PARTIAL FREENESS OF RANDOM MATRICES

JIAHAO CHEN AND ALAN EDELMAN

Abstract. We investigate the implications of free probability for random matrices. From rules for calculating all possible joint moments of two free random matrices, we develop a notion of partial freeness which is quantified by the breakdown of these rules. We provide a combinatorial interpretation for partial freeness as the presence of closed paths in Hilbert space defined by particular joint moments. We also discuss how asymptotic moment expansions provide an error term on the density of states. We present MATLAB code for the calculation of moments and free cumulants of arbitrary random matrices.

Contents

Introduction

Free probability has received much attention in algebra since its discovery as an algebraic structure for noncommuting operators.[20] Subsequently, it has also found a place in combinatorics with the relationship between free cumulants and noncrossing partitions.[12, 13] Its implications for random matrices, however, have not been as well studied. The discussion of free probability for random matrices usually centers around the onset of asymptotic freeness for certain random matrices in the infinite limit.[22, 2] By contrast, the situation for finite random matrices have not been studied much. This has motivated us to study the applicability of free probabilistic ideas to random matrix theory, concentrating in particular on its numerical and computational aspects.

In this paper, we provide a random matrix theoretic perspective on free probability theory, hoping that this will be accessible to readers familiar with linear algebra and elementary statistics, without requiring intricate knowledge of operator algebras or combinatorics. We illustrate the application of free probability to a specific problem, namely the calculation of the eigenvalues of sums of matrices. In general, the eigenvalues of the sum of two matrices $A + B$ are not the sum of the eigenvalues of the individual matrices A and B;[8] as matrices do not generally commute, the addition of eigenvalues must take into account the relative orientations of eigenvectors. In the limiting case where the rotation matrix between the two bases is a random matrix of Haar measure, the matrices are said to be in generic position and the eigenvalue spectrum converges to the additive free convolution $A \boxplus B$ of two random matrices A and B as the matrix sizes increase to infinity.[12]

A natural question to ask is how accurately $A \boxplus B$ approximates the exact eigenvalue spectrum, or density of states (d.o.s.), of the sum $A + B$ when the individual matrices are known to be noncommuting, but not necessarily free. We seek to quantify this statement in this paper. Our synthesis of existing results in the literature, together with a small number of new results, produces a coherent framework that provides insight into the inner workings of free probability on random matrices. We will also show that in addition to helping classify two random matrices as being free or not free relative to each other, we can characterize them as having an intermediate, graduated property which we call partial freeness. Furthermore, we are able to quantify the leading-order discrepancy between freeness and partial freeness. This has already helped us explain the unexpected accuracy of approximations to the Hamiltonians of disordered condensed matter systems.[4]

We begin with a brief, self-contained review of the free independence (i.e. freeness) from a random matrix theory perspective, and provide an elementary illustration of how computing the additive free convolution using an integral transform allows us to calculate the d.o.s. for the sum of free random matrices in a numerically exact manner. Next, we recap how the additive free convolution can also be approached via the moments of random matrices, and in particular how both classical and free independence can be interpreted as imposing precise rules for the decomposition of joint moments of arbitrary orders. We then define formally the notion of partial freeness and describe a procedure for detecting it numerically from samples of random matrices. Finally, we demonstrate with examples how partial freeness reveals insight into the optimal partition of a random matrix into the sum of two components such that their free convolution most closely resembles the original d.o.s.

1. Freeness of two matrices

We use the notation $\langle \cdot \rangle$ for the normalized expected trace (n.e.t.) $\frac{1}{N}\mathbb{E}\,\mathrm{Tr}\cdot$ of a $N \times N$ matrix.

Definition 1. The random matrices A and B are free (or synonymously, freely independent) with respect to the n.e.t. if for all $k \in \mathbb{N}$,

$$\langle p_1(A)\, q_1(B)\, p_2(A)\, q_2(B) \cdots p_k(A)\, q_k(B) \rangle = 0 \tag{1.1}$$

for all polynomials $p_1, q_1, p_2, q_2, \ldots p_k, q_k$ such that $\langle p_1(A) \rangle = \langle q_1(B) \rangle = \cdots = 0$.[20, Definition 4.2]

Fact 2. *The preceding is equivalent to defining free independence using the special case of the centering polynomials* $p_i(x) = x^{n_i} - \langle x^{n_i} \rangle$, $q_i(x) = x^{m_i} - \langle x^{m_i} \rangle$, $i = 1, \ldots, k$ *for positive integers* $n_1, m_1, \ldots, n_k, m_k$.[20, Proposition 4.3]

This fact necessarily follows from the definition of the centering polynomials. That this is sufficient follows from the linearity of the n.e.t. Thus the two statements are equivalent.

In principle, two matrices A and B can be checked for freeness by checking that all centered joint moments

$$\langle (A^{n_1} - \langle A^{n_1} \rangle) (B^{m_1} - \langle B^{m_1} \rangle) \cdots (A^{n_k} - \langle A^{n_k} \rangle) (B^{m_k} - \langle B^{m_k} \rangle) \rangle$$

vanish for all positive exponents $n_1, m_1, \ldots, n_k, m_k$. However, this is numerically impractical due to the need to check joint moments of all orders, as well as the presence of fluctuations from sampling error, which causes the higher order joint moments to converge more slowly with increasing degree. In practice, it is far easier to check that the d.o.s. of the exact sum $A + B$ converges to the p.d.f. defined by the free convolution $A \boxplus B$,[21] which we will now define.

1.1. The free convolution.

Definition 3. The R-transform of the p.d.f. f_A, denoted by R_A, is defined via the Cauchy transform:[21, 12]

$$w = \lim_{\epsilon \downarrow 0} \int_{\mathbb{R}} \frac{f_A(z)}{R_A^{-1}(w) - (z + i\epsilon)} dz \tag{1.2}$$

where R_A^{-1} is the functional inverse of R_A in the usual sense, i.e. $R_A^{-1}(R_A(w)) = R_A\left(R_A^{-1}(w)\right) = w$. Some intuition for the R-transform may be achieved by expanding the Cauchy integral as a formal power series:[1]

$$G_A(w) = \lim_{\epsilon \downarrow 0} \int_{\mathbb{R}} \frac{f_A(z)}{w - z - i\epsilon} dz = \sum_{k=0}^{\infty} \frac{\mu_k(A)}{w^{k+1}} \tag{1.3}$$

where μ_k is the kth moment

$$\mu_k(A) = \int_{\mathbb{R}} x^k f_A(x)\, dx = \left\langle A^k \right\rangle \tag{1.4}$$

In other word, the Cauchy transform of a p.d.f. is a generating function of its moments. We then have that the R-transform R_A complements the Cauchy transform G_A in the sense that

$$R_A\left(\frac{1}{G_A(w)}\right) = w \tag{1.5}$$

The free cumulants ν_k are the coefficients produced by the expansion of the R-transform as a formal power series in $1/w$, i.e.

$$R_A(w) = \sum_{k=0}^{\infty} \frac{\nu_k}{w^{k+1}} \tag{1.6}$$

with $\nu_k = 1$. The free cumulants are particular combinations of moments $\nu_k = \nu_k(\mu_1, \ldots, \mu_k)$ which will be made more explicit later.

Definition 4. The free convolution $A \boxplus B$ is defined via its R-transform:

$$R_{A \boxplus B}(w) = R_A(w) + R_B(w) - \frac{1}{w} \tag{1.7}$$

[1] Physicists may recognize $G_A(w)$ as the retarded Green function corresponding to the Hamiltonian A.

The free cumulants can be seen as linearizing the free convolution,[12, 13] in the sense that for all $k > 0$,

$$\nu_k(A \boxplus B) = \nu_k(A) + \nu_k(B) \tag{1.8}$$

and the subtraction of $1/w$ produces a properly normalized p.d.f. by conserving $\nu_0(A \boxplus B) = \mu_0(A \boxplus B) = 1$.

In Section 1.3, we show an example of calculating $f_{A \boxplus B}$ analytically via the R-transform. In general, such analytic calculations are hindered by the functional inversions required in (1.2). This has inspired interesting work in calculating $A \boxplus B$ numerically, such as in the RMTool package.[14] We discuss instead an alternate strategy starting directly from numerical samples of random matrices. In situations where only the numerical samples are known, it may be convenient instead to use the Result described in the next section.

1.2. Free convolution from random rotations.

Definition 5. A square matrix Q is a unitary/orthogonal/symplectic random matrix of Haar measure if for any constant unitary/orthogonal/symplectic matrix P, the integral of any function over dQ is identical to the integral over $d(PQ)$ or that over $d(QP)$.

Example 6. Unitary matrices of dimension $N = 1$ are simply scalar unit complex phases of the form $e^{i\theta}$. Haar measure over $e^{i\theta}$ can be written simply as $\mathrm{d}\theta/2\pi$. This is manifestly rotation invariant, as multiplying $e^{i\theta}$ by any constant phase factor $e^{i\phi}$ simply changes the measure to $\mathrm{d}(\theta + \phi)/2\pi = \mathrm{d}\theta/2\pi$.

Haar measure generalizes the concept of uniformity to higher dimensions by preserving the notion of invariance with respect to arbitrary rotations. Consequently, the eigenvalues of Q lie uniformly on the unit circle on the complex plane.[6] Explicit samples can be generated numerically by performing QR decompositions on $N \times N$ matrices sampled from the Gaussian orthogonal (unitary) ensemble.[5]

Fact 7. *For a pair of Hermitian (real symmetric) random matrices A and B, the d.o.s. of $A + QBQ^\dagger$, where Q is a unitary (orthogonal) random matrix of Haar measure, coincides with the p.d.f. of $A \boxplus B$ in the limit of infinitely large matrices $N \to \infty$.*

Consider the diagonalization of $A = Q_A \Lambda_A Q_A^\dagger$ and $B = Q_B \Lambda_B Q_B^\dagger$. The d.o.s. of $A + QBQ^\dagger$ is identical to that of $\Lambda_A + \left(Q_A^\dagger Q Q_B\right) \Lambda_B \left(Q_B^\dagger Q^\dagger Q_A\right)$, since these matrices are related by the similarity transformation $Q_A^\dagger (\cdot) Q_A$. However, the Haar property of Q means that the d.o.s. of this matrix is identical to that of $\Lambda_A + Q \Lambda_B Q^\dagger$. This affords us another intepretation of free convolution: it describes the statistics resulting from adding two random matrices when the basis of one matrix is randomly rotated or "spun around" relative to the other. The information about the relative orientations of the two bases is effectively ignored.

The freeness of random matrices is often discussed in the context of it emerging only in the limit of infinitely large matrices, i.e. $N \to \infty$, where it is called asymptotic freeness.[22] Perhaps the most well-known example is that two matrices sampled from the Gaussian ensembles (orthogonal, unitary or symplectic) are free.[12] Nevertheless, finite-dimensional

random matrices can exhibit freeness as well. We now provide some examples and illustrate the analytic calculation of the free convolution using the R-transform.

1.3. Examples of free finite-dimensional matrices.

Example 8. Consider the 2×2 real symmetric random matrices

$$A(t) = U(t)\,\sigma_z U(-t) \tag{1.9a}$$

$$B(t) = U(-t)\,\sigma_z U(t) \tag{1.9b}$$

where σ_z is the Pauli matrix $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ and $U(t)$ is the rotation matrix $\begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix}$, and the rotation angle t is uniformly sampled on the interval $[0, \pi]$. By construction, the d.o.s. of $A(t)$ and $B(t)$ are identical; their eigenvalues have the p.d.f.

$$f_A(x) = f_B(x) = \frac{1}{2}\left(\delta(x+1) + \delta(x-1)\right) \tag{1.10}$$

where $\delta(x)$ is the Dirac delta distribution. Furthermore, for any particular t, the sum of $A(t)$ and $B(t)$ can be written in the basis where $A(t)$ is diagonal as

$$M(t) = \sigma_z + U(-2t)\,\sigma_z U(2t) \tag{1.11}$$

By construction, $U(2t)$ is of uniform Haar measure and so the d.o.s. of $M(t)$ is given exactly by the additive free convolution of $A(t)$ and $B(t)$.[22, 12] The R-transforms of f_A and f_B are

$$R_A(w) = R_B(w) = \frac{1 + \sqrt{1+4w^2}}{2w} \tag{1.12}$$

where we have taken only the positive root so that that the R-transforms remain nonnegative. We then obtain

$$R_{A \boxplus B}(w) = R_A(w) + R_B(w) - \frac{1}{w} = \frac{-1 + \sqrt{1+4w^2}}{2w} \tag{1.13}$$

Finally, we calculate the p.d.f. using the Plemelj inversion formula:

$$f_{A \boxplus B}(x) = \frac{1}{\pi}\left[\operatorname{Im} R^{-1}_{A \boxplus B}(w)\right]_{w=x} \tag{1.14a}$$

$$= \frac{1}{\pi\sqrt{1-x^2}} \tag{1.14b}$$

which is the arcsine distribution on the interval $[-1, 1]$. The odd moments all vanish by the even symmetry of $f_{A \boxplus B}$, and the even moments are the central binomial coefficients $\mu_{2n}(A \boxplus B) = \binom{2n}{n}$.

This example illustrates how the free convolution of two discrete probability distributions can be a continous probability distribution. It is useful to contrast this result with the classical convolution $A \star B$,

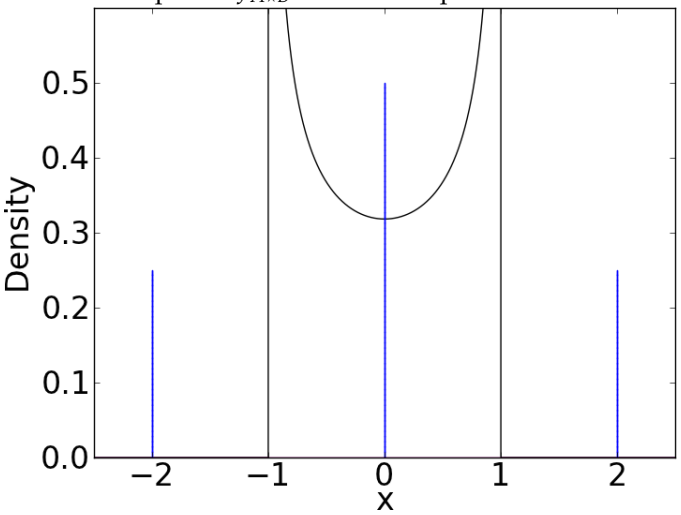
$$f_{A \star B}(x) = f_A \star f_B = \int_{\mathbb{R}} f_A(y)\, f_B(x-y)\, dy \tag{1.15a}$$

$$= \frac{1}{4}\left(\delta(x+2) + 2\delta(x) + \delta(x-2)\right) \tag{1.15b}$$

which is simply a discrete binomial distribution. The results of the two convolutions are plotted in figure 1.1.

Perhaps surprisingly, freeness can also be a property between two finite, deterministic matrices.

FIGURE 1.1. The d.o.s. $f_{A \boxplus B}(x)$ and $f_{A \star B}(x)$ for the free (solid black line) and classical convolutions (dashed blue line) of the matrices in Example 1, as given in (1.14b) and (1.15b) respectively. The heights of the lines in the plot of $f_{A \star B}$ indicate the point masses.



Example 9. The $2N \times 2N$ deterministic matrices

$$A = \begin{pmatrix} 0 & 1 & & & & \\ 1 & 0 & & & & \\ & & 0 & 1 & & \\ & & 1 & 0 & & \\ & & & & \ddots & \end{pmatrix} \tag{1.16a}$$

$$B = \begin{pmatrix} 0 & & & & & 1 \\ & 0 & 1 & & & \\ & 1 & 0 & & & \\ & & & \ddots & & \\ 1 & & & & & 0 \end{pmatrix} \tag{1.16b}$$

are free for all $N \geq 2$. Each consists of N copies of the Pauli matrix $\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, with B having a basis shifted relative to A with circulant (periodic) boundary conditions. The d.o.s. are the same as in the previous example and so the calculations of $A \boxplus B$ and $A \star B$ proceed identically.

The matrix B', being B without the circulant entries on the lower left and upper right, is not free relative to A. However, A and B' are still asymptotically free as $N \to \infty$.

1.4. Comparison of free and classical convolutions.

We conclude this introductory survey by fleshing out the analogy between the free additive convolution $\boxplus$ and the classical convolution $\star$. We note that the Fourier transform $\widehat{\cdot}$ turns classical convolutions into products according to the convolution theorem, i.e. $\widehat{f_A \star f_B} = \widehat{f_A}\,\widehat{f_B}$; furthermore, taking the logarithm allows this to be written as a linear sum:

$$\log \widehat{f_A \star f_B} = \log \widehat{f_A} + \log \widehat{f_B} \tag{1.17}$$

Furthermore if f_A and f_B are p.d.f.s, then $\log \widehat{f_A}$ and $\log \widehat{f_B}$ can be identified as the corresponding cumulant generating functions, i.e. $\log \widehat{f_A}$ can be expanded in a formal

power series

$$\log \widehat{f_A}(w) = \sum_{k=0}^{\infty} \frac{\kappa_n(A)\, w^n}{n!} \tag{1.18}$$

where $\kappa_n(A)$ is the nth cumulant of f_A, and similarly for f_B. We have also have that $\kappa_n(A \star B) = \kappa_n(A) + \kappa_n(B)$ for $n \geq 1$, i.e. that cumulants linearize the convolution.

For this reason, the R-transform is often described as the free analogue of the log-Fourier transform, with the Cauchy transform being the analogue of the Fourier transform and the functional inversion playing the part analogous to the logarithm. Naturally, the free analogue of the cumulant κ_n is the free cumulant ν_n.

Finally, we note that the sum of two random variables a and b, sampled with p.d.f.s f_A and f_B respectively, itself has the resulting p.d.f. $f_{A\star B}$.[7] If we have matrices A and B with d.o.s. f_A and f_B respectively, then the matrix $A + \Pi B \Pi^T$, where Π is a random permutation matrix, has d.o.s. $f_A \star f_B$. This is equivalent to the p.d.f. formed by picking an eigenvalue of A at random and an eigenvalue of B at random. In this sense, the discrete random permutation Π which generates the classical convolution is the analogue of the continuous random rotation Q of uniform Haar measure in free convolution.

Table 1 summarizes the analogies between the free and classical convolutions.

TABLE 1. Comparison of the free and classical convolutions.

$A \boxplus B$	$A \star B$
R-transform R	log-Fourier transform $\log \hat{f}$
Cauchy transform G	Fourier transform $\hat{f}$
functional inversion	logarithm
free cumulants ν_n	(classical) cumulants κ_n
Plemelj inversion	inverse Fourier transform
$A + QBQ^\dagger$	$A + \Pi B \Pi^T$
uniform Haar measure Q	random permutations Π

2. D.o.s. of sums of random matrices

We now consider general $N \times N$ random matrices A and B which are not necessarily free, and investigate moments of the sum, $\mu_n(A+B)$. The moments $\{\mu_n\}$ capture all the information contained in the corresponding p.d.f., provided that they all exist. For random matrices, this simply means that all powers of the matrix must have a finite n.e.t. For simplicity, we assume in the rest of this paper that all necessary moments exist.

Recall that the primordial definition of freeness (1.1) concerns itself with joint moments of A and B. We will now show how each moment of $A+B$ subdivides into sums over joint moments of A and B, and how this affords us fine-grain detail into the inner workings of free and classical independence, and whether or not they are obeyed by the matrices.

2.1. Calculating moments of $A+B$ from their joint moments.

We consider the expansion of the d.o.s. of the exact sum $A+B$ in terms of the moments $\{\mu_n(A+B)\}$. From the definition of the moment of a random matrix, these moments are given by

$$\mu_n = \left\langle (A+B)^n \right\rangle \tag{2.1a}$$

$$= \left\langle A^n + A^{n-1}B + A^{n-2}BA + \cdots + B^n \right\rangle \tag{2.1b}$$

$$= \langle A^n \rangle + n \left\langle A^{n-1}B \right\rangle + n \left\langle A^{n-2}B^2 \right\rangle + \cdots + \langle B^n \rangle \tag{2.1c}$$

where the second equality follows directly from the noncommutative binomial expansion of (2.1b), and the third equality follows from the linearity of $\langle\cdot\rangle$ and its cyclic invariance property, i.e. $\langle AB \rangle = \langle BA \rangle$.

We refer to (2.1c) as the *word expansion of* μ_n. As written, there are 2^n terms in (2.1b) but some of them yield the same n.e.t. in (2.1c) identically because of cyclic invariance. It turns out that the equivalence classes defined by grouping identical terms in this manner are exactly those of combinatorial necklaces.[9, 15]

Definition 10. A (n,k)-word W is a string of n symbols, each of which can have any of k values. A (n,k)-necklace $[\mathcal{N}]$ is the equivalence class over (n,k)-words W with respect to cyclic permutations Π of length n, i.e.

$$[\mathcal{N}] = \{w \in W | \exists \pi \in \Pi : \mathcal{N} = \pi w\} \tag{2.2}$$

We note that there are efficient algorithms for enumerating all (n,k)-necklaces for a given n and k.[16, 17] Furthermore, the total number of terms in the word expansion (2.1c) is well-known:

Fact 11. *The number of (n,k)-necklaces is*

$$N(n,k) = \frac{1}{n}\sum_{d|n} \phi(d)\, k^{n/d} = \frac{1}{n}\sum_{i=1}^{n} k^{\gcd(i,n)} \tag{2.3}$$

where $d|n$ means that d divides n, ϕ is the Euler totient function, gcd *is the greatest common divisor.*[9, 15] *By definition, $\phi(d)$ is the number of integers m in the range $1 \leq m \leq d$ that are relatively prime to d, i.e. $\gcd(d,m) = 1$.*

In addition, we can determine the multiplicity of each term in (2.1c), which is identical to the number of words in the equivalence class defined by each corresponding particular necklace. We state this very simple result without proof and provide an example.

Fact 12. *Let $m = \#([\mathcal{N}])$ be the number of (n,k)-words belonging to the equivalence class that defines the necklace $[\mathcal{N}]$. Then m is the length of the longest cyclic permutation that leaves any word $W \in \mathcal{N}$ unchanged, i.e. it is the length of the longest subword S of a word $W \in \mathcal{N}$ such that $W = S^{n/m}$.*

Example 13. The necklace $[AABAAB] = \left[A^2BA^2B\right]$ is an equivalence class over $(6,2)$-words of size 3, since applying a (one-symbol) cyclic permutation three times leaves A^2BA^2B unchanged:

$$AABAAB \mapsto ABAABA \mapsto BAABAA \mapsto AABAAB \tag{2.4}$$

i.e. $\#\left(\left[A^2BA^2B\right]\right) = 3$ which follows from the fact that $AABAAB = (AAB)^2$.

The algorithm for the enumeration of necklaces, together with knowledge of their multiplicities, allows us to sum the

joint moments in the word expansion (2.1c) to obtain μ_n. We now proceed to calculate each of these joint moments.

2.2. **Decomposition rules for joint moments.** We have now reduced the problem of calculating μ_n to that of calculating joint moments; each has the form $\langle A^{n_1}B^{m_1}\cdots A^{n_k}B^{m_k}\rangle$ for positive integers $n_1, m_1, \ldots, n_k, m_k$. We can understand classical and free independence as prescriptions for computing such joint moments in terms of the pure moments $\langle A\rangle$, $\langle A^2\rangle$, ..., and $\langle B\rangle$, $\langle B^2\rangle$, ... of A and B respectively.

Fact 14. *For classically independent random matrices A and B, we have*

$$\langle A^{n_1}B^{m_1}\cdots A^{n_k}B^{m_k}\rangle = \left\langle A^{n_1+\cdots+n_k}B^{m_1+\cdots+m_k}\right\rangle \tag{2.5a}$$

$$= \left\langle A^{n_1+\cdots+n_k}\right\rangle \left\langle B^{m_1+\cdots+m_k}\right\rangle \tag{2.5b}$$

i.e. A and B behave as if they commute.[12]

The analogous rule for free independence is somewhat more complicated, but can be derived from the primordial definition of freeness.

To cast the decomposition rule for free independence in the same form as that for classical independence in (2.5b), we can use again the linearity of the n.e.t. to expand the joint moment

$$\begin{aligned} 0 &= \langle (A^{n_1} - \langle A^{n_1}\rangle)(B^{m_1} - \langle B^{m_1}\rangle)\cdots \\ &\quad \times (A^{n_k} - \langle A^{n_k}\rangle)(B^{m_k} - \langle B^{m_k}\rangle)\rangle \qquad (2.6a)\\ &= \langle A^{n_1}B^{m_1}\cdots A^{n_k}B^{m_k}\rangle \\ &\quad - \langle A^{n_1}\rangle \langle A^{n_2}B^{m_2}\cdots A^{n_k}B^{m_k+m_1}\rangle \\ &\quad - \langle B^{m_1}\rangle \langle A^{n_1+n_2}B^{m_2}\cdots A^{n_k}B^{m_k}\rangle - \cdots \\ &\quad + (-1)^{\sum_{i=1}^{k}(n_i+m_i)} \langle A^{n_1}\rangle \cdots \langle A^{n_k}\rangle \langle B^{m_1}\rangle \cdots \langle B^{m_k}\rangle \qquad (2.6b) \end{aligned}$$

which can be rearranged immediately to give a recurrence relation for the joint moment $\langle A^{n_1}B^{m_1}\cdots A^{n_k}B^{m_k}\rangle$ in terms of lower order joint moments.

3. Partial freeness

We now present our new main concept, namely a proposal for defining partial freeness.

Definition 15. Two random matrices A and B are partially free of order p if the first statistically significant difference between $A+B$ and $A \boxplus B$ occurs at the pth moment μ_p, i.e. (2.6b) holds for all joint moments of the form

$$\langle A^{n_1}B^{m_1}\cdots A^{n_k}B^{m_k}\rangle$$

with positive integers $n_1, m_1, \ldots, n_k$, $\sum_{i=1}^{n}(n_i+m_i) = q$, for all $q < p$, but there exists at least one joint moment for $q = p$ for which (2.6b) does not hold. We will say that A and B are free to p moments.

We have immediately:

Fact 16. *Let A and B be a pair of $N \times N$ diagonalizable random matrices with $A = Q_A\Lambda_A Q_A^\dagger$ and $B = Q_B\Lambda_B Q_B^\dagger$. If $\mathbb{E}\left[\left(Q_B^\dagger Q_A\right)_{ij}\right] = 1/N$ for each matrix element of $Q_B^\dagger Q_A$, then A and B are free to $p > 3$ moments.*

This is a restatement of the matching three moments theorem of Refs. [10, 11].

Our definition is a natural generalization of the concept of freeness, and they coincide if all the moments match.

Claim 17. Two random matrices A and B are free if they are partially free to all orders $p \to \infty$.

This follows immediately from the definition of partial freeness above and the definition of freeness in (1.1).

Example 18. The $N \times N$ random matrix A, a diagonal matrix with elements i.i.d. standard Gaussian random variates, and B, the tridiagonal matrix

$$\begin{pmatrix} 0 & 1 & & 0 \\ 1 & \ddots & \ddots & \\ & \ddots & \ddots & 1 \\ 0 & & 1 & 0 \end{pmatrix}$$

are partially free of order 8.

This can be verified by explicit calculation of the first eight moments. Again by even symmetry all the odd moments of $A+B$ vanish while even moments are 1, 3, 17, 125, 1099, 11187, 129759, ... Furthermore we can identify the leading order deviation as arising from the term $\left\langle (AB)^4\right\rangle = 1$.

This example illustrates that (partial) freeness can also be a property of a pair of random and deterministic matrices.

Since free probability affords us a method for calculating the d.o.s. of the sum of two matrices $A+B$ using the free convolution $A \boxplus B$ when A and B are free, it is natural to ask how good an approximation the free convolution $A \boxplus B$ is to the d.o.s. of the sum $A+B$ when A and B are not free, but only partially free. We have proposed to quantify this statement using asymptotic moment expansions,[4] which we will describe in the next section.

4. Distinguishing between two distributions using asymptotic moment expansions

We have described partial freeness in terms of discrepancies in the moments between the exact p.d.f. and that from the free convolution. This suggests to us that a natural framework for expressing the difference between partially free matrices and (completely) free matrices can be found in the setting of asymptotic moment expansions,[23] which are parameterized naturally using the moments (or cumulants) of the two distributions being compared . The two standard approaches are the Gram-Charlier series of Type A and the Edgeworth series,[18] in which a probability density f is written as an expansion about another, reference probability density $\tilde{f}$.

4.1. **The Gram-Charlier Type A series.** The Gram-Charlier series arises immediately from the orthogonal polynomial expansion with respect to $\tilde{f}$ as the weight:[19, Chapter IX]

$$f(x) = \sum_{n=0}^{\infty} c_n \phi_n(x)\, \tilde{f}(x) \tag{4.1}$$

where the coefficients can be shown, by the orthonormality of the orthogonal polynomials, to be

$$\int_{\mathbb{R}} \phi_m(x)\, f(x)\, dx = \sum_{n=0}^{\infty} c_n \int_{\mathbb{R}} \phi_m(x)\, \phi_n(x)\, \tilde{f}(x)\, dx = c_m \tag{4.2}$$

i.e. the mth coefficient is the expected value of the mth orthogonal polynomial with respect to the probability density f. By expressing the orthogonal polynomials in the monomial basis,

$$\phi_m(x) = \sum_{k=0}^{m} a_{mk} x^k \tag{4.3}$$

$$c_m = \sum_{k=0}^{m} a_{mk} \int_{\mathbb{R}} x^k f(x)\, dx = \sum_{k=0}^{m} a_{mk} \mu_k \tag{4.4}$$

we get an explicit expansion of the Gram-Charlier coefficients as linear combinations of the moments μ_k of f. The Gram-Charlier Type A series is simply a special case of the orthogonal polynomial expansion just described as specialized to the case of a Gaussian weight:

$$\tilde{f}(x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \tag{4.5}$$

in which the corresponding family of orthogonal polynomials $\{\phi_n\}$ is simply the (probabilist's) Hermite polynomials, conventionally denoted by $\{He_n\}$.

4.2. **Edgeworth series.** The Gram-Charlier series comes from applying the operator

$$T(x) = \sum_{n=0}^{\infty} c_n \phi_n(x) \tag{4.6}$$

that tranforms $\tilde{f}$ into $f = T\tilde{f}$. The Edgeworth series is derived by rewriting T as a differential operator, as can be derived using the relations between a probability density, its characteristic function $\chi(x)$ and the moment generating function, and its cumulant generating function:

$$\chi(t) = \int_{\mathbb{R}} e^{itx} f(x)\, dx = \mathbb{E}_{f(x)}\left(e^{itx}\right) \tag{4.7}$$

$$= \sum_{n=0}^{\infty} \frac{\mu_n}{n!} (it)^n = \exp\left(\sum_{n=1}^{\infty} \frac{\kappa_n}{n!} (it)^n\right) \tag{4.8}$$

Writing down the analogous relations for $\tilde{f}$ and dividing yields

$$\frac{\chi(t)}{\tilde{\chi}(t)} = \exp\left(\sum_{n=1}^{\infty} \frac{\kappa_n - \tilde{\kappa}_n}{n!} (it)^n\right) \tag{4.9}$$

which, after rearrangement and taking the inverse Fourier transform, yields

$$f(x) = \exp\left(\sum_{n=1}^{\infty} \frac{\kappa_n - \tilde{\kappa}_n}{n!} \left(-\frac{d}{dx}\right)^n\right) \tilde{f}(x) \tag{4.10}$$

As before, the Edgeworth series is usually presented for the Gaussian case $\tilde{f} = \Phi$. Although these series are formally identical, they yield different partial sums when truncated to a finite number of terms. The Edgeworth form is generally considered more compact than the Gram-Charlier series, as only the former is a true asymptotic series.[18, 3]

4.3. **Deriving the Gram-Charlier series from the Edgeworth series.** Rederiving the Gram-Charlier form from the Edgeworth series reveals additional interesting relationships. One such relation follows from the identity

$$\exp\left(\sum_{n=1}^{\infty} \frac{a_n}{n!} t^n\right) = \sum_{n=0}^{\infty} \frac{B_n(a_1, \ldots, a_n)}{n!} t^n = \sum_{n=0}^{\infty} \frac{B_n(\{a_n\})}{n!} t^n \tag{4.11}$$

where B_n is the complete Bell polynomial of order n.[1] Setting $t = -d/dx$ gives immediately

$$T(x) = \exp\left(\sum_{n=1}^{\infty} \frac{(\kappa_n - \tilde{\kappa}_n)}{n!} \left(-\frac{d}{dx}\right)^n\right) \tag{4.12}$$

$$= \sum_{n=0}^{\infty} \frac{B_n(\{\kappa_n - \tilde{\kappa}_n\})}{n!} \left(-\frac{d}{dx}\right)^n \tag{4.13}$$

We will call (4.13) the direct series of T.

Futher specializing again to the Gaussian reference, we can use Rodrigues's formula

$$He_n(x)\, \Phi(x) = (-1)^n \left(\frac{d}{dx}\right)^n \Phi(x) \tag{4.14}$$

so that

$$T(x)\, \Phi(x) = \sum_{n=0}^{\infty} \frac{B_n(\{\kappa_n - \tilde{\kappa}_n\})}{n!} He_n(x)\, \Phi(x) \tag{4.15}$$

The first few coefficients $c_n = B_n(\{\kappa_n - \tilde{\kappa}_n\})$ have been tabulated explicitly,[18] but to our knowledge the relationship to the Bell polynomials have not been previously discussed in the literature.

4.4. **Quantifying the effect of differing moments.** The Edgeworth series yields a useful result for error quantification. If the first $k-1$ moments of two p.d.f.s f and $\tilde{f}$ are the same, but the kth moments differ, then the leading term in the Edgeworth series is

$$f(x) = \tilde{f}(x) + \frac{(-1)^k B_k(\{\kappa_k - \tilde{\kappa}_k\})}{k!} \tilde{f}^{(k)}(x) + \mathcal{O}\left(\tilde{f}^{(k+1)}\right) \tag{4.16a}$$

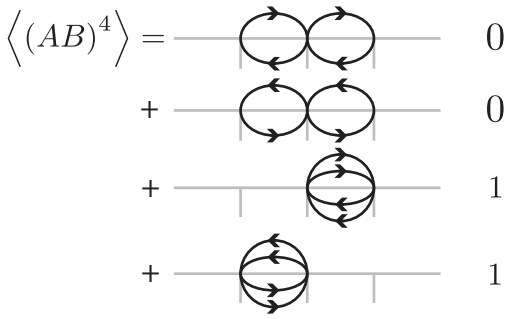
$$= \tilde{f}(x) + \frac{(-1)^k (\mu_k - \tilde{\mu}_k)}{k!} \tilde{f}^{(k)}(x) + \mathcal{O}\left(\tilde{f}^{(k+1)}\right) \tag{4.16b}$$

4.5. **The locus of differences in moments.** The word expansion (2.1c) allows us to refine the error analysis in terms of specific joint moments that contribute to $\mu_k - \tilde{\mu}_k$. It is instructive to interpret each term in (2.1c), being a trace of a product of k matrices, as a closed path with up to k hops as allowed by the structure of the matrices being multiplied.

Example 19. Consider A and B as in Example 18, which are partially free of degree 8 and whose discrepancy in the eighth moments relative to complete freeness is in the term $\left\langle (AB)^4 \right\rangle$. Note that B is exactly the adjacency matrix of the one–dimensional chain $\cdot - \cdot - \cdots - \cdot - \cdot$ with N nodes and periodic boundary conditions, and we can interpret $\left\langle (AB)^4 \right\rangle$ as corresponding to specific paths on this lattice. These paths must have exactly four hops, as A, being diagonal, does not permit hops, whereas B, having nonzero

entries only on the super- and sub-diagonals, require exactly one hop either to the immediate left or the immediate right. This gives rise to four different paths as illustrated in Figure 4.1. The paths neatly represent the terms arising

FIGURE 4.1. Paths contributing to the term $\left\langle (AB)^4 \right\rangle$ in Example 19.



from decomposing the joint moment into moments of scalar i.i.d. standard Gaussians g_i, i.e.:

$$\left\langle (AB)^4 \right\rangle = \langle g_i g_{i-1} g_i g_{i+1} \rangle + \langle g_i g_{i+1} g_i g_{i-1} \rangle + \langle g_i g_{i-1} g_i g_{i-1} \rangle + \langle g_i g_{i+1} g_i g_{i+1} \rangle \tag{4.17a}$$

$$= 2\mathbb{E}\left(g_i\right)^2 \mathbb{E}\left(g_i^2\right) + 2\mathbb{E}\left(g_i^2\right)^2 = 2 \tag{4.17b}$$

5. Computational implementation

5.1. **Numerical calculations on empirical samples.** The formalism we have developed can be applied directly to empirically sampled pairs of random matrices without knowledge of their underlying distributions. To demonstrate this, we have developed a proof of concept tool and implemented it in a MATLAB program:

(1) Generate t pairs of samples $\{(A_i, B_i)\}_{i=1}^t$ of $n \times n$ Hermitian random matrices A and B.
(2) For each pair:
   (a) Calculate the exact eigenvalues of A_i and B_i.
   (b) Calculate the eigenvalues of the sample $A_i + B_i$ of the exact sum $A + B$.
   (c) Calculate n samples of the free convolution $f_{A \boxplus B}$ using the eigenvalues of $M_i = A_i + Q_i B_i Q_i^\dagger$ using a numerically generated Haar orthogonal (or unitary) matrix Q_i.
   (d) Calculate n samples from the classical convolution $f_{A \star B}$ using the eigenvalues of $L_i = A_i + \Pi_i^{-1} B_i \Pi_i$ using a random permutation matrix Π_i.
(3) Calculate the first $2K$ moments of A and B as well as the first K moments of $\mu_k(A \boxplus B)$, $\mu_k(A \star B)$, and $\mu_k(A + B)$.
(4) Calculate the degree k for which A and B are partially free by testing for the smallest k such that the moments of the free convolution differ from the exact result, i.e. test the hypothesis $\mu_k(A + B) \neq \mu_k(A \boxplus B)$.
(5) Use Sawada's algorithm[17] to enumerate all unique terms $T_j = \left\langle A^{m_{1j}} B^{n_{1j}} \cdots A^{m_{kj}} B^{n_{kj}} \right\rangle$ in $\left\langle (A+B)^k \right\rangle$. For each term T_j:
   (a) Calculate
   $$T_j^{(cl)} = \left\langle A^{m_{1j}+\cdots+m_{kj}} \right\rangle \left\langle B^{n_{1j}+\cdots+n_{kj}} \right\rangle$$
   which would be its value expected from classical independence. Test the hypothesis of equality $T_j = T_j^{(cl)}$.
   (b) Calculate the normalized expected trace of the centered term
   $$T_j^{(c)} = \langle (A^{m_{1j}} - \langle A^{m_{1j}} \rangle)(B^{n_{1j}} - \langle B^{n_{1j}} \rangle) \cdots \rangle$$
   which would be expected to vanish if A and B were truly free. Test the hypothesis of equality $T_j^{(c)} = 0$.
(6) Calculate the kth derivative $f_{A \boxplus B}^{(k)}$ using numerical finite difference.
(7) Plot f_{A+B}, $f_{A \boxplus B}$ and $f_{A \boxplus B} + (\mu_k - \tilde{\mu}_k)/k! \cdot f_{A \boxplus B}^{(k)}$.

This algorithm tests for partial freeness of degree $k \leq K$, attempts to identify the locus of discrepancy by testing all possible $(k, 2)$-words, and calculates the leading order correction term to the density of states. In practice, we account for sampling error in the hypothesis tests in Steps 4 and 5 by calculating the standard error of each term being tested, and evaluating the p–value for each such hypothesis. For example, the standard error of the kth moment is

$$SE(\mu_k) = \sqrt{\frac{\mu_{2k} - \mu_k^2}{t}} \tag{5.1}$$

and the standard error of a term T_j in the expansion of $(A+B)^k$ is

$$SE(T_j) = \sqrt{\frac{\left\langle \left(A^{m_{1j}} B^{n_{1j}} \cdots A^{m_{kj}} B^{n_{kj}} \right)^2 \right\rangle - T_j^2}{t}} \tag{5.2}$$

The calculation of necessary standard errors require information up to the $2K$th moment for calculating variances stemming from the Kth moment, which is why $2K$ moments of A and B are calculated in Step 3.

5.2. **Symbolic computation of moments and joint moments.** If the analytic forms of the random matrices A and B are known, a computer algebra system can be used to calculate the necessary moments symbolically. Some examples of the calculation of the moments and joint moments in Mathematica is shown in Algorithm (1). The code provides simple functions for calculating normalized expected traces of an arbitrary joint moment or centered joint moment in terms of the distribution of matrix elements. For simplicity, only the i.i.d. case of one scalar probability distribution with moments $\{m_n\}$ is illustrated, although this approach can be extended to more complicated situations as necessary.

6. Summary

We have proposed an extension of the notion of freeness of random matrices.

The ideas in this paper are developed for two random matrices but our analysis generalizes readily to multiple additive free convolutions.

Algorithm 1 Mathematica code for calculating normalized expected traces of joint matrix products and centered joint matrix products for finite dimensional random matrices.

```
NN = 100; (* Size of matrix *)

(* The following generates the map which
formally evaluates the expectation of the
G random variables assuming that they
are i.i.d. with vanishing mean. *)
MomentsOfG := Flatten[{
 Table[Subscript[G, j]^i -> Subscript[m, i],
  {i, 2, NN}, {j, 1, NN}],
 Table[Subscript[G, j] -> 0, {j, 1, NN}]
}];

ExpectationOfG[x_] := x /. MomentsOfG;

(* Normalized expected trace *)
AngleBracket[x_] := ExpectationOfG[
 Tr[x]/NN // Expand ];

(* centering operator *)
c[x_] := (x - AngleBracket[x] IdentityMatrix[NN])

(* Example 19 *)
A = DiagonalMatrix[Array[Subscript[G, #] &, NN]];
B = SparseArray[{Band[{1, 2}] -> 1}, {NN, NN}];
B[[1]][[-1]] = 1; (* Add circulant boundary *)
B = B + Transpose[B];

AngleBracket[c[A.A].c[B.B]]
(* Output: 0 *)

AngleBracket[MatrixPower[c[A].c[B], 4]]
(* Output: 2 Subscript[m, 2]^2 *)

(* Specialized to standard Gaussian Gs *)
GaussianG = Array[Subscript[m, #] ->
  Moment[NormalDistribution[], #] &, NN];
AngleBracket[MatrixPower[A+B], 4]] /. GaussianG

(* Output: 1099 *)
```

This work was funded by NSF SOLAR Grant No. 1035400. We gratefully acknowledge useful discussions with N. Raj Rao (Michigan) that had led us to pursue this avenue of investigation.

References

[1] Eric Temple Bell. Partition polynomials. *The Annals of Mathematics*, 29:38–46, 1927–1928.

[2] Philippe Biane. Free probability for probabilists. In *Quantum probability communications*, volume 11, chapter 3, pages 55–71. 1998.

[3] S. Blinnikov and R. Moessner. Expansions for nearly Gaussian distributions. *Astronomy and Astrophysics Supplement Series*, 130(1):193–205, May 1998.

[4] Jiahao Chen, Eric Hontz, Jeremy Moix, Matthew Welborn, Troy Van Voorhis, Alberto Suárez, Ramis Movassagh, and Alan Edelman. *arXiv:1202.5839*.

[5] Persi Diaconis. What is a Random Matrix? *Notices of the American Mathematical Society*, 52(11):1348–9, April 2005.

[6] Persi Diaconis and Mehrdad Shahshahani. On the Eigenvalues of Random Matrices. *Journal of Applied Probability*, 31(1994):49, 1994.

[7] William Feller. *An introduction to probability theory and its applications*, volume 1. Wiley, 2 edition, 1971.

[8] Allen Knutson and Terence Tao. Honeycombs and Sums of Hermitian Matrices. *Notices of the American Mathematical Society*, 48(2):175–186, 2001.

[9] P. A. MacMahon. Applications of a Theory of Permutations in Circular Procession to the Theory of Numbers. *Proceedings of the London Mathematical Society*, 23(449):305–313, November 1892.

[10] Ramis Movassagh and Alan Edelman. Isotropic Entanglement. *arXiv:quant-ph/1012.5039*, page 29, December 2010.

[11] Ramis Movassagh and Alan Edelman. Density of States of Quantum Spin Systems from Isotropic Entanglement. *Physical Review Letters*, 107(9):097205, August 2011.

[12] Alexandru Nica and Roland Speicher. *Lectures on the Combinatorics of Free Probability*. London Mathematical Society Lecture Note Series. London, 2006.

[13] Jonathan Novak and Piotr Śniady. What is a Free Cumulant? *Notices of the American Mathematical Society*, 58(2):300–301, 2011.

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

[15] John Riordan. The Combinatorial Significance of a Theorem of Pólya. *Journal of the Society for Industrial and Applied Mathematics*, 5(4):225–237, 1957.

[16] Frank Ruskey, Terry Min, and Y I H Wang. Necklaces. *Journal of Algorithms*, 430:414–430, 1992.

[17] Joe Sawada. Generating Bracelets in Constant Amortized Time. *SIAM Journal on Computing*, 31(1):259, 2001.

[18] Alan Stuart and J Keith Ord. *Kendall's advanced theory of statistics*. Edward Arnold, London, 1994.

[19] Gábor Szegő. *Orthogonal polynomials*. Colloquium Publications. American Mathematical Society, Providence, RI, 4 edition, 1975.

[20] Dan Voiculescu. Symmetries of some reduced free product C*-algebras. In Huzihiro Araki, Calvin Moore, Serban-Valentin Stratila, and Dan-Virgil Voiculescu, editors, *Operator algebras and their connections with topology and ergodic theory*, volume 1132 of *Lecture Notes in Mathematics*, pages 556–588. Springer, 1985.

[21] Dan Voiculescu. Addition of Certain Non-commuting Random Variables. *Journal of functional analysis*, 66(3):323–346, 1986.

[22] Dan Voiculescu. Limit laws for Random matrices and free products. *Inventiones Mathematicae*, 104(1):201–220, December 1991.

[23] DL Wallace. Asymptotic approximations to distributions. *Annals of Mathematical Statistics*, 29(3):635–654, 1958.

Department of Chemistry, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge MA 02139
E-mail address: jiahao@mit.edu

Department of Mathematics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge MA 02139
E-mail address: edelman@math.mit.edu