concepts in free probability, like freeness with amalgamation [30], could also lead to interesting new applications. Of particular interest would be an extension to bivariate polynomials, which is possibly related to the concept of *second order freeness* introduced by Mingo and Speicher [24]. Such an extension would have the potential to place more advanced results such as [1, 20] under a similar umbrella, which would be a notable advance in the understanding of how such results fit into the free probability framework.

It would be interesting to see direct applications to random matrix theory, particularly in the realm of universality. Since such results typically use the relationship between polynomials and random matrices (and free probability) in a somewhat ad-hoc way, one might hope that a theory

connecting the two would be useful. The concept of finite free entropy seems to be directly related to such pursuits, since its manifestation employs a *logarithmic potential*, a topic that has led to a number of results in the asymptotic root distributions of polynomials [29]. In the reverse, random matrix theory could be useful in establishing concentration of measure results in this paradigm — something typically necessary for widespread applications.

Lastly, we mention possible implications in quantum information theory. In fact, results of this sort have already been applied in such a context: when A and B are in finite free position and B has rank 1, the (single) nontrivial eigenvector of B coincides with the so-called *trace vector* that was introduced by Murray and von Neumann in their initial work in C^* algebras [25]. Trace vectors have been used to obtain results in private quantum channels [7], and so the fact that finite freeness is a vast generalization of this concept gives promise to the possibility of further applications.

7.2 Acknowledgements

References

[1] N. Anari and S. O. Gharan. The Kadison–Singer problem for strongly Rayleigh measures and applications to asymmetric TSP. *arXiv preprint arXiv:1412.1143*, 2014.

[2] M. Anshelevich. Free stochastic measures via noncrossing partitions. *Advances in Mathematics*, 155(1):154–179, 2000.

[3] R. Bapat. Mixed discriminants of positive semidefinite matrices. *Linear Algebra and its Applications*, 126:107–124, 1989.

[4] J. Borcea and P. Brändén. The Lee-Yang and Pólya-Schur programs, II. theory of stable polynomials and applications. *Communications on Pure and Applied Mathematics*, 62(12):1595–1631, 2009.

[5] J. Bourgain and L. Tzafriri. Invertibility of 'large' submatrices with applications to the geometry of Banach spaces and harmonic analysis. *Israel journal of mathematics*, 57(2):137–224, 1987.

[6] M. Bownik, P. Casazza, A. W. Marcus, and D. Speegle. Improved bounds in Weaver's and Feichtinger's conjectures. *arXiv preprint arXiv:1508.07353*, 2015.

[7] A. Church, D. W. Kribs, R. Pereira, and S. Plosker. Private quantum channels, conditional expectations, and trace vectors. *Quantum Information & Computation*, 11(9-10):774–783, 2011.

[8] P. Deift. *Orthogonal polynomials and random matrices: a Riemann–Hilbert approach*, volume 3. American Mathematical Soc., 2000.

[9] C. Deninger. Determinants on von Neumann algebras, Mahler measures and Ljapunov exponents. *Journal für die reine und angewandte Mathematik (Crelles Journal)*, 2011(651):165–185, 2011.

[10] H. Dette and W. J. Studden. Some new asymptotic properties for the zeros of Jacobi, Laguerre, and Hermite polynomials. *Constructive Approximation*, 11(2):227–238, 1995.

[11] A. Edelman and N. R. Rao. Random matrix theory. *Acta Numerica*, 14:233–297, 2005.

[12] L. Erdős. Universality of Wigner random matrices: a survey of recent results. *Russian Mathematical Surveys*, 66(3):507, 2011.

[13] B. Fuglede and R. V. Kadison. On determinants and a property of the trace in finite factors. *Proceedings of the National Academy of Sciences*, 37(7):425, 1951.

[14] J. W. Helton and V. Vinnikov. Linear matrix inequality representation of sets. *Communications on pure and applied mathematics*, 60(5):654–674, 2007.

[15] A. Lewis, P. Parrilo, and M. Ramana. The Lax conjecture is true. *Proc. Amer. Math. Soc.*, 133(9):2495–2499, 2005.

[16] V. A. Marchenko and L. A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Matematicheskii Sbornik*, 114(4):507–536, 1967.

[17] A. W. Marcus, D. A. Spielman, and N. Srivastava. Ramanujan graphs and the solution of the Kadison–Singer problem. In *Proceedings of the ICM*, volume III, pages 375–386, 2014.

[18] A. W. Marcus, D. A. Spielman, and N. Srivastava. Finite free convolutions of polynomials. *arXiv preprint arXiv:1504.00350*, 2015.

[19] A. W. Marcus, D. A. Spielman, and N. Srivastava. Interlacing families I: bipartite Ramanujan graphs of all degrees. *Ann. of Math.*, 182-1:307–325, 2015.

[20] A. W. Marcus, D. A. Spielman, and N. Srivastava. Interlacing families II: mixed characteristic polynomials and the Kadison–Singer problem. *Ann. of Math.*, 182-1:327–350, 2015.

[21] A. W. Marcus, D. A. Spielman, and N. Srivastava. Interlacing families IV: bipartite Ramanujan graphs of all sizes. *arXiv preprint arXiv:1505.08010*, 2015.

[22] A. W. Marshall, I. Olkin, and B. Arnold. *Inequalities: theory of majorization and its applications*. Springer Science & Business Media, 2010.

[23] M. L. Mehta and M. Gaudin. On the density of eigenvalues of a random matrix. *Nuclear Physics*, 18:420–427, 1960.

[24] J. A. Mingo and R. Speicher. Second order freeness and fluctuations of random matrices, I. Gaussian and Wishart matrices and cyclic Fock spaces. *Journal of Functional Analysis*, 235(1):226–270, 2006.

[25] F. Murray and J. von Neumann. On rings of operators, II. *Transactions of the American Mathematical Society*, 41(2):208–248, 1937.

[26] A. Nica and R. Speicher. *Lectures on the combinatorics of free probability*, volume 13. Cambridge University Press, 2006.

[27] R. Pereira. Matrix-theoretical derivations of some results of Borcea–Shapiro on hyperbolic polynomials. *Comptes Rendus Mathematique*, 341(11):651–653, 2005.

[28] N. R. Rao and A. Edelman. The polynomial method for random matrices. *Foundations of Computational Mathematics*, 8(6):649–702, 2008.

[29] E. Saff and V. Totik. *Logarithmic potentials with external fields*, volume 316. Springer Science & Business Media, 2013.

[30] D. Shlyakhtenko. Random Gaussian band matrices and freeness with amalgamation. *International Mathematics Research Notices*, 1996(20):1013–1025, 1996.

[31] G. Szegő. *Orthogonal polynomials*, volume 23. American Mathematical Soc., 1939.

[32] D. V. Voiculescu. *Free probability theory.* American Mathematical Society, 1997.