Free Probability, Random Matrices, and Representations of Non-Commutative Rational Functions

Tobias Mai and Roland Speicher

Abstract A fundamental problem in free probability theory is to understand distributions of "non-commutative functions" in freely independent variables. Due to the asymptotic freeness phenomenon, which occurs for many types of independent random matrices, such distributions can describe the asymptotic eigenvalue distribution of corresponding random matrix models when their dimension tends to infinity. For non-commutative polynomials and rational functions, an algorithmic solution to this problem is presented. It relies on suitable representations for these functions.

1 Introduction

We want to understand distributions of functions in non-commuting variables. This phrase needs some explanations.

Firstly, let us specify what our "non-commuting variables" will usually be. We are mostly interested in either (random) matrices of size $N \times N$ or in operators on Hilbert spaces; one of our main points later will be that such operators correspond usually to the limit $N \to \infty$ of our random matrices.

Then, which "functions" of those variables do we want to consider? Since our variables do in general not commute, taking functions in such non-commuting variables is not a straightforward thing. In fact, we see that this question actually splits into two: we first need to clarify what our non-commutative functions should be as objects in their own right and secondly, we must explain how these non-commutative functions can be evaluated in the given collection of non-commuting

Tobias Mai
Universität des Saarlandes, Fachrichtung Mathematik, Postfach 151150, 66041 Saarbrücken e-mail: mai@math.uni-sb.de

Roland Speicher
Universität des Saarlandes, Fachrichtung Mathematik, Postfach 151150, 66041 Saarbrücken e-mail: speicher@math.uni-sb.de

variables. Everything which goes beyond polynomials is a non-trivial issue. Here, we will mostly address non-commutative polynomials and non-commutative rational functions, but our hope is that in the long run we will also have access to non-commutative analytic functions. The basis for a non-commutative analogue of complex function theory, intended to provide some sort of multivariate functional calculus in analogy to the well-known analytic functional calculus for a single operator, was laid in the 1970's in work of Joseph L. Taylor [33, 34]; but only recently this was revived and is under heavy development (with motivations coming from different directions, in particular free probability theory, but also control theory). We refer the reader who is interested in this subject to [25]. In this article we will not go beyond non-commutative rational functions.

Finally, we should be precise what we mean by "distribution" of our functions in our variables. There are essentially two versions of this. In the most general setting, we have to talk about algebraic/combinatorial distributions, which is just given by the collection of moments of our considered random variables. In more restricted analytic settings this might be identified with an analytic distribution, which is just a probability measure. To make this more precise we first have to set our frame.

Definition 1. A *non-commutative probability space* $(\mathscr{A},\varphi)$ consists of a complex algebra $\mathscr{A}$ with unit $1_{\mathscr{A}}$ and a linear functional $\varphi : \mathscr{A} \to \mathbb{C}$ satisfying $\varphi(1_{\mathscr{A}}) = 1$. Elements $x \in \mathscr{A}$ are called *non-commutative random variables* and φ is usually addressed as *expectation*.

Example 1. Let us give some examples for this.

1. The classical setting is captured in this algebraic form via $(L^{\infty}(\Omega,P),E)$, where (Ω,Σ,P) is a classical probability space and E the usual expectation that is given by $E[X] = \int_{\Omega} X(\omega)\,dP(\omega)$.
2. A typical genuine non-commutative example is $(M_N(\mathbb{C}),\mathrm{tr}_N)$, where tr_N is the normalized trace on $M_N(\mathbb{C})$; i.e., $\mathrm{tr}((a_{ij})_{i,j=1}^N) = \frac{1}{N}\sum_{k=1}^N a_{kk}$.
3. The combination of the two examples leads to one of our most important examples, given by random matrices $(L^{\infty}(\Omega)\otimes M_N(\mathbb{C}), E\otimes \mathrm{tr}_N)$. More on this later.

Definition 2. We call $(\mathscr{A},\varphi)$ a *C^*-probability space* if $\mathscr{A}$ is a unital C^*-algebra and φ is a state (i.e. $\varphi(x^*x) \geq 0$ for all $x \in \mathscr{A}$). The former means that $\mathscr{A}$ consists of bounded operators on some Hilbert space $\mathscr{H}$ and a state on $\mathscr{A}$ can, via the GNS construction, be realized in the form $\varphi(x) = \langle \Omega, x\Omega\rangle$ for some unit vector $\Omega \in \mathscr{H}$.

Now we can be more precise on what we mean with "distributions" in such a setting. In the general algebraic frame, we can only talk about the collection of moments, whereas in the analytic setting, we can identify this in the case of one self-adjoint operator with a probability measure. Usually, it is clear which of the two we are using; if we want to be precise, we should distinguish between "combinatorial" and "analytic" distribution.

Definition 3. 1. Let $(\mathscr{A},\varphi)$ be a non-commutative probability space. Let $(x_i)_{i\in I}$ be a family of non-commutative random variables. We call the collection of all mixed moments

$$\{\varphi(x_{i_1}\cdots x_{i_k}) \mid k \in \mathbb{N}, i_1,\ldots,i_k \in I\}$$

their *(joint) distribution* and denote it by $\mu_{(x_i)_{i\in I}}$.

2. Let $(\mathcal{A},\varphi)$ be a C^*-probability space. For any $x = x^* \in \mathcal{A}$, the distribution of x can be identified with the unique *Borel probability measure* μ_x on the real line $\mathbb{R}$ that satisfies

$$\varphi(x^k) = \int_{\mathbb{R}} t^k \, d\mu_x(t) \qquad \text{for all } k \in \mathbb{N}_0.$$

Note that in the classical multivariate case (for several commuting selfadjoint variables $x_1,\ldots,x_n$ in a C^*-setting) we can identify the combinatorial distribution $\mu_{x_1,\ldots,x_n}$ also with an analytic object, which is then just a probability measure on $\mathbb{R}^n$. In the general case, where our variables $x_1,\ldots,x_n$ do not commute, this is not possible any more. It is tempting to think of the distribution $\mu_{x_1,\ldots,x_n}$ in such a situation as a "non-commutative probability measure", but actually we have no idea what this should mean. As a kind of analytic substitute, we will try to analyze the distribution of $(x_1,\ldots,x_n)$ by investigating the analytic distributions of all $p(x_1,\ldots,x_n)$ for a large class of selfadjoint functions of $x_1,\ldots,x_n$. Clearly, the more functions we can deal with, the better we understand $\mu_{x_1,\ldots,x_n}$. Looking on polynomials and rational functions is a first step in this direction.

2 Random Matrices

Random matrices are $N \times N$ matrices, whose entries are chosen randomly (according to a prescribed distribution). Usually, one looks on sequences of such matrices for growing N. One of the basic observations in the subject is that for $N \to \infty$ something interesting happens. Before becoming more concrete on this, let us give a bit of history of the subject.

2.1 Some History

1928	Wishart introduced random matrices in statistics, for finite N;
1955	Wigner introduced random matrices in physics, for a statistical description of nuclei of heavy atoms, and investigated the $N \to \infty$ asymptotics of these "Wigner matrices";
1967	Marchenko and Pastur described the $N \to \infty$ asymptotics of "Wishart matrices";
1972	Montgomery and Dyson discovered relation between zeros of the Riemann zeta function and eigenvalues of random matrices;
since 2000	random matrix theory developed into a central subject in mathematics, with many different connections.

Fig. 1 The Oberwolfach workshop "Random Matrices" in 2000 was one of the first general appearances of the subject in mathematics. *Source: Archives of the Mathematisches Forschungsinstitut Oberwolfach.*

2.2 Wigner's semi-circle law

As said before, random matrices are sequences of $N \times N$ matrices whose entries are chosen randomly (according to a prescribed distribution). A fundamental observation in the subject is that many random matrices show for $N \to \infty$ almost surely a *deterministic* (and interesting) behaviour. Let us give an example for this via one of the most important random matrix ensembles, the Wigner matrices, which were introduced by Eugene Wigner in 1955 [38].

Definition 4. A *Wigner random matrix* $X_N = \frac{1}{\sqrt{N}}\left(x_{ij}\right)_{i,j=1}^{N}$ is a real random matrix, which is symmetric ($X_N^* = X_N$, i.e. $x_{ij} = x_{ji}$ for all $i, j = 1, \ldots, N$) and apart from this symmetry condition all its entries $\{x_{ij} \mid 1 \leq i \leq j \leq N\}$ are chosen independent and identically distributed.

Surprisingly, the common distribution of the entries does not matter for many results. The nicest distribution is the Gaussian (which leads to what is called "Gaussian orthogonal ensemble" GOE). We will here instead produce our random matrices by independent coin tosses for the entries; i.e., our common distribution for the entries is the symmetric Bernoulli distribution $\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_{+1}$. Here is one realization (via independent coin tosses by the authors) for such a 10×10 Wigner matrix:

$$\frac{1}{\sqrt{10}} \begin{pmatrix} 1 & -1 & -1 & 1 & -1 & 1 & -1 & -1 & -1 & 1 \\ -1 & 1 & -1 & -1 & 1 & 1 & -1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 & -1 & 1 & 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 & 1 & 1 & -1 & -1 & -1 & 1 \\ -1 & 1 & -1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & 1 & 1 & -1 & 1 & 1 \\ -1 & -1 & 1 & -1 & -1 & 1 & 1 & 1 & -1 & 1 \\ -1 & 1 & 1 & -1 & -1 & -1 & 1 & -1 & -1 & -1 \\ -1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & 1 \end{pmatrix}$$

The main quantity one is usually interested in for (random) matrices are the eigenvalues. For a matrix $A \in M_N(\mathbb{C})$, the information about its eigenvalues $\lambda_1, \ldots, \lambda_N$ (counted with multiplicity) is encoded in the *empirical eigenvalue distribution*

$$\mu_A = \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i}.$$

Note that this probability measure is nothing but the analytical distribution of A with respect to the normalized trace tr_N.

The left picture in Figure 2 shows the histogram of the 10 eigenvalues for the above matrix. Of course, since the matrix is random, the eigenvalue distribution is also random, so depends on the chosen realization. The right picture in Figure 2 is the histogram of the 10 eigenvalues of another such matrix created by coin tosses.

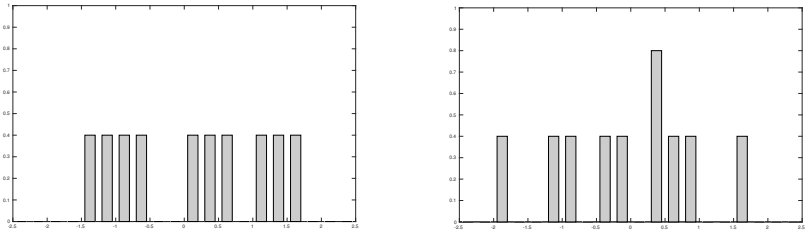

Fig. 2 The histogram of the 10 eigenvalues of a 10×10-Wigner matrix; for two different realizations of the matrix.

Clearly, the two pictures do not have much similarity. But now let's do the same for two different realizations of a 100×100 matrix, see Figure 3, and for two different realizations of a 3000×3000 matrix, see Figure 4. (Instead of tossing coins we preferred in those cases to use matlab for producing the matrices.)

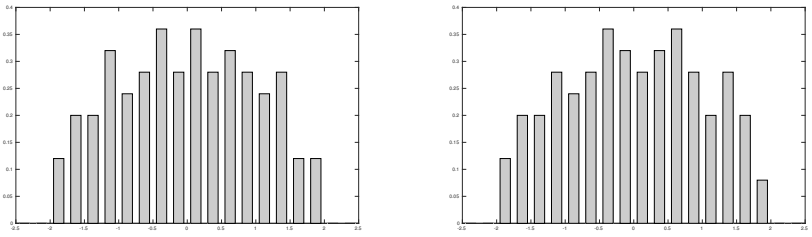

Fig. 3 The histogram of the 100 eigenvalues of a 100×100-Wigner matrix; for two different realizations of the matrix.

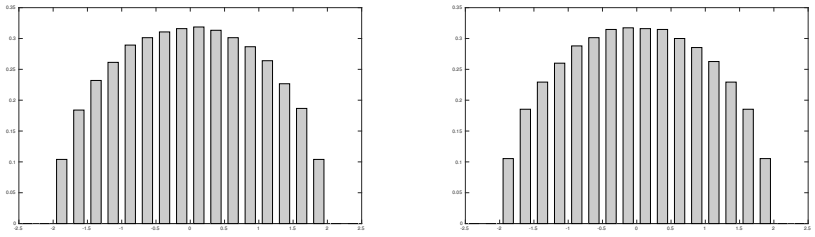

Fig. 4 The histogram of the 3000 eigenvalues of a 3000×3000-Wigner matrix; for two different realizations of the matrix.

Those histograms should make clear what we mean with the statement that for $N \to \infty$ the eigenvalue distribution of a Wigner matrix converges almost surely to a deterministic limit μ (which is called "semicircle distribution"); more precisely, we have that μ_{X_N} converges in the weak topology for probability measures to μ (and this happens for almost all realizations of X_N). This almost sure convergence is a concrete instance of concentration phenomena in high dimensions and is usually not too hard to prove. What is more interesting is the determination and description of this deterministic limit μ. Let us address the question how we can describe the limit.

2.3 Convergence in distribution to the large N limit

In the above treated one-matrix case X_N, the usual classical way of describing the almost sure limit of μ_{X_N} is by a probability measure μ. Here is an alternative to this, which we will favor in the following: instead of just describing μ, we try to find some nice operator x on a Hilbert space $\mathscr{H}$ with state φ such that the distribution of x with respect to φ coincides with μ; i.e. that $\mu = \mu_x$; then we can say that X_N converges to x in distribution. Note that this is like in the classical central limit theorem where often one prefers to talk about the convergence of normalized sums to a normal variable instead of just saying that the distribution of the normalized sums converge to a normal distribution.

Of course, this is just language. However, in the multi-variate non-commutative case this shift in perspective is more fundamental. So let us consider two independent copies X_N, Y_N of our Wigner matrices. As those do not commute, there is no nice analytic object describing their joint distribution (which is given by all mixed moments with respect to tr_N) and hence the determination of the almost sure limit of μ_{X_N,Y_N} would consist in trying to find some (combinatorial) description of the limits of the moments. Again, we propose an alternative: try to find some nice operators x, y on a Hilbert space with some state φ, such that almost surely

$$\lim_{N\to\infty} \mathrm{tr}_N(q(X_N, Y_N)) = \varphi(q(x,y))$$

for all monomials, and hence for all polynomials, q. Then we can say again that the pair (X_N, Y_N) converges in distribution to the pair (x,y).

The question is of course how big are our chances to have such nice limiting operators. Note that the important point here is "nice"; by the GNS-construction we can always find some abstract operators somewhere out there with the correct limiting moments. What we really want are operators, which can be handled and are useful.

The surprising fact in this context is the fundamental observation of Voiculescu [35] from 1991 that indeed limits of random matrices can often be described by "nice" and "interesting" operators on Hilbert spaces. (Actually, those operators describe usually interesting von Neumann algebras; which was the initial starting point of Voiculescu.)

2.4 Semi-circle law and one-sided shift

Here is again the histogram of our large Wigner matrices compared to the semicircle density.

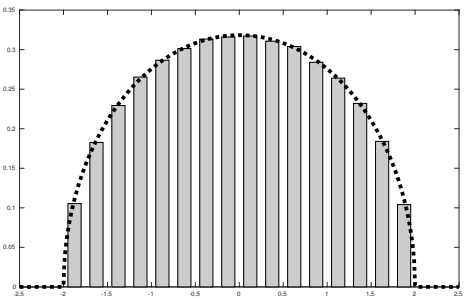

Fig. 5 Wigner's Theorem [38] says that the empirical eigenvalue distribution of Wigner matrices converges to the semicircle distribution; the Theorem of Füredi and Komlós [15] says that we also have almost sure convergence of the operator norms; i.e., there are no outlier eigenvalues outside the limit spectrum.

We claim that the real part of one of the most important Hilbert space operators – if suitably rescaled – has actually this semicircle distribution. More precisely, in this case our limit operator x can be written in the form $x = l + l^*$, where l is the one-sided shift on the Hilbert space $\bigoplus_{n\geq 0} \mathbb{C}e_n$ with orthonormal basis $(e_n)_{n\in\mathbb{N}_0}$; the action of the shift is given by the action on the basis: $le_n = e_{n+1}$ for all $n \in \mathbb{N}_0$; this implies that the action of the adjoint operator l^* is given by: $l^*e_{n+1} = e_n$ for all $n \in \mathbb{N}_0$ and $l^*e_0 = 0$. A canonical state on the algebra generated by those operators is the vector state $\varphi(a) = \langle e_0, ae_0\rangle$, corresponding to the distinguished basis element e_0. It turns out (and is actually a nice exercise) that the moments of x with respect to φ are given by the moments of the semicircle distribution; namely both are equal to

the famous *Catalan numbers*. More concretely, odd moments are zero in both cases and for even moments we have

$$\varphi(x^{2n}) = \frac{1}{2n+1}\binom{2n}{n} = \frac{1}{2\pi}\int_{-2}^{+2} t^{2n}\sqrt{4-t^2}\,dt.$$

In our language, we can now express the theorem of Wigner from 1955 [38] by saying that $X_N \to x$. Wigner did not equate the limiting moments of the Wigner matrices to the moments of our operator x, but just calculated them as the Catalan numbers.

We want to point out that the eigenvalue distribution μ_{X_N} gives only the averaged behaviour over all eigenvalues and its limiting behaviour does not allow to infer what happens to the largest eigenvalues of our Wigner matrices. Wigner's semicircle law would still allow that there is one exceptional large eigenvalue which has nothing to do with the limiting spectrum $[-2,+2]$. The mass $1/N$ of such an eigenvalue would disappear in the limit. However, there have been strengthenings of Wigner's result, which also tell us that such outliers are almost surely non-existent. More precisely, Füredi and Komlós showed in 1981 [15] that almost surely the largest eigenvalue of X_N converges to the edge of the spectrum, namely 2. Since the operator norm of the limit operator is 2, $\|x\| = 2$, we can paraphrase the result of Füredi and Komlós in our language as $\|X_N\| \to \|x\|$ almost surely.

2.5 Several independent Wigner matrices and full Fock space

Let us now consider the multi-variate situation. Voiculescu showed in [35] that the limit of two independent Wigner matrices X_N, Y_N can be described by a canonical multi-dimensional version of the one-sided shift; namely, by two copies of the one-sided shift in different directions. More precisely, we consider now the full Fock space $\mathscr{F}(\mathscr{H})$ over an underlying Hilbert space $\mathscr{H}$, given by

$$\mathscr{F}(\mathscr{H}) := \bigoplus_{n=0}^{\infty} \mathscr{H}^{\otimes n},$$

where $\mathscr{H}^{\otimes 0}$ is a one-dimensional Hilbert space which we write in the form $\mathscr{H}^{\otimes 0} = \mathbb{C}\Omega$ for some distinguished unit vector of norm one; Ω is usually called the *vacuum vector*. On this full Fock space one has, for each $f \in \mathscr{H}$, a *creation operator* $l(f)$ given by

$$l(f)\Omega = f, \qquad l(f)f_1 \otimes \cdots \otimes f_n = f \otimes f_1 \otimes \cdots \otimes f_n.$$

The adjoint of $l(f)$ is the *annihilation operator* $l^*(f)$, i.e. $l(f)^* = l^*(f)$, which is given by

$$l^*(f)\Omega = 0, \qquad l^*(f)f_1 \otimes \cdots \otimes f_n = \langle f, f_1\rangle f_2 \otimes \ldots f_n,$$

where, in particular, $l^*(f)f_1 = \langle f, f_1\rangle\Omega$. Let g_1 and g_2 be two orthogonal unit vectors in $\mathscr{H}$; then we put $x := l(g_1) + l^*(g_1)$ and $y := l(g_2) + l^*(g_2)$. Again, we have a canonical state given by the vacuum vector Ω, $\varphi(a) = \langle \Omega, a\Omega\rangle$. It turns now out (as a special case of Voiculescu's result [35] on asymptotic freeness) that we have $(X_N, Y_N) \to (x, y)$. Note that both x and y have with respect to φ a semicircular distribution; the basis vectors e_n from the one-sided shift correspond in the present setting to $g_1^{\otimes n}$ (for x) or to $g_2^{\otimes n}$ (for y).

Let us point out that in the same way as the one sided-shift l is one of the most important operators in single operator theory, the creation operators $l(g_1)$ and $l(g_2)$ are actually important (and nice) operators in the theory of operator algebras; they are closely related to the *Cuntz algebra*, which is one of the most important C^*-algebras and their real parts generate as von Neumann algebra the *free group factor* $L(\mathbb{F}_2)$, which is a main object of interest in Voiculescu's free probability theory.

As indicated at the end of Section 1, in order to get a better understanding of the limit distribution $\mu_{x,y}$ we will now try to deal with selfadjoint polynomials $p(x,y)$ in x and y. The convergence in distribution of (X_N, Y_N) to (x,y) implies that also $p(X_N, Y_N)$ converges to $p(x,y)$ for all such polynomials. For example, consider the polynomial $p(x,y) = xy + yx + x^2$. Then the theorem of Voiculescu tells us that $p(X_N, Y_N) \to p(x,y)$. This is again something which can be visualized by comparing the histogram of eigenvalues of $p(X_N, Y_N) = X_N Y_N + Y_N X_N + X_N^2$ with the analytic distribution of the selfadjoint operator $p(x,y) = xy + yx + x^2$; see Figure 6. At the moment it should not be clear to the reader how to get the distribution of $p(x,y)$ explicitly; understanding how we can get the dotted curve in Figure 6 will be one of the main points of the rest of this article.

Again, the behavior of the largest eigenvalue of $p(X_N, Y_N)$ is not captured by Voiculescu's result on the convergence in distribution of (X_N, Y_N) to (x,y). As in the classical case, there is some strengthening, which addresses this question. Namely, Haagerup and Thorbjørnsen have shown in [18] that we have almost sure convergence of the largest eigenvalue $\|p(X_N, Y_N)\|$ to the corresponding limit quantity $\|p(x,y)\|$. Whereas in the one-dimensional case we are only dealing with one limiting probability measure, for which the edge of the spectrum is clear, in the present, multivariate case we want now a statement covering a whole family of probability measures μ_p varying with the considered polynomial p; since we have no concrete description of those measures, there is also no explicit description of the edge of the support of those measures in useful classical terms – in the non-commutative setting, however, this can be easily described as the operator norm of $p(x,y)$.

2.6 Are those limit operators x, y really useful?

Still one might have the feeling that talking about operators as the limit of the X_N, Y_N instead of distributions of limits of $p(X_N, Y_N)$ might be more an issue of language than real insights. So the question remains: What are those limit operators good for? Here are some supporting facts in favour of their relevance.

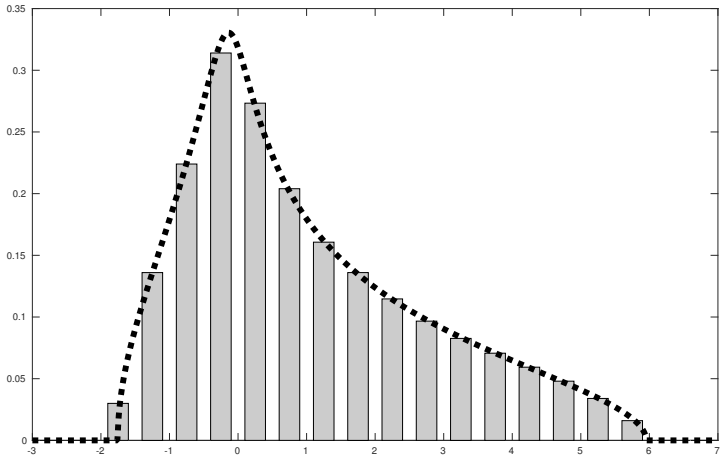

Fig. 6 Voiculescu's multivariate version of Wigner's Theorem says that the empirical eigenvalue distribution of $p(X_N, Y_N)$ for two independent Wigner matrices converges to the distribution of $p(x,y)$; the Theorem of Haagerup and Thorbjørnsen says that we also have almost sure convergence of the operator norms; i.e., there are no outlier eigenvalues outside the limit spectrum. Here we have $p(x,y) = xy + yx + x^2$.

Theorem 1 (Voiculescu 1991). *For many random matrix models X_N, Y_N (like for independent Wigner matrices) the limit operators x, y are free in the sense of Voiculescu's free probability theory*

We are not going to explain how "freeness" is defined (for this the reader should consult some of the references [21, 30, 31, 36] for the subject); instead we want to emphasize that free probability theory has developed a couple of tools to work effectively with free random variables. In particular, for x and y free we have

- *free convolution*: the distribution of $x+y$ can effectively be calculated in terms of the distribution of x and the distribution of y;
- *matrix-valued free convolution*: the matrix-valued distribution of $\alpha_0 \otimes 1 + \alpha_1 \otimes x + \alpha_2 \otimes y$ (where the coefficients $\alpha_0, \alpha_1, \alpha_2$ are now not just complex numbers, but matrices of arbitrary size) can be calculated in terms of the distribution of x and the distribution of y.

Still, this does not sound like a convincing argument in favour of x, y. What we want is to be able to deal with arbitrary polynomials in x and y. The above tells us that we can deal with linear polynomials in x and y, which seems to be much less. However, the fact the we have included the matrix-valued version above has striking consequences if we combine this with some powerful techniques of purely algebraic nature, which we summarize here for the seek of simplicity under the name "linearization". The latter is of such a fundamental relevance that we will treat it in a section of its own.

3 Linearization and the calculation of the distribution of $p(x,y)$

3.1 Idea of linearization

The linearization philosophy says that we can transform a *non-linear problem* into a *matrix-valued linear problem*. More precisely, if we want to understand a non-linear polynomial $p(x_1,\ldots,x_m)$ in non-commuting variables $x_1,\ldots,x_m$, then we can assign to this (in a non-unique way) a linear polynomial $\hat{p} := \alpha_0 \otimes 1 + \alpha_1 \otimes x_1 + \cdots + \alpha_m \otimes x_m$ (where we have to allow matrix-valued coefficients), such that $\hat{p}$ contains all "relevant information" about $p(x_1,\ldots,x_m)$. Relevant information for us is the spectrum of the operators, hence we would like to decide whether $p(x_1,\ldots,x_m)$, and more generally $z - p(x_1,\ldots,x_m)$ for $z \in \mathbb{C}$, is invertible. To see how such questions on invertibility can be shifted from $p(x_1,\ldots,x_m)$ to some $\hat{p}$ let us consider some examples.

Example 2. Consider first the simple polynomial $p(x,y) = xy$. We try to decide for which $z \in \mathbb{C}$ the element $z - xy$ is invertible. For this we write

$$\begin{pmatrix} z-xy & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \begin{pmatrix} z & -x \\ -y & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ y & 1 \end{pmatrix}. \tag{1}$$

Of course, $z - xy$ is invertible if and only if the matrix on the left-hand side is invertible. On the right-hand side we have a product of three matrices; however, the first and the third are always invertible, as one has for all x and all y

$$\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & -x \\ 0 & 1 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 0 \\ y & 0 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 \\ -y & 1 \end{pmatrix}.$$

Hence $z - xy$ is invertible if and only if the middle matrix

$$\begin{pmatrix} z & -x \\ -y & 1 \end{pmatrix} = \begin{pmatrix} z & 0 \\ 0 & 0 \end{pmatrix} - \begin{pmatrix} 0 & x \\ y & -1 \end{pmatrix} = \Lambda(z) - \hat{p}$$

is invertible, where we put

$$\Lambda(z) = \begin{pmatrix} z & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad \hat{p} = \begin{pmatrix} 0 & x \\ y & -1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix} \otimes 1 + \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \otimes x + \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \otimes y.$$

But this $\hat{p}$ is now a matrix-valued linear polynomial in x and y. Furthermore, we infer from the identity (1) that the resolvent $(z-xy)^{-1}$ appears as the $(1,1)$-entry of the 2×2-matrix $(\Lambda(z) - \hat{p})^{-1}$.

Example 3. Let us consider now the more interesting $p(x,y) = xy + yx + x^2$ and ask for which $z \in \mathbb{C}$ the element $z - p(x,y)$ becomes is invertible. Again we have a factorization into linear terms on matrix level

$$\begin{pmatrix} z-p(x,y) & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & y+\frac{x}{2} & x \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} z & -x & -y-\frac{x}{2} \\ -x & 0 & 1 \\ -y-\frac{x}{2} & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ y+\frac{x}{2} & 1 & 0 \\ x & 0 & 1 \end{pmatrix}. \quad (2)$$

As before the first and third term are triangular matrices which are always invertible; indeed,

$$\begin{pmatrix} 1 & 0 & 0 \\ y+\frac{x}{2} & 1 & 0 \\ x & 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ -y-\frac{x}{2} & 1 & 0 \\ -x & 0 & 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & y+\frac{x}{2} & x \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & -y-\frac{x}{2} & -x \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Hence $p(x,y) = xy+yx+x^2$ is invertible if and only if the 3×3-matrix valued linear polynomial

$$\begin{pmatrix} z & -x & -y-\frac{x}{2} \\ -x & 0 & 1 \\ -y-\frac{x}{2} & 1 & 0 \end{pmatrix} = \begin{pmatrix} z & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} - \begin{pmatrix} 0 & x & y+\frac{x}{2} \\ x & 0 & -1 \\ y+\frac{x}{2} & -1 & 0 \end{pmatrix} = \Lambda(z) - \hat{p}$$

is invertible, where we put

$$\Lambda(z) = \begin{pmatrix} z & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

and

$$\hat{p} = \begin{pmatrix} 0 & x & y+\frac{x}{2} \\ x & 0 & -1 \\ y+\frac{x}{2} & -1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{pmatrix} \otimes 1 + \begin{pmatrix} 0 & 1 & \frac{1}{2} \\ 1 & 0 & 0 \\ \frac{1}{2} & 0 & 0 \end{pmatrix} \otimes x + \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \otimes y.$$

One should note that also the value of the inverse can be read of from inverting $\hat{p}$. Namely, we can easily infer from (2) that the resolvent $(z-p(x,y))^{-1}$ is the $(1,1)$-entry of the 3×3-matrix $(\Lambda(z)-\hat{p})^{-1}$.

All the above can now actually be generalized to any polynomial $p(x,y)$. In view of the previous examples, this requires, of course, to have some general rule to produce matricial factorizations like in (1) and (2). For clarifying these relations, it is helpful to consider a block decomposition of the considered linearization $\hat{p}$ of the form

$$\hat{p} = \begin{pmatrix} 0 & u \\ v & Q \end{pmatrix}, \quad (3)$$

where the zero block in the upper left corner is of size 1×1 and all other blocks are of appropriate size. In each of the previous examples, we may observe

1. that the block Q is invertible without any conditions on x and y and
2. that its inverse Q^{-1} satisfies $p(x,y) = -uQ^{-1}v$.

Furthermore, we see that with these notations, the factorizations (1) and (2) take now the general form

$$\begin{pmatrix} z-p(x,y) & 0 \\ 0 & -Q \end{pmatrix} = \begin{pmatrix} 1 & -uQ^{-1} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} z & -u \\ -v & -Q \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -Q^{-1}v & 1 \end{pmatrix}. \tag{4}$$

In this abstract frame, we can repeat the computations, which were carried out in the previous examples; this yields

$$\begin{pmatrix} (z-p(x,y))^{-1} & 0 \\ 0 & -Q^{-1} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ Q^{-1}v & 1 \end{pmatrix} \left(\begin{pmatrix} z & 0 \\ 0 & 0 \end{pmatrix} - \hat{p} \right)^{-1} \begin{pmatrix} 1 & uQ^{-1} \\ 0 & 1 \end{pmatrix} \tag{5}$$

and finally

$$(z-p(x,y))^{-1} = \left[(\Lambda(z)-\hat{p})^{-1} \right]_{1,1} \qquad \text{with} \qquad \Lambda(z) = \begin{pmatrix} z & 0 \\ 0 & 0 \end{pmatrix}. \tag{6}$$

In fact, the validity of the factorization (4) and thus the validity of the formulas in (5) and (6) only depend on the properties formulated in Item 1 and Item 2. This is known under the name *Schur-complement formula* and it allows us to generalize our arguments given above to any non-commutative polynomial $p(x_1,\ldots,x_m)$ in finitely many variables $x_1,\ldots,x_m$. For this, however, we need to be sure that $p(x_1,\ldots,x_m)$ enjoys a representation of the form

$$p(x_1,\ldots,x_m) = -uQ^{-1}v \tag{7}$$

with vectors u,v and an invertible matrix Q of compatible sizes, which are (affine) linear in the variables $x_1,\ldots,x_m$. According to (3), finding such a representation of $p(x_1,\ldots,x_m)$ is all we need in order to produce a linearization $\hat{p}$.

Theorem 2. *Each non-commutative polynomial $p(x_1,\ldots,x_m)$ admits a representation of the form* (7)*. It can be constructed in the following way:*

1. If $p(x_1,\ldots,x_m)$ is a monomial *of the form*

$$p(x_1,\ldots,x_m) = \lambda x_{i_1} x_{i_2} \cdots x_{i_k}$$

with $\lambda \in \mathbb{C}$, $k \geq 1$, and $i_1,\ldots,i_k \in \{1,\ldots,m\}$, then

$$p(x_1,\ldots,x_m) = -\begin{pmatrix} 0 & 0 & \ldots & 0 & \lambda \end{pmatrix} \begin{pmatrix} & & & x_{i_1} & -1 \\ & & & x_{i_2} & -1 \\ & & \iddots & \iddots & \\ x_{i_k} & -1 & & & \\ -1 & & & & \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

2. If polynomials $p_1(x_1,\ldots,x_m),\ldots,p_k(x_1,\ldots,x_m)$ have representations

$$p_j(x_1,\ldots,x_m) = -u_j Q_j^{-1} v_j \qquad \text{for } j = 1,\ldots,k,$$

then their sum

$$p(x_1,\ldots,x_m) := p_1(x_1,\ldots,x_m)+\cdots+p_k(x_1,\ldots,x_m)$$

is represented by

$$p(x_1,\ldots,x_m) = -\begin{pmatrix}u_1 & \ldots & u_k\end{pmatrix}\begin{pmatrix}Q_1 & & 0\\ & \ddots & \\ 0 & & Q_k\end{pmatrix}^{-1}\begin{pmatrix}v_1\\ \vdots\\ v_k\end{pmatrix}.$$

3. If p is selfadjoint and

$$p(x_1,\ldots,x_m) = -uQ^{-1}v$$

any representation, then

$$p(x_1,\ldots,x_m) = -\begin{pmatrix}\frac{1}{2}u & v^*\end{pmatrix}\begin{pmatrix}0 & Q^*\\ Q & 0\end{pmatrix}^{-1}\begin{pmatrix}\frac{1}{2}u^*\\ v\end{pmatrix}$$

yields another representation, which induces via (3) *a self-adjoint linearization $\hat{p}$ of $p(x_1,\ldots,x_m)$.*

The previous theorem constitutes the alternative approach of Anderson [2] to the "linearization trick" of [18, 17]. Whereas the original algorithm in [18, 17] was quite complicated and did not preserve selfadjointness, Anderson's version streamlines their arguments and respects also selfadjointness. Let us summarize.

Theorem 3 (Haagerup, Thorbjørnsen 2005 (+Schultz 2006); Anderson 2012). *Every polynomial $p(x_1,\ldots,x_m)$ has a (non-unique) linearization*

$$\hat{p} = \alpha_0\otimes 1+\alpha_1\otimes x_1+\cdots+\alpha_m\otimes x_m,$$

such that

$$(z-p(x_1,\ldots,x_m))^{-1} = \left[(\Lambda(z)-\hat{p})^{-1}\right]_{1,1},\quad \textit{where}\qquad \Lambda(z) = \begin{pmatrix}z & 0 & \ldots & 0\\ 0 & 0 & \ldots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \ldots & 0\end{pmatrix}.$$

If p is selfadjoint, then $\hat{p}$ can also be chosen selfadjoint (meaning that the matrices $\alpha_0,\alpha_1,\ldots,\alpha_m$ appearing in $\hat{p}$ are all hermitian).

3.2 Calculation of the distribution of $p(x,y)$

Let us now come back to our problem of calculating the distribution of a selfadjoint polynomial $p(x,y)$ in two free variables x and y. The distribution μ_p of $p = p(x,y)$ is a probability measure and the information about such probability measures is often encoded in certain functions: whereas in classical probability theory the function of

our choice is usually the Fourier transform, in free probability and random matrix theory it is more adequate to use the so-called *Cauchy transform*; for any Borel probability measure μ on the real line $\mathbb{R}$, this is the analytic function

$$G_\mu : \mathbb{C}^+ \to \mathbb{C}^-,\ z \mapsto \int_{\mathbb{R}} \frac{1}{z-t}\, d\mu(t),$$

which is defined on the upper complex half plane $\mathbb{C}^+ = \{z \in \mathbb{C} | \Im(z) > 0\}$ and whose values lie all in the lower complex half plane $\mathbb{C}^- = \{z \in \mathbb{C} | \Im(z) < 0\}$. Note that the Cauchy transform G_μ differs only by a minus sign from the so-called *Stieltjes transform* $S_\mu : \mathbb{C}^+ \to \mathbb{C}^+$, which is the more familiar object in random matrix theory. It is an important fact that the measure μ can be recovered from G_μ by the so-called *Stieltjes inversion formula*: for each $\varepsilon > 0$, we have an absolutely continuous probability measure μ_ε given by

$$d\mu_\varepsilon(t) = -\frac{1}{\pi}\Im(G_\mu(t+i\varepsilon))\, dt,$$

and μ_ε converges weakly to μ as $\varepsilon \searrow 0$, meaning that

$$\int_{\mathbb{R}} f(t)\, d\mu(t) = \lim_{\varepsilon \searrow 0} \int_{\mathbb{R}} f(t)\, d\mu_\varepsilon(t)$$

for all bounded continuous functions $f : \mathbb{R} \to \mathbb{C}$. In this sense, knowing the Cauchy transform of a probability measure is as good as the measure itself. Note that for the analytic distribution μ_x of some selfadjoint x in a C^*-probability space $(\mathscr{A}, \varphi)$ we have $G_{\mu_x}(z) = \varphi((z-x)^{-1})$; we often write G_x instead of G_{μ_x}.

The method of linearization formulated in Theorem 3 now allows us to connect the wanted Cauchy transform of $p(x,y)$ with its selfadjoint linearization $\hat{p}$ according to

$$G_{p(x,y)}(z) = \varphi\big((z-p(x,y))^{-1}\big) = \big[(\varphi \otimes 1)\big((\Lambda(z)-\hat{p})^{-1}\big)\big]_{1,1}, \tag{8}$$

where $\varphi \otimes 1$ acts entrywise as φ on each entry of the corresponding matrix. This puts the original scalar-valued problem concerning $p(x,y)$ into the setting of operator-valued free probability, where the expression $(\varphi \otimes 1)\big((\Lambda(z)-\hat{p})^{-1}\big)$ can be interpreted as (a boundary value of) the operator-valued Cauchy transform of $\hat{p}$: an *operator-valued non-commutative probability space* $(\mathscr{A}, E, \mathscr{B})$ consists of a complex unital algebra $\mathscr{A}$ with a distinguished subalgebra $1_{\mathscr{A}} \in \mathscr{B} \subseteq \mathscr{A}$ and a linear map $E : \mathscr{A} \to \mathscr{B}$, called *conditional expectation*, which satisfies $E[b] = b$ for all $b \in \mathscr{B}$ and $E[b_1 a b_2] = b_1 E[a] b_2$ for all $a \in \mathscr{A}$, $b_1, b_2 \in \mathscr{B}$; this generalizes Definition 1. If $\mathscr{A}$ and $\mathscr{B}$ are even C^*-algebras and if E is positive in the sense that $E[a^*a] \geq 0$ holds for each $a \in \mathscr{A}$, then $(\mathscr{A}, E, \mathscr{B})$ is called an *operator-valued C^*-probability space*, in analogy to Definition 2. In the latter case, if we take any selfadjoint $X \in \mathscr{A}$, then the *$\mathscr{B}$-valued Cauchy transform of X* is defined by

$$G_X : \mathbb{H}^+(\mathscr{B}) \to \mathbb{H}^-(\mathscr{B}),\ b \mapsto E\big[(b-X)^{-1}\big],$$

where the upper respectively lower half plane in $\mathscr{B}$ are given by

$$\mathbb{H}^+(\mathscr{B}) = \{b \in \mathscr{B} \mid \Im(b) > 0\} \qquad \text{and} \qquad \mathbb{H}^-(\mathscr{B}) = \{b \in \mathscr{B} \mid \Im(b) < 0\}$$

with $\Im(b) = \frac{1}{2i}(b - b^*)$. Below, we will also use the so-called *h-transform of* X, which is given by

$$h_X : \mathbb{H}^+(\mathscr{B}) \to \overline{\mathbb{H}^+(\mathscr{B})},\ b \mapsto G_X(b)^{-1} - b.$$

Now, if N is the matrix size of the linearization $\hat{p}$, then the underlying C^*-probability space $(\mathscr{A}, \varphi)$ induces via $(M_N(\mathbb{C}) \otimes \mathscr{A}, \varphi \otimes 1, M_N(\mathbb{C}))$ an operator-valued C^*-probability space, in which we may interpret (8) as

$$G_{p(x,y)}(z) = \lim_{\varepsilon \searrow 0} \big[G_{\hat{p}}(\Lambda_\varepsilon(z))\big]_{1,1}, \qquad \text{where} \qquad \Lambda_\varepsilon(z) = \begin{pmatrix} z & 0 & \dots & 0 \\ 0 & i\varepsilon & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & i\varepsilon \end{pmatrix}.$$

Note that the only reason for having introduced the limit $\varepsilon \searrow 0$ is that we can move the point $\Lambda(z)$ to $\Lambda_\varepsilon(z)$, which clearly belongs to the natural domain $\mathbb{H}^+(M_N(\mathbb{C}))$ of the $M_N(\mathbb{C})$-valued Cauchy transform $G_{\hat{p}}$. Hence, what we need in order to calculate the distribution of the non-linear scalar polynomial $p(x,y)$ is to calculate the operator-valued distribution (via its operator-valued Cauchy transform) of the operator-valued linear polynomial

$$\hat{p} = \alpha_0 \otimes 1 + \alpha_1 \otimes x + \alpha_2 \otimes y.$$

But this is exactly the realm of operator-valued free convolution, for which we have a well-developed analytic theory [4]. We only need to note that if x and y are free, then $X = \alpha_0 \otimes 1 + \alpha_1 \otimes x$ and $Y = \alpha_2 \otimes y$ are free in the operator-valued sense with

$$G_X(b) = \int_{\mathbb{R}} (b - \alpha_0 - t\alpha_1)^{-1}\, d\mu_x(t) \qquad \text{and} \qquad G_Y(b) = \int_{\mathbb{R}} (b - t\alpha_2)^{-1}\, d\mu_y(t).$$

Theorem 4 (Belinschi, Mai, Speicher, 2013). *Consider an operator-valued C^*-probability space $(\mathscr{A}, E, \mathscr{B})$ and self-adjoint variables $X, Y \in \mathscr{A}$, which are free in the operator-valued sense. Then the operator-valued Cauchy transform of $X+Y$ can be calculated from the operator-valued Cauchy transforms G_X and G_Y in the following way: there exists a unique pair of (Fréchet-)holomorphic maps $\omega_1, \omega_2 : \mathbb{H}^+(\mathscr{B}) \to \mathbb{H}^+(\mathscr{B})$, such that*

$$G_X(\omega_1(b)) = G_Y(\omega_2(b)) = G_{X+Y}(b), \quad b \in \mathbb{H}^+(\mathscr{B})$$

holds, where the subordination functions ω_1 and ω_2 *can easily be computed via the following fixed point iterations on $\mathbb{H}^+(\mathscr{B})$*

$$\begin{aligned} w &\mapsto h_Y(b + h_X(w)) + b \qquad \text{for } \omega_1(b), \\ w &\mapsto h_X(b + h_Y(w)) + b \qquad \text{for } \omega_2(b). \end{aligned}$$

By applying this algorithm to $p(x,y) = xy + yx + x^2$ and its linearization $\hat{p} = X + Y$ with

$$X = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{pmatrix} \otimes 1 + \begin{pmatrix} 0 & 1 & \frac{1}{2} \\ 1 & 0 & 0 \\ \frac{1}{2} & 0 & 0 \end{pmatrix} \otimes x, \qquad Y = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \otimes y$$

we produced the distribution of $p(x,y)$ as shown in Figure 6. One should realize that the solution of the fixed point equations has to be done by numerical methods. Usually there is no hope of finding explicit solutions of those equations. Hence it is important to have a description of the solution which is amenable to easily implementable and controllable numerical methods. The fixed point equations from Theorem 4 provide such a controllable convergent scheme.

3.3 Historical remark

After the successful implementation of the above program it was brought to our attention by J. William Helton and Victor Vinnikov that the linearization trick is not new at all, but a well-known idea in many other mathematical communities, known under various names like

- Higman's trick ("The units of group rings": Higman 1940 [22])
- recognizable power series (automata theory: Kleene 1956 [28]; Schützenberger 1961 [32]; Fliess 1974 [14]; Berstel and Reutenauer 1984 [7])
- linearization by enlargement (ring theory: Cohn 1985 [10, 11]; Cohn and Reutenauer 1994 [12, 13]; Malcolmson 1978 [29])
- descriptor realization (control theory: Kalman 1963 [26, 27]; Ball, Malakorn, and Groenewald 2005 [3]; Helton, McCullough, and Vinnikov 2006 [20]; Kaliuzhnyi-Verbovetskyi and Vinnikov 2009/2012 [23, 24]; Volcic 2015 [37])

However, in most of those contexts dealing with polynomials is (in contrast to our application in free probability) kind of trivial and the real domain for the linearization idea are non-commutative rational functions. Since our algorithm for calculating the distribution of a polynomial in free variables is actually an algorithm on the level of linearizations, this implies right away that all we have said before should work equally well for non-commutative rational functions in free variables. Let us address these issues in the next section.

4 Distributions of non-commutative rational functions in free variables

Let us start with giving a bit of background on non-commutative rational functions before we address their distributions.

4.1 Non-commutative rational functions

Non-commutative rational functions were introduced by Amitsur [1] in 1966, whose methods were developed further by Bergman [6] in 1970, and they were studied extensively by Cohn [10, 11], Cohn and Reutenauer [12, 13], and Malcolmson [29]; see also [23, 24, 37].

Roughly speaking, non-commutative rational functions are given by rational expressions in non-commuting variables, like

$$r(x,y) := (4-x)^{-1} + (4-x)^{-1}y\big((4-x) - y(4-x)^{-1}y\big)^{-1}y(4-x)^{-1},$$

where two expressions are considered to be identical, when they can be transformed into each other by algebraic manipulations. The set of all non-commutative rational functions forms a skew field, the so-called *free field*. This – although it conveys the right idea – does not provide a rigorous definition of non-commutative rational functions to work with, since we presuppose here the existence of the free field as an algebraic frame, in which we can perform our algebraic manipulations. As a kind of substitute for this we can use matrix evaluations:

- Given a non-commutative rational expression r in m variables, we denote by $\mathrm{dom}(r)$ the subset of $\coprod_{n\in\mathbb{N}} M_n(\mathbb{C})^m$ consisting of all m-tuples $(X_1,\ldots,X_m)$, for which the evaluation $r(X_1,\ldots,X_m)$ is defined; if $\mathrm{dom}(r)\neq\emptyset$, we call the rational expression r *non-degenerate*.
- Two non-degenerate rational expressions r_1 and r_2 in m variables are considered to be equivalent if we have

$$r_1(X_1,\ldots,X_m) = r_2(X_1,\ldots,X_m) \qquad \text{for all } (X_1,\ldots,X_m)\in \mathrm{dom}(r_1)\cap\mathrm{dom}(r_2).$$

One can show that the free field is obtained as the set of all equivalence classes of non-degenerate rational expressions, with operations defined on representatives; we refer the reader to [24] for more details.

In the terminology of [10, 11], the free field is more precisely the universal skew field of fractions for the ring of non-commutative polynomials in the variables $x_1,\ldots,x_m$. That non-commutative rational functions form a skew field means that each $r(x_1,\ldots,x_m)\neq 0$ is invertible. However, deciding whether $r(x_1,\ldots,x_m)=0$ is not an easy task. For example, one has non-trivial rational identities, like

$$x_2^{-1} + x_2^{-1}(x_3^{-1}x_1^{-1} - x_2^{-1})^{-1}x_2^{-1} - (x_2 - x_3x_1)^{-1} = 0.$$

In the commutative situation, every rational function can be written as a fraction, i.e., the quotient of two polynomials. This is not true any more in the non-commutative case, and in general nested inversions are needed. So in the expression $r(x,y)$ from above we have a two-fold nested inversion. There are other ways of writing $r(x,y)$, but none of them can do without such a nested inversion. Whereas dealing with non-commutative rational functions just on the scalar level seems to be be quite involved, going over to a matrix-level makes things again easier. In fact, it turns out

that any non-commutative rational function can always be realized in the form of (7), namely in terms of matrices of polynomials, such that only one inverse is involved; in addition, we can achieve that the polynomials in the realization are linear. More precisely, we can always find a representation of the form

$$r(x_1,\ldots,x_m) = -uQ(x_1,\ldots,x_m)^{-1}v, \tag{9}$$

where u,v are scalar row and column vectors, respectively, and $Q(x_1,\ldots,x_m)$ is a matrix of corresponding size, whose entries are affine linear polynomials in the variables $x_1,\ldots,x_m$. For example, our $r(x,y)$ from above can be represented as

$$r(x,y) = \begin{pmatrix}\frac{1}{2} & 0\end{pmatrix}\begin{pmatrix}-1+\frac{1}{4}x & \frac{1}{4}y \\ \frac{1}{4}y & -1+\frac{1}{4}x\end{pmatrix}^{-1}\begin{pmatrix}\frac{1}{2}\\ 0\end{pmatrix}. \tag{10}$$

Such representations appear for instance in [29, 13]; in [13], where they go under the name *pure linear representations*, they were used for an alternative construction of the free field. For non-commutative rational functions $r(x_1,\ldots,x_m)$, which are *regular at zero* (meaning that $0 \in \mathrm{dom}(r)$ holds – at least after suitable algebraic manipulations), such representations are called *non-commutative descriptor realizations*; see [26, 27, 3, 20, 23, 24, 19, 37].

4.2 Linearization for non-commutative rational functions

As we have learned above, a realization like in (9) is according to (3) more or less the same as a linearization; the realization (10) of $r(x,y)$ yields directly a linearization $\hat{r}$ of the form

$$\hat{r} = \begin{pmatrix}0 & \frac{1}{2} & 0 \\ \frac{1}{2} & -1+\frac{1}{4}x & \frac{1}{4}y \\ 0 & \frac{1}{4}y & -1+\frac{1}{4}x\end{pmatrix}.$$

Since this fits into the frame of our machinery for the calculation of distributions, one is tempted to believe that these methods extend also to non-commutative rational functions. This is indeed the case, but there is one hidden subtlety, which requires clarification: representations like in (10) provide formulas for the non-commutative rational function $r(x_1,\ldots,x_m)$ and are thus valid only over the free field; it is not clear, if those formulas remain valid under evaluation of the involved rational expression r, i.e., when the variables $x_1,\ldots,x_m$ of the free field are replaced by elements from any non-commutative probability space $(\mathscr{A},\varphi)$.

Such questions were addressed in [19]. The focus there was mainly on the case of non-commutative rational functions regular at zero, but most of the arguments pass directly to the frame of [13]. In any case, it turns out that rational identities are not necessarily preserved under evaluations on general algebras $\mathscr{A}$. However, it works well for the important class of *stably finite* algebras $\mathscr{A}$ (sometimes also addressed as *weakly finite*): if for $(X_1,\ldots,X_m) \in \mathscr{A}^m$ both $r(X_1,\ldots,X_m)$ is defined

and $Q(X_1,\ldots,X_m)$ is invertible over $\mathscr{A}$, then $r(X_1,\ldots,X_m) = -uQ(X_1,\ldots,X_m)^{-1}v$ holds true. Notably, it was proven in [19] that, under certain conditions on the representation (9), the invertibility of $Q(X_1,\ldots,X_m)$ is automatically given, whenever $r(X_1,\ldots,X_m)$ is defined. It can be shown that if $(\mathscr{A},\varphi)$ is a C^*-probability space, endowed with a faithful tracial state φ, then $\mathscr{A}$ must be stably finite.

When working in such a setting, our machinery applies. For the given r, the linearization $\hat{r}$ splits into a term X depending only on x and a term Y depending only on y: i.e., we have $\hat{r} = X + Y$ with

$$X = \begin{pmatrix} 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \otimes 1 + \begin{pmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{4} \end{pmatrix} \otimes x, \qquad Y = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & \frac{1}{4} \\ 0 & \frac{1}{4} & 0 \end{pmatrix} \otimes y.$$

If x and y are free, then X and Y are free in the operator-valued sense, and this is again an operator-valued free convolution problem, which can be solved as before by applying Theorem 4. The dotted curve in Figure 7 shows the result of such a calculation for our r from above.

4.3 Rational functions of random matrices and their limit

In Figure 7 we compare again the distribution of $r(x,y)$ with the histogram of the eigenvalues of $r(X_N,Y_N)$ for independent Wigner matrices X_N, Y_N. One point one has to realize in the context of rational functions is that they cannot be evaluated on all operators. Clearly, we should only plug in operators for which all needed inverses make sense. So we have chosen an r which has two free semicirculars x and y in its domain. (For example, we have to invert $4-x$; this is okay, because the spectrum of x is $[-2,2]$.) If we approximate x,y by X_N,Y_N we hope that $r(X_N,Y_N)$ also makes sense, at least for sufficiently large N. This is indeed the case, but relies on the fact that we also have good control on the largest eigenvalues. More precisely we have the following statement [39].

Proposition 1 (Yin 2017). *Consider selfadjoint random matrices X_N,Y_N which converge to selfadjoint operators x,y in the following strong sense: for any selfadjoint polynomial p we have almost surely*

- *$p(X_N,Y_N) \to p(x,y)$ in distribution,*
- *$\lim_{N\to\infty} \|p(X_N,Y_N)\| = \|p(x,y)\|$.*

Then this strong convergence remains also true for rational functions: Let r be a selfadjoint non-commutative rational expression, such that $r(x,y)$ is defined. Then we have almost surely that

- *$r(X_N,Y_N)$ is defined eventually for large N,*
- *$r(X_N,Y_N) \to r(x,y)$ in distribution,*
- *$\lim_{N\to\infty} \|r(X_N,Y_N)\| = \|r(x,y)\|$.*

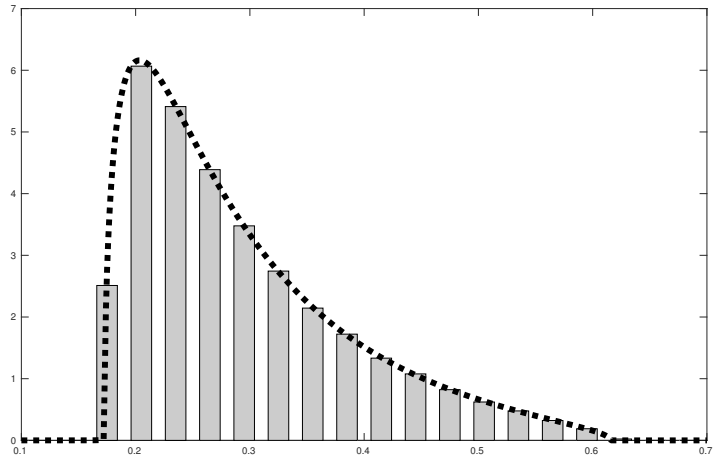

Fig. 7 As in Figure 6, we consider the convergence of two independent Wigner matrices X_N, Y_N to two free semicircular operators x, y. We have then also for non-commutative rational functions r the almost sure convergence of $r(X_N, Y_N)$ to $r(x, y)$ in distribution as well as the convergence of the operator norms - again, there are no outlier eigenvalues outside the limiting spectrum. Here we have $r(x,y) = (4-x)^{-1} + (4-x)^{-1}y\big((4-x) - y(4-x)^{-1}y\big)^{-1}y(4-x)^{-1}$.

5 Non-selfadjoint case: Brown measure

The reader might wonder about our restriction to the case of self-adjoint polynomials (or rational functions). Why not consider arbitrary polynomials in random matrices or their limit operators, like

$$p(x_1, x_2, x_3, x_4) = x_1x_2 + x_2x_3 + x_3x_4 + x_4x_1?$$

Of course, we can (say for four independent Wigner matrices) just plug in our random matrices and calculate their eigenvalues. Those are now not real anymore, we will get instead a number of points in the complex plane, as in the right plot of Figure 8.

The limit of such four independent Wigner matrices is given by four free semicircular elements s_1, s_2, s_3, s_4. The relevant information about $p := p(s_1, s_2, s_3, s_4)$ is given by its $*$-distribution, i.e., all moments in p and p^*. As p and p^* do not commute, this information cannot fully be captured by an analytic object, like a probability measure on $\mathbb{C}$. There is no straightforward substitute for the eigenvalue distribution for a non-normal operator. The full information about the non-normal operator p is given by its $*$-distribution, which is a highly non-trivial algebraic object. There is however a projection of this non-commutative algebraic object into the analytic classical world; namely, there exists a probability measure ν_p on $\mathbb{C}$, which

captures some information about the $*$-distribution of p, and which is a canonical candidate for the limit of the eigenvalue distribution for the corresponding random matrix approximations. This ν_p is was introduced by Brown [9] in 1981 for operators in finite von Neumann algebras and is called the *Brown measure* of the operator p: let (M,τ) be a *tracial W^*-probability space*, i.e., a non-commutative probability space build out of a von Neumann algebra M and a faithful normal tracial state τ on M; for any given $x \in M$,

- the *Fuglede-Kadison determinant* $\Delta(x)$ is determined by

$$\log(\Delta(x)) = \int_{\mathbb{R}} \log(t)\, d\mu_{|x|}(t) \in \mathbb{R} \cup \{-\infty\},$$

  where $\mu_{|x|}$ denotes the analytic distribution of the operator $|x| = (x^*x)^{\frac{1}{2}}$ in the sense of Definition 3, Item 2;
- the *Brown measure* ν_x is the compactly supported Radon probability measure on $\mathbb{C}$, which is uniquely determined by the condition

$$\int_{\mathbb{C}} \psi(z)\, d\nu_x(z) = \frac{1}{2\pi} \int_{\mathbb{C}} \nabla^2 \psi(z) \log(\Delta(x-z))\, d\Re(z)\, d\Im(z)$$

  for all compactly supported C^∞-functions $\psi : \mathbb{C} \to \mathbb{C}$, where $\nabla^2\psi$ denotes the Laplacian of ψ, i.e., $\nabla^2\psi = \frac{\partial^2\psi}{\partial\Re(z)^2} + \frac{\partial^2\psi}{\partial\Im(z)^2}$.

It can be shown that the support of ν_x is always contained in the spectrum of the operator x. For matrices, the Brown measure coincides with the eigenvalue distribution. For self-adjoint operators, the Brown measure is just the analytic distribution of the operator.

It turns out that we can refine the algorithm for selfadjoint polynomials or rational functions also to the non-selfadjoint case in order to calculate the Brown measure of arbitrary polynomials or rational functions in free variables. The result of this machinery for the polynomial $p(s_1,s_2,s_3,s_4)$ from above is shown in the left plot of Figure 8. For more facts about the Brown measure as well as for the details on how free probability allows to calculate it, see [16, 8, 5, 19].

One problem in this context is that the construction of the Brown measure is not continuous with respect to convergence in $*$-distribution, i.e., knowing that our independent Wigner matrices X_1,X_2,X_3,X_4 converge in $*$-distribution to s_1,s_2,s_3,s_4 does not guarantee that the Brown measure of $p(X_1,X_2,X_3,X_4)$ converges to the Brown measure of $p(s_1,s_2,s_3,s_4)$. It is an open conjecture that this is indeed the case for all polynomials or even rational functions in independent Wigner matrices.

Acknowledgement

This work was supported by the ERC Advanced Grant "Non-commutative Distributions in Free Probability" (grant no. 339760).

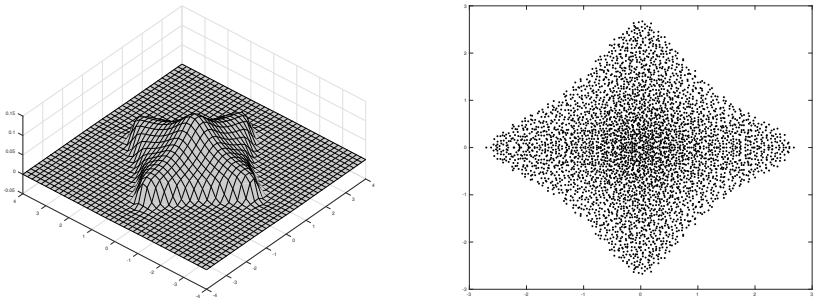

Fig. 8 The right plot shows the complex eigenvalues of the polynomial $p(X_1,X_2,X_3,X_4)$ in four independent Wigner matrices, each of size $N = 4000$. The left plot shows the Brown measure of the corresponding limit operator $p(s_1,s_2,s_3,s_4)$, calculated with our operator-valued free probability machinery. Here we have $p(x_1,x_2,x_3,x_4) = x_1x_2 + x_2x_3 + x_3x_4 + x_4x_1$.

References

1. Amitsur, S.A.: Rational identities and applications to algebra and geometry. J. Algebra **3**, 304–359 (1966)
2. Anderson, G.W.: Convergence of the largest singular value of a polynomial in independent Wigner matrices. Ann. Probab. **41**(3B), 2103–2181 (2013)
3. Ball, J.A., Groenewald, G., Malakorn, T.: Structured noncommutative multidimensional minear systems. SIAM J. Control Optim. **44**(1), 1474–1528 (2005)
4. Belinschi, S.T., Mai, T., Speicher, R.: Analytic subordination theory of operator-valued free additive convolution and the solution of a general random matrix problem. J. Reine Angew. Math. (2013). DOI 10.1515/crelle-2014-0138
5. Belinschi, S.T., Sniady, P., Speicher, R.: Eigenvalues of non-hermitian random matrices and Brown measure of non-normal operators: hermitian reduction and linearization method. arXiv:1506.02017 (2015)
6. Bergman, G.W.: Skew fields of non-commutative rational functions, after Amitsur. Semin. M.P. Schützenberger, A. Lentin, M. Nivat 1969/70, Probl. Math. Theor. Automates, No.16, 18 p. (1970). (1970)
7. Berstel, J., Reutenauer, C.: Rational series and their languages. Berlin etc.: Springer-Verlag (1988)
8. Biane, P., Lehner, F.: Computation of some examples of Brown's spectral measure in free probability. Colloq. Math. **90**(2), 181–211 (2001)
9. Brown, L.G.: Lidskii's theorem in the type II case. Geometric methods in operator algebras, Proc. US-Jap. Semin., Kyoto/Jap. 1983, Pitman Res. Notes Math. Ser. 123, 1-35 (1986). (1986)
10. Cohn, P.M.: Free rings and their relations. 2nd ed. London Mathematical Society Monographs, No.19. London etc.: Academic Press (Harcourt Brace Jovanovich, Publishers). XXII, 588 p. (1985) (1985)
11. Cohn, P.M.: Free ideal rings and localization in general rings. Cambridge: Cambridge University Press (2006)
12. Cohn, P.M., Reutenauer, C.: A normal form in free fields. Can. J. Math. **46**(3), 517–531 (1994)
13. Cohn, P.M., Reutenauer, C.: On the construction of the free field. Int. J. Algebra Comput. **9**(3-4), 307–323 (1999)
14. Fliess, M.: Sur divers produits de séries formelles. Bull. Soc. Math. Fr. **102**, 181–191 (1974)

15. Füredi, Z., Komloś, J.: The eigenvalues of random symmetric matrices. Combinatorica **1**, 233–241 (1981)
16. Haagerup, U., Larsen, F.: Brown's spectral distribution measure for *R*-diagonal elements in finite von Neumann algebras. J. Funct. Anal. **176**(2), 331–367 (2000)
17. Haagerup, U., Schultz, H., Thorbjørnsen, S.: A random matrix approach to the lack of projections in $C^*_{\mathrm{red}}(\mathbb{F}_2)$. Adv. Math. **204**(1), 1–83 (2006)
18. Haagerup, U., Thorbjørnsen, S.: A new application of random matrices: $\mathrm{Ext}(C^*_{\mathrm{red}}(F_2))$ is not a group. Ann. Math. (2) **162**(2), 711–775 (2005)
19. Helton, J.W., Mai, T., Speicher, R.: Applications of Realizations (aka Linearizations) to Free Probability. arXiv:1511.05330v1 (2015)
20. Helton, J.W., McCullough, S.A., Vinnikov, V.: Noncommutative convexity arises from linear matrix inequalities. J. Funct. Anal. **240**(1), 105–191 (2006)
21. Hiai, F., Petz, D.: The semicircle law, free random variables and entropy. Providence, RI: American Mathematical Society (AMS) (2000)
22. Higman, G.: The units of group-rings. Proc. Lond. Math. Soc. (2) **46**, 231–248 (1940)
23. Kaliuzhnyi-Verbovetskyi, D.S., Vinnikov, V.: Singularities of rational functions and minimal factorizations: the noncommutative and the commutative setting. Linear Algebra Appl. **430**(4), 869–889 (2009)
24. Kaliuzhnyi-Verbovetskyi, D.S., Vinnikov, V.: Noncommutative rational functions, their difference-differential calculus and realizations. Multidimensional Syst. Signal Process. **23**(1-2), 49–77 (2012)
25. Kaliuzhnyi-Verbovetskyi, D.S., Vinnikov, V.: Foundations of free noncommutative function theory. Providence, RI: American Mathematical Society (AMS) (2014)
26. Kalman, R.E.: Mathematical description of linear dynamical systems. J. Soc. Ind. Appl. Math., Ser. A, Control **1**, 152–192 (1963)
27. Kalman, R.E.: Realization theory of linear dynamical systems. Control Theory Top. Funct. Anal., Vol. II, Lect. int. Semin. Course, Trieste 1974, 235-256 (1976). (1976)
28. Kleene, S.C.: Representation of events in nerve nets and finite automata. Automata Studies, pages 341. Princeton University Press, Princeton, NJ, USA (1956)
29. Malcolmson, P.: A prime matrix ideal yields a skew field. J. Lond. Math. Soc., II. Ser. **18**, 221–233 (1978)
30. Mingo, J., Speicher, R.: Free Probability and Random Matrices. to appear in Fields Monograph Series, Springer (2016)
31. Nica, A., Speicher, R.: Lectures on the combinatorics of free probability. Cambridge: Cambridge University Press (2006)
32. Schützenberger, M.P.: On the definition of a family of automata. Inf. Control **4**, 245–270 (1961)
33. Taylor, J.L.: A general framwork for a multi-operator functional calculus. Adv. Math. **9**, 183–252 (1972)
34. Taylor, J.L.: Functions of several noncommuting variables. Bull. Am. Math. Soc. **79**, 1–34 (1973)
35. Voiculescu, D.: Limit laws for random matrices and free products. Invent. Math. **104**(1), 201–220 (1991)
36. Voiculescu, D., Dykema, K.J., Nica, A.: Free random variables. A noncommutative probability approach to free products with applications to random matrices, operator algebras and harmonic analysis on free groups. Providence, RI: American Mathematical Society (1992)
37. Volcic, J.: Matrix coefficient realization theory of noncommutative rational functions. arXiv:1505.07472v1 (2015)
38. Wigner, E.P.: Characteristic vectors of bordered matrices with infinite dimensions. Ann. Math. (2) **62**, 548–564 (1955)
39. Yin, S.: Rational functions in strongly convergent random matrices. Preprint (2017)