Free Probability Theory and Random Matrices

Roland Speicher

1 Motivation of Freeness via Random Matrices

In this chapter we want to motivate the definition of freeness on the basis of random matrices.

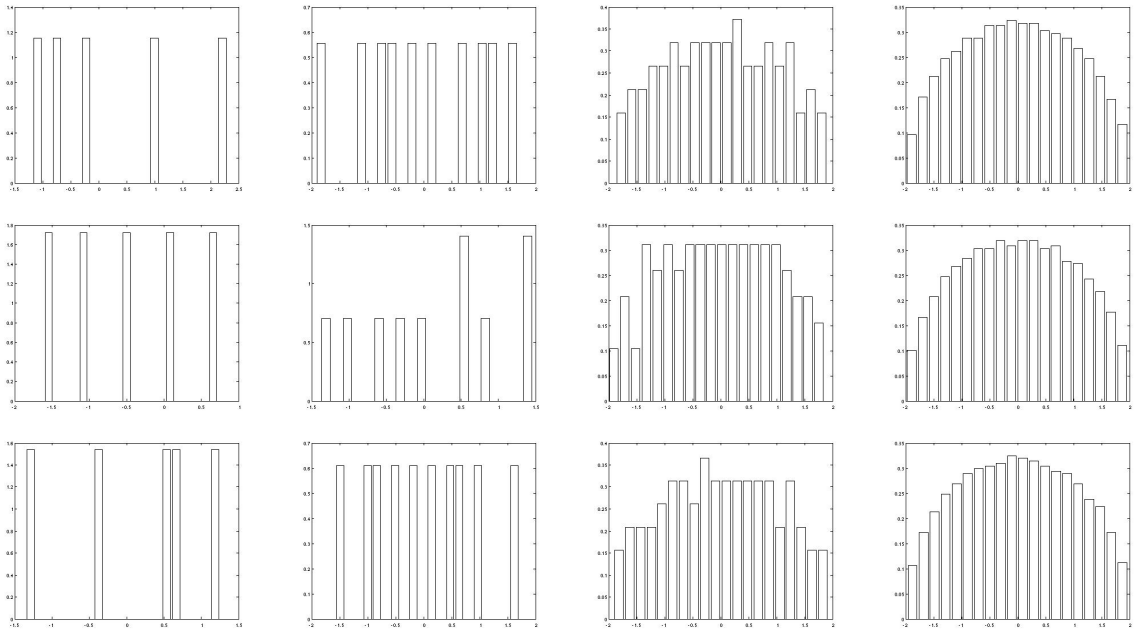
1.1 Asymptotic Eigenvalue Distribution of Random Matrices

In many applications it's necessary to compute the eigenvalue distribution of $N \times N$ random matrices as $N \to \infty$. For many basic random matrix ensembles we have almost sure convergence to some limiting eigenvalue distribution. Another nice property is the fact that this distribution often can be efficiently calculated.

Example 1.1. We consider an $N \times N$ **Gaussian random matrix**. This is a selfadjoint matrix $A = \frac{1}{\sqrt{N}}(a_{ij})_{i,j=1}^{N}$ such that the entries $\{a_{ij}\}_{i \geq j}$ are independent identically distributed complex (real for $i = j$) Gaussian random variables fullfilling the relations

- $\mathbb{E}[a_{ij}] = 0$,
- $\mathbb{E}[a_{ij}\bar{a}_{ij}] = 1$.

We show the eigenvalue distribution of some realizations of Gaussian random matrices of different dimensions:



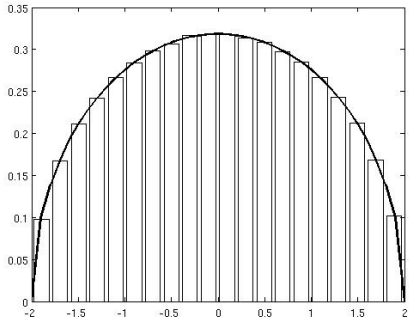
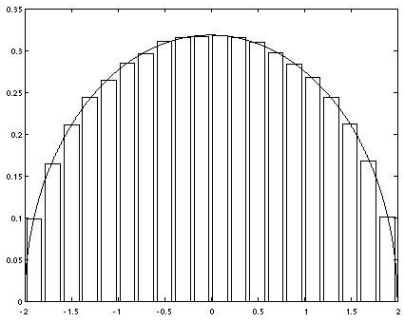
$N = 5$ $N = 10$ $N = 100$ $N = 1000$

One sees that for large N the eigenvalue histogram looks always the same, independent of the actual realized matrix from the ensemble. Actually, for such matrices we have almost sure convergence to **Wigner's**

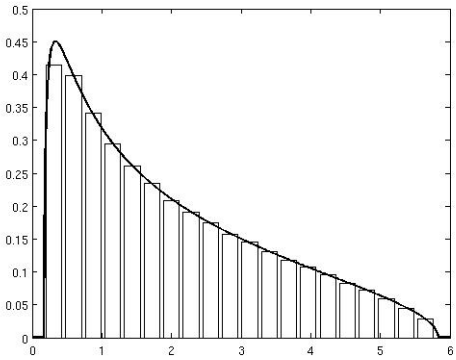
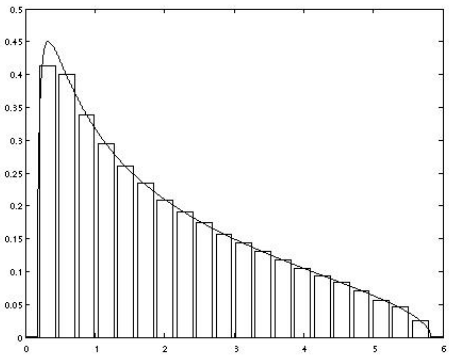
semicircle law given by the densitiy

$$\rho(t) = \frac{1}{2\pi}\sqrt{4-t^2}.$$

The following picture shows two realizations of 4000×4000 Gaussian random matrices, compared to the semicircle.

Example 1.2. Now we consider a **Wishart random matrix**; this is of the form $A = XX^*$, where X is an $N \times M$ matrix with independent Gausssian entries. If we keep the ratio M/N fixed, its eigenvalue distribution converges almost surely to the **Marchenko-Pastur distribution**. The following picture shows two realizations for $N = 3000, M = 6000$.

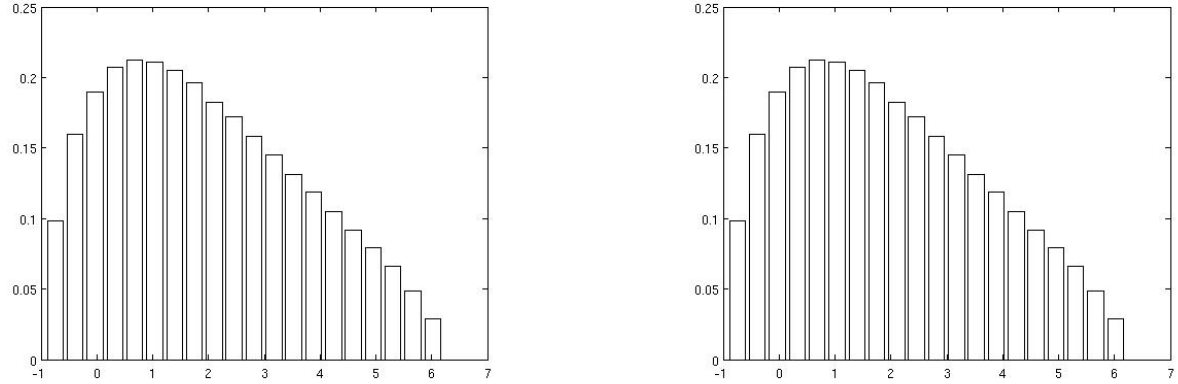
1.2 Functions of Several Random Matrices

As we saw it's possible to calculate the eigenvalue distribution of some special random matrices. But what happens if we take the sum or product of two such matrices? Or what happens for example if we take corners of Wishart random matrices? Generally spoken we want to study the eigenvalue distribution of non-trivial functions $f(A, B)$ (like $A + B$ or AB) of two $N \times N$ random matrices A and B. Such a distribution will of course depend on the relation between the eigenspaces of A and B. However we might expect that we have almost sure convergence to a deterministic result if $N \to \infty$ and if the eigenspaces are almost surely in a "typical" or "generic" position. This is the realm of **free probability theory**.

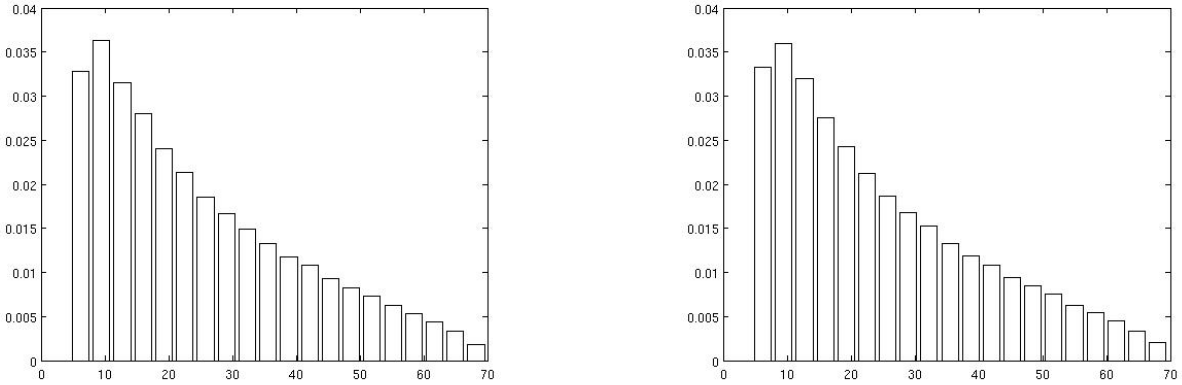
Example 1.3. Consider $N \times N$ random matrices A and B such that A and B have asymptotic eigenvalue distributions for $N \to \infty$, A and B are independent (i.e., entries of A are independent from entries of B) and B is an unitarily invariant ensemble (i.e., the joint distribution of its entries does not change under unitary conjugation). Then, almost surely, the eigenspaces of A and of B are in generic position.

In such a generic case we expect that the asymptotic eigenvalue distribution of functions of A and B should almost surely depend in a deterministic way on the asymptotic eigenvalue distribution of A and the asymptotic eigenvalue distribution of B. Let us look on some examples, where we compare two different realisations of the same situation.

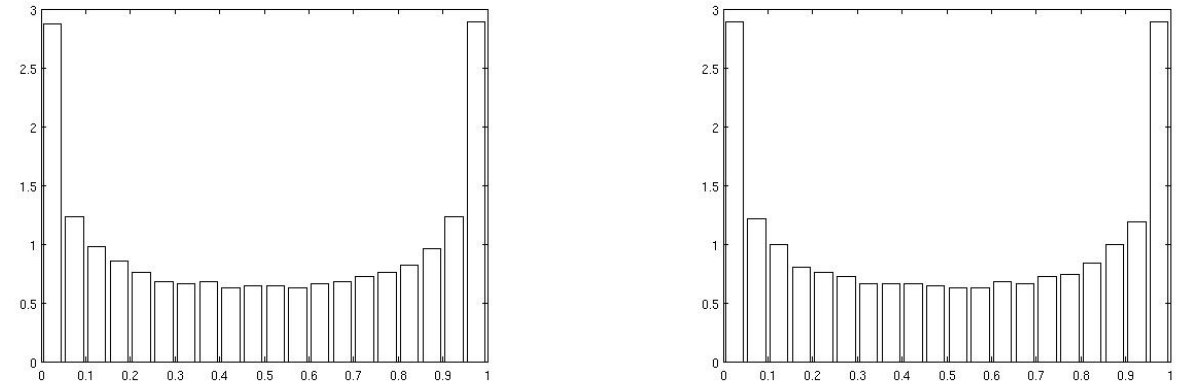
Example 1.4. Sum of independent Gaussian and Wishart ($M = 2N$) random matrices, for $N = 3000$



Example 1.5. Product of two independent Wishart ($M = 5N$) random matrices, for $N = 2000$



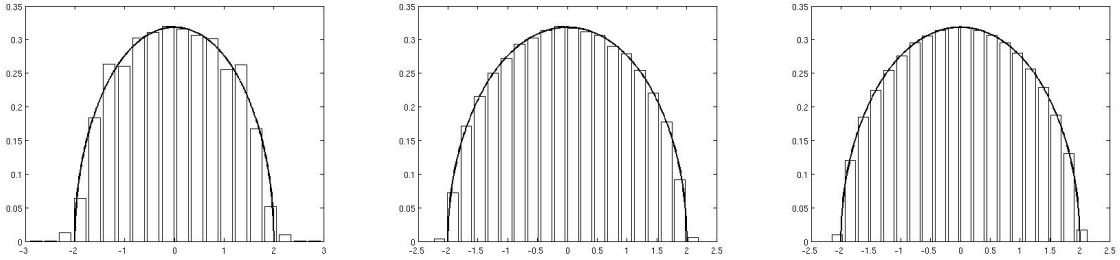
Example 1.6. Upper left corner of size $N/2 \times N/2$ of a randomly rotated $N \times N$ projection matrix, with half of the eigenvalues 0 and half of the eigenvalues 1, for $N = 2048$



Our goal is to get a **conceptual way of understanding** the asymptotic eigenvalue distributions in such cases and also to find an **algorithm for calculating** the corresponding asymptotic eigenvalue distributions. Instead of eigenvalue distributions of typical realizations we will now look at eigenvalue distributions averaged over the ensembles. One advantage of doing this is the much faster convergence to an asymptotic eigenvalue distribution and it's also theoretically easier to deal with the averaged situation. Note however, this is just for convenience; the following can also be justified for typical realizations. (In order to see that averaged convergence can be refined to almost sure convergence one has to derive sufficiently good estimates for the

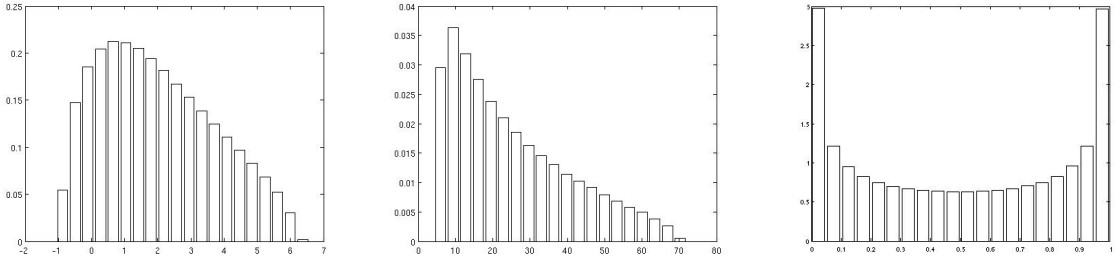
variance of the considered quantities - this can usually be achieved.)

Example 1.7. Convergence of eigenvalue distribution of $N \times N$ Gaussian random matrix, averaged over 10.000 realizations, to the semicircle,



$N = 5$ $N = 20$ $N = 50$

Example 1.8. Sums, products, corners, averaged over 10.000 realizations, for moderate $N = 100$



1.3 The Moment Method

There are different methods to analyze limit distributions. One analytical technique is the so called resolvent method. The main idea of this method is to derive an equation for the resolvent of the limit distribution. The advantage of this method is that there is a powerful complex analysis machinery to deal with such equations. This method also allows us to look at probability measures without moments. On the other side one cannot deal directly with several matrices A, B; one has to treat each function $f(A, B)$ separately. But there is also a combinatorical ansatz to calculate all moments of the limit distribution. This allows us, in principle, to deal directly with several matrices A and B by looking on mixed moments. In these lectures we will concentrate on the moment method.

By $\mathrm{tr}(A)$ we denote the normalized trace of an $N \times N$ matrix A. If we want to understand the eigenvalue distribution of a matrix A, it suffices to know the trace $\mathrm{tr}(A^k)$ of all powers of A. Because of the invariance of the trace under conjugation, we have for $k \in \mathbb{N}$ that

$$\frac{1}{N}\left(\lambda_1^k + \cdots + \lambda_N^k\right) = \mathrm{tr}(A^k).$$

Therefore instead of studying the averaged eigenvalue distribution of a random matrix A we can look at expectations of traces of powers, $E[\mathrm{tr}(A^k)]$.

Now we consider random matrices A and B in generic position and we want to understand $f(A, B)$ in a uniform way for many f. The above considerations indicate that we have to understand for all $k \in \mathbb{N}$ the moments $E\left[\mathrm{tr}\left(f(A,B)^k\right)\right]$. For example if we want to analyze the distribution of $A+B$, AB, $AB-BA$, etc. we have to look at moments $E\left[\mathrm{tr}\left((A+B)^k\right)\right]$, $E\left[\mathrm{tr}\left((AB)^k\right)\right]$, $E\left[\mathrm{tr}\left((AB-BA)^k\right)\right]$ etc. Thus we need to understand as basic objects **mixed moments** $E\left[\mathrm{tr}\left(A^{n_1}B^{m_1}A^{n_2}B^{m_2}\cdots\right)\right]$. In the following we will use the notation $\varphi(A) := \lim_{N\to\infty} E[\mathrm{tr}(A)]$. With this notation the goal is to understand $\varphi\left(A^{n_1}B^{m_1}A^{n_2}B^{m_2}\cdots\right)$ in terms of $\left(\varphi(A^k)\right)_{k\in\mathbb{N}}$ and $\left(\varphi(B^k)\right)_{k\in\mathbb{N}}$. In order to get an idea how this might look like in a generic situation, we will consider the simplest case of two such random matrices.

1.4 The Example of Two Independent Gaussian Random Matrices

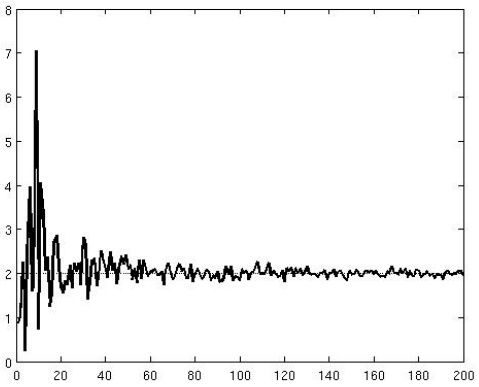
Consider two independent Gaussian random matrices A and B. Then it is fairly easy to see (and folklore in physics) that in the limit $N \to \infty$ the moment $\varphi\left(A^{n_1}B^{m_1}A^{n_2}B^{m_2}\cdots\right)$ is given by the number of non-crossing/planar pairings of the pattern

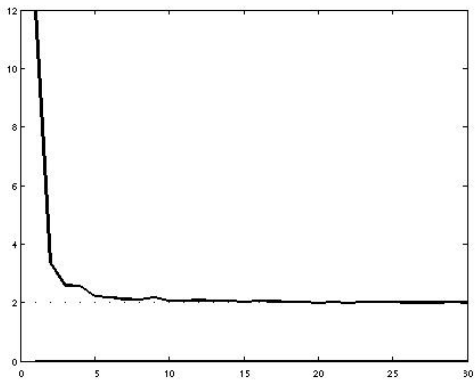
$$\underbrace{A\cdot A\cdots A}_{n_1\text{-times}}\cdot\underbrace{B\cdot B\cdots B}_{m_1\text{-times}}\cdot\underbrace{A\cdot A\cdots A}_{n_2\text{-times}}\cdot\underbrace{B\cdot B\cdots B}_{m_2\text{-times}}\cdots,$$

which do not pair A with B.

For example we have $\varphi(AABBABBA) = 2$ because there are two such non-crossing pairings:

The following pictures shows the value of $\mathrm{tr}(A_N A_N B_N B_N A_N B_N B_N A_N)$ as a function of N, calculated for two indendent Gaussian random matrices A_N and B_N; the first picture is for one realization of the situation, whereas in the second we have, for each N, averaged over 100 realizations.





Let us now come back to the general situation: $\varphi\left(A^{n_1}B^{m_1}A^{n_2}B^{m_2}\cdots\right)$ is the number of non-crossing pairings which do not pair A with B. Some musing about this reveals that this implies that $\varphi[(A^{n_1} - \varphi(A^{n_1})\cdot 1)\cdot(B^{m_1} - \varphi(B^{m_1})\cdot 1)\cdot(A^{n_2} - \varphi(A^{n_2})\cdot 1)\cdots]$ is exactly the number of non-crossing pairings which do not pair A with B and for which, in addition, each group of A's and each group of B's is connected with some other group. However, since our pairs are not allowed to cross, this number is obviously zero.

1.5 Freeness and Random Matrices

One might wonder: why did we trade an explicit formula for an implicit one? The reason is that the actual equation for the calculation of mixed moments $\varphi\left(A^{n_1}B^{m_1}A^{n_2}B^{m_2}\cdots\right)$ is different for different random matrix models. However, the relation $\varphi[(A^{n_1} - \varphi(A^{n_1})\cdot 1)\cdot(B^{m_1} - \varphi(B^{m_1})\cdot 1)\cdot(A^{n_2} - \varphi(A^{n_2})\cdot 1)\cdots] = 0$ between mixed moments remains the same for matrix ensembles in generic position and constitutes the definition of **freeness**. This was introduced by Voiculescu in 1985 and can be regarded as the starting point of free probability theory.

Definition 1.9. A and B are **free** (with respect to φ) if we have for all $n_1, m_1, n_2, \cdots \geq 1$ that

$$\varphi\Big(\big(A^{n_1} - \varphi(A^{n_1})\cdot 1\big)\cdot\big(B^{m_1} - \varphi(B^{m_1})\cdot 1\big)\cdot\big(A^{n_2} - \varphi(A^{n_2})\cdot 1\big)\cdots\Big) = 0$$

and

$$\varphi\Big(\big(B^{n_1} - \varphi(B^{n_1})\cdot 1\big)\cdot\big(A^{m_1} - \varphi(A^{m_1})\cdot 1\big)\cdot\big(B^{n_2} - \varphi(B^{n_2})\cdot 1\big)\cdots\Big) = 0.$$

(With the above we mean of course a finite number of factors.)

This definition of freeness makes our notion of "generic position" rigorous. What was just a vague idea in Example 1.3 becomes now a theorem.

Theorem 1.10 (Voiculescu 1991)**.** *Consider $N \times N$ random matrices A and B such that*

- *A and B have an asymptotic eigenvalue distribution for $N \to \infty$,*
- *A and B are independent,*
- *B is a unitarily invariant ensemble.*

Then, for $N \to \infty$, A and B are free.

2 Free Probability and Non-crossing Partitions

As we mentioned in the last chapter, the theory of asymptotically large random matrices is closeley related to free probability. The starting point of free probability was the definition of freeness, given by Voiculescu in 1985. However, this happened in the context of operator algebras, related to the isomorphism problem of free group factors. Only a few years later, in 1991, Voiculescu dicovered the relation between random matrices and free probability, as outlined in the last chapter. These connections between operator algebras and random matrices lead, among others, to deep results on free group factors. In 1994 Speicher developped a combinatorical theory of freeness, based on free cumulants. In the following we concentrate on this combinatorial way of understanding freeness.

2.1 Definition of Freeness

Let $(\mathcal{A}, \phi)$ be a **non-commutative probability space**, i.e., $\mathcal{A}$ is a unital algebra and $\phi : \mathcal{A} \to \mathbb{C}$ is a unital linear functional (i.e., $\phi(1) = 1$).

Definition 2.1. Unital subalgebras $\mathcal{A}_i$ $(i \in I)$ are **free** or **freely independent**, if $\phi(a_1 \cdots a_n) = 0$ whenever

- $a_i \in \mathcal{A}_{j(i)}$ with $j(i) \in I$, for $i = 1, \ldots, n$
- $j(1) \neq j(2) \neq \cdots \neq j(n)$
- $\phi(a_i) = 0 \quad \forall i$

Random variables $x_1, \ldots, x_n \in \mathcal{A}$ are free, if their generated unital subalgebras $\mathcal{A}_i := \text{algebra}(1, x_i)$ are so.

Remark 2.2. Freeness between A and B is, by definition, an infinite set of equations relating various moments in A and B. However, one should notice that freeness between A and B is actually a **rule for calculating mixed moments** in A and B from the moments of A and the moments of B. The following example shows such a calculation. That this works also for general mixed moments should be clear. Hence, freeness is a rule for calculating mixed moments, analogous to the concept of independence for random variables. Thus freeness is also called **free independence**.

Example 2.3. We want to calculate the mixed moments $\varphi(A^n B^m)$ of some free random variables A and B. By freeness it follows that $\phi[(A^n - \varphi(A^n)1)(B^m - \varphi(B^m)1)] = 0$. Thus we get by using the properties of our expectation ϕ the equation

$$\varphi(A^n B^m) - \varphi(A^n \cdot 1)\varphi(B^m) - \varphi(A^n)\varphi(1 \cdot B^m) + \varphi(A^n)\varphi(B^m)\varphi(1 \cdot 1) = 0,$$

and hence

$$\varphi(A^n B^m) = \varphi(A^n) \cdot \varphi(B^m).$$

Remark 2.4. The above is the same result as for independent classical random variables. However, this is misleading. Free independence is a different rule from classical independence; free independence occurs typically for **non-commuting random variables**, like operators on Hilbert spaces or (random) matrices.

Example 2.5. Let A and B be some free random variables. By definition of freeness we get

$$\varphi[(A-\varphi(A)1)\cdot(B-\varphi(B)1)\cdot(A-\varphi(A)1)\cdot(B-\varphi(B)1)]=0,$$

which results in

$$\varphi(ABAB)=\varphi(AA)\cdot\varphi(B)\cdot\varphi(B)+\varphi(A)\cdot\varphi(A)\cdot\varphi(BB)-\varphi(A)\cdot\varphi(B)\cdot\varphi(A)\cdot\varphi(B).$$

We see that this result is different from the one for independent classical random variables.

2.2 Understanding the Freeness Rule: the Idea of Cumulants

The main idea in this section is to write moments in terms of other quantities, which we call **free cumulants**. We will see that freeness is much easier to describe on the level of free cumulants, namely by the vanishing of mixed cumulants. There is also a nice relation between moments and cumulants, given by summing over **non-crossing or planar partitions**

Definition 2.6. A **partition** of $\{1,\ldots,n\}$ is a decomposition $\pi=\{V_1,\ldots,V_r\}$ with

$$V_i\neq\emptyset,\qquad V_i\cap V_j=\emptyset\quad(i\neq y),\qquad \bigcup_i V_i=\{1,\ldots,n\}$$

The V_i are the **blocks** of $\pi\in\mathcal{P}(n)$. π is **non-crossing** if we do not have

$$p_1<q_1<p_2<q_2$$

such that p_1,p_2 are in a same block, q_1,q_2 are in a same block, but those two blocks are different.
By $\mathbf{NC(n)}$ we will denote the set of all non-crossing partions of $\{1,\ldots,n\}$.

Let us remark that $NC(n)$ is actually a lattice with respect to refinement order.

2.3 Moments and Cumulants

Definition 2.7. For a unital linear functional $\varphi:\mathcal{A}\to\mathbb{C}$ we define **cumulant functionals** $\kappa_n:\mathcal{A}^n\to\mathbb{C}$ (for all $n\geq 1$) as multi-linear functionals by the moment-cumulant relation

$$\varphi(A_1\cdots A_n)=\sum_{\pi\in NC(n)}\kappa_\pi[A_1,\ldots,A_n].$$

κ_π is here a product of cumulants, one term for each block of π; see the following examples.

Remark 2.8. Classical cumulants are defined by a similar formula, where only $NC(n)$ is replaced by $\mathcal{P}(n)$.

Next we want to calculate some examples for cumulants:

For $n=1$ there exists only one partition, $\sqcap\!\!\mid$, so that the first moment and the first cumulant are the same: $\varphi(A_1)=\kappa_1(A_1)$.

For $n=2$ there are two partitions, $\sqcup$ and $\sqcap\!\!\sqcap$, and also each of them is non-crossing. By the moment-cumulant formula we get

$$\varphi(A_1A_2)=\kappa_2(A_1,A_2)+\kappa_1(A_1)\kappa_1(A_2),\qquad\text{and thus}\qquad\kappa_2(A_1,A_2)=\varphi(A_1A_2)-\varphi(A_1)\varphi(A_2).$$

In the same recursive way, we are able to compute the third cumulant. There are five partitions of three elements; still, they are all non-crossing:

So the moment-cumulant formula gives

$$\begin{aligned}\varphi(A_1A_2A_3)=&\kappa_3(A_1,A_2,A_3)+\kappa_1(A_1)\kappa_2(A_2,A_3)\\&+\kappa_2(A_1,A_2)\kappa_1(A_3)+\kappa_2(A_1,A_3)\kappa_1(A_2)+\kappa_1(A_1)\kappa_1(A_2)\kappa_1(A_3)\end{aligned}$$

and hence

$$\kappa_3(A_1, A_2, A_3) = \varphi(A_1A_2A_3) - \varphi(A_1)\varphi(A_2A_3) - \varphi(A_1A_2)\varphi(A_3) - \varphi(A_1A_3)\varphi(A_2) + 2\varphi(A_1)\varphi(A_2)\varphi(A_3).$$

The first difference to the classical theory occurs now for $n = 4$; there are 15 partitions of 4 elements, but one is crossing and there are only 14 non-crossing partitions:

Hence the moment-cumulant formula yields

$$\begin{aligned}\varphi(A_1A_2A_3A_4) &= \kappa_4(A_1, A_2, A_3, A_4) + \kappa_1(A_1)\kappa_3(A_2, A_3, A_4) + \kappa_1(A_2)\kappa_3(A_1, A_3, A_4) + \kappa_1(A_3)\kappa_3(A_1, A_2, A_4) \\ &\quad + \kappa_3(A_1, A_2, A_3)\kappa_1(A_4) + \kappa_2(A_1, A_2)\kappa_2(A_3, A_4) + \kappa_2(A_1, A_4)\kappa_2(A_2, A_3) + \kappa_1(A_1)\kappa_1(A_2)\kappa_2(A_3, A_4) \\ &\quad + \kappa_1(A_1)\kappa_2(A_2, A_3)\kappa_1(A_4) + \kappa_2(A_1, A_2)\kappa_1(A_3)\kappa_1(A_4) + \kappa_1(A_1)\kappa_2(A_2, A_4)\kappa_1(A_3) \\ &\quad + \kappa_2(A_1, A_4)\kappa_1(A_2)\kappa_1(A_3) + \kappa_2(A_1, A_3)\kappa_1(A_2)\kappa_1(A_4) + \kappa_1(A_1)\kappa_1(A_2)\kappa_1(A_3)\kappa_1(A_4).\end{aligned}$$

As before, this can be resolved for κ_4 in terms of moments. Whereas κ_1, κ_2, and κ_3 are the same as the corresponding classical cumulans, the free cumulant κ_4 (and all the higher ones) is different from its classical counterpart.

2.4 Freeness $\hat{=}$ Vanishing of Mixed Cumulants

The following theorem is essential for the theory of free cumulants.

Theorem 2.9 (Speicher 1994)**.** *The fact that A and B are free is equivalent to the fact that $\kappa_n(C_1, \ldots, C_n) = 0$ whenever*

- *$n \geq 2$,*
- *$C_i \in \{A, B\}$ for all i,*
- *there are i, j such that $C_i = A$, $C_j = B$.*

This theorem states losely spoken that the free product of some random variables can be understood as a direct sum of cumulants. For A and B free, $\varphi(A^{n_1}B^{m_1}A^{n_2}\cdots)$ is given by a sum over planar diagrams where the blocks do not connect free random variables.

Example 2.10. If A and B are free, then we have

$$\varphi(ABAB) = \kappa_1(A)\kappa_1(A)\kappa_2(B, B) + \kappa_2(A, A)\kappa_1(B)\kappa_1(B) + \kappa_1(A)\kappa_1(B)\kappa_1(A)\kappa_1(B),$$

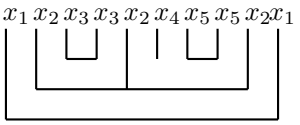
corresponding to the three non-crossing partitions of $ABAB$ which connect A with A and B with B:

2.5 Factorization of Non-Crossing Moments

The iteration of the rule

$$\varphi(A_1BA_2) = \varphi(A_1A_2)\varphi(B) \qquad \text{if } \{A_1, A_2\} \text{ and } B \text{ free}$$

leads to the simple factorization of all "non-crossing" moments in free variables. For example, if $x_1, \ldots, x_5$ are free, then we have for the moment corresponding to



the factorization $\varphi(x_1x_2x_3x_3x_2x_4x_5x_5x_2x_1) = \varphi(x_1x_1) \cdot \varphi(x_2x_2x_2) \cdot \varphi(x_3x_3) \cdot \varphi(x_4) \cdot \varphi(x_5x_5)$. This is the same as for independent classical random variables. The difference between classical and free shows only up for "crossing moments".

3 Sum of Free Variables

Let A, B be two free random variables. Then, by freeness, the moments of $A+B$ are uniquely determined by the moments of A and the moments of B. But is there an effective way to calculate the distribution of $A+B$ if we know the distribution of A and the distribution of B?

3.1 Free Convolution

Before we answer this question we want to fix some notation: We say the distribution of $A+B$ is the **free convolution** of the distribution μ_A of A and the distribution μ_B of B and denote it by $\mu_{A+B} = \mu_A \boxplus \mu_B$.

In principle, freeness determines this, but the concrete nature of this distribution on the level of moments is not apriori clear.

Example 3.1. We compute some moments of $A+B$:

$$\begin{aligned}
\varphi\big((A+B)^1\big) &= \varphi(A) + \varphi(B) \\
\varphi\big((A+B)^2\big) &= \varphi(A^2) + 2\varphi(A)\varphi(B) + \varphi(B^2) \\
\varphi\big((A+B)^3\big) &= \varphi(A^3) + 3\varphi(A^2)\varphi(B) + 3\varphi(A)\varphi(B^2) + \varphi(B^3) \\
\varphi\big((A+B)^4\big) &= \varphi(A^4) + 4\varphi(A^3)\varphi(B) + 4\varphi(A^2)\varphi(B^2) + 2\big(\varphi(A^2)\varphi(B)\varphi(B) + \varphi(A)\varphi(A)\varphi(B^2) \\
&\qquad - \varphi(A)\varphi(B)\varphi(A)\varphi(B)\big) + 4\varphi(A)\varphi(B^3) + \varphi(B^4)
\end{aligned}$$

We see that there is no "obvious" rule to compute the moments of $A+B$ out of the moments of A and the moments of B.

However, by our last theorem, there is an easy rule on the level of free cumulants: if A and B are free then

$$\kappa_n(A+B, A+B, \ldots, A+B) = \kappa_n(A, A, \ldots, A) + \kappa_n(B, B, \ldots, B)$$

because all mixed cumulants in A and B vanish.

Proposition 3.2. *If A and B are free random variables, then the relation $\kappa_n^{A+B} = \kappa_n^A + \kappa_n^B$ holds for all $n \geq 1$.*

The combinatorial relation between moments and cumulants can also be rewritten easily as a relation between corresponding formal power series.

3.2 Relation between Moments and Free Cumulants

We denote the n-th moment of A by $m_n := \varphi(A^n)$ and the n-th free cumulant of A by $\kappa_n := \kappa_n(A, A, \ldots, A)$. Then, the combinatorical relation between them is given by the **moment-cumulant formula**

$$m_n = \varphi(A^n) = \sum_{\pi \in NC(n)} \kappa_\pi.$$

Example 3.3. The third moment of a random variable A is given in terms of its free cumulants by

$$m_3 = \kappa_{\sqcup\!\sqcup\!\sqcup} + \kappa_{\sqcup\!\sqcap} + \kappa_{\sqcap\!\sqcup} + \kappa_{\sqcup\!\sqcup} + \kappa_{\sqcap\!\sqcap} = \kappa_3 + 3\kappa_2\kappa_1 + \kappa_1^3$$

The next theorem is an important step to develop an effective algorithm for computing the distribution of $A+B$.

Theorem 3.4 (Speicher 1994)**.** *Consider formal power series $M(z) = 1 + \sum_{k=1}^{\infty} m_n z^n$ and $C(z) = 1 + \sum_{k=1}^{\infty} \kappa_n z^n$. Then the relation $m_n = \sum_{\pi \in NC(n)} \kappa_\pi$ between the coefficients is equivalent to the relation $M(z) = C[zM(z)]$.*

Proof. A non-crossing partition can be described by its first block and by the non-crossing partitions of the points between the legs of the first block. This leads to the following recursive relation between cumulants and moments.

$$m_n = \sum_{\pi \in NC(n)} \kappa_\pi = \sum_{s=1}^{n} \sum_{\substack{i_1,\ldots,i_s \geq 0 \\ i_1+\cdots+i_s+s=n}} \sum_{\pi_1 \in NC(i_1)} \cdots \sum_{\pi_s \in NC(i_s)} \kappa_s \kappa_{\pi_1} \cdots \kappa_{\pi_s} = \sum_{s=1}^{n} \sum_{\substack{i_1,\ldots,i_s \geq 0 \\ i_1+\cdots+i_s+s=n}} \kappa_s m_{i_1} \cdots m_{i_s}$$

Plugging this into the formal power series $M(z)$ gives

$$M(z) = 1 + \sum_n m_n z^n = 1 + \sum_n \sum_{s=1}^{n} \sum_{\substack{i_1,\ldots,i_s \geq 0 \\ i_1+\cdots+i_s+s=n}} k_s z^s m_{i_1} z^{i_1} \cdots m_{i_s} z^{i_s} = 1 + \sum_{s=1}^{\infty} \kappa_s z^s \big(M(z)\big)^s = C[zM(z)]$$

□

Remark 3.5. Classical cumulants c_k are combinatorially defined by $m_n = \sum_{\pi \in \mathcal{P}(n)} c_\pi$. In terms of exponential generating power series $\tilde{M}(z) = 1 + \sum_{n=1}^{\infty} \frac{m_n}{n!} z^n$ and $\tilde{C}(z) = \sum_{n=1}^{\infty} \frac{c_n}{n!} z^n$ this is equivalent to $\tilde{C}(z) = \log \tilde{M}(z)$.

3.3 The Cauchy transform

For a free random variable A we define the Cauchy transform G by

$$G(z) := \varphi(\frac{1}{z - A}) = \int \frac{1}{z - t} d\mu_A(t).$$

We can expand this into a formal power series and get

$$G(z) := \sum \frac{\varphi(A^n)}{z^{n+1}} = \frac{1}{z} M(1/z).$$

Therefore, instead of $M(z)$, we can consider the Cauchy transform. This transform has many advantages over $M(z)$. If μ_A is a probability measure, its Cauchy transform is an analytic function $G : \mathbb{C}^+ \to \mathbb{C}^-$ and we can recover μ_A from G by using the **Stieltjes inversion formula**:

$$d\mu(t) = -\frac{1}{\pi} \lim_{\varepsilon \to 0} \Im G(t + i\varepsilon) dt.$$

3.4 The R-transform

Voiculescu defined the following variant of our cumulant generating series $C(z)$. He had shown the existence of the free cumulants of a random variable, but without having a combinatorial interpretation for them.

Definition 3.6. For a random variable $A \in \mathcal{A}$ we define it's **R-transform** by

$$R(z) = \sum_{n=1}^{\infty} \kappa_n(A, \ldots, A) z^{n-1}.$$

Then by a simple application of our last theorem we get the following result. The original proof of Voiculescu was much more analytical.

Theorem 3.7 (Voiculescu 1986, Speicher 1994)**.** *1) For a random variable we have the relation* $\frac{1}{G(z)} + R[G(z)] = z$ *between its Cauchy and R-transform.*
2) If A and B are free, then we have $R_{A+B}(z) = R_A(z) + R_B(z)$.

3.5 Calculation of Free Convolution

The relation between Cauchy transform and R-transform, and the Stieltjes inversion formula give an effective algorithm for calculating free convolutions; and thus also for the calculation of the asymptotic eigenvalue distribution of sums of random matrices in generic position. Let $\mu,\ \nu$ be probability measures on $\mathbb{R}$ with Cauchy-transform G_μ and G_ν respectively. Then we use the first part of our last theorem to calculate the corresponding R-transforms R_μ and R_ν. Then we use the identity $R_{\mu\boxplus\nu} = R_\mu + R_\nu$ and go over to $G_{\mu\boxplus\nu}$, by invoking once again the relation between R and G. Finally we use the Stieltjes inversion formula to recover $\mu \boxplus \nu$ itself.

Example 3.8. What is the Free Binomial $\left(\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_{+1}\right)^{\boxplus 2}$?
First we set

$$\mu := \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_{+1}, \qquad \nu := \mu \boxplus \mu.$$

In this example the Cauchy transform is given by

$$G_\mu(z) = \int \frac{1}{z-t} d\mu(t) = \frac{1}{2}\left(\frac{1}{z+1} + \frac{1}{z-1}\right) = \frac{z}{z^2-1}.$$

Now we apply the relation between Cauchy and R-transform to get the following algebraic equation

$$z = G_\mu[R_\mu(z) + 1/z] = \frac{R_\mu(z) + 1/z}{(R_\mu(z) + 1/z)^2 - 1}.$$

This quadratic equation can easily be solved and we get as a solution

$$R_\mu(z) = \frac{\sqrt{1+4z^2} - 1}{2z}.$$

(The other solution can be excluded by looking at the asymptotic behavior.) Now we use the additive property of the R-transform to get

$$R_\nu(z) = 2R_\mu(z) = \frac{\sqrt{1+4z^2} - 1}{z}.$$

Now

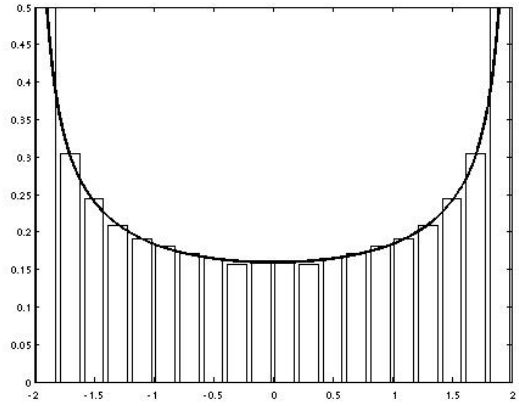
$$R_\nu(z) = \frac{\sqrt{1+4z^2} - 1}{z} \qquad \text{gives} \qquad G_\nu(z) = \frac{1}{\sqrt{z^2-4}}$$

and thus by Stieltjes inversion formula we get the **arcsine-distribution**

$$d\nu(t) = -\frac{1}{\pi}\Im\frac{1}{\sqrt{t^2-4}}dt = \begin{cases} \frac{1}{\pi\sqrt{4-t^2}}, & |t| \leq 2 \\ 0, & \text{otherwise} \end{cases}$$

The following figure compares this analytic result with the histogram of 2800 eigenvalues of $A + UBU^*$, where A and B are both diagonal matrices, with 1400 eigenvalues +1 and 1400 eigenvalues -1, and U is a

randomly chosen unitary matrix.



It is no coincidence that this figure looks similar to the ones of Example 1.6. It was shown by Nica and Speicher that the upper left corner of a randomly rotated matrix has, up to scaling, the same distribution as the free sum of two of such matrices.

3.6 The R-transform as an Analytic Object

In the last sections we only considered the R-transform as a formal power series. But for explicit calculations it is necessary to study the analytic properties of this object. It's easy to see that the R-transform can be established as an analytic function via power series expansions around the point infinity in the complex plane. But usually there are some problems in the context of concrete computations. One problem is that the R-transform can, in contrast to the Cauchy transform, in general not be defined on all of the upper complex half-plane, but only in some truncated cones (which depend on the considered variable). Another problem is that the equation $\frac{1}{G(z)} + R[G(z)] = z$ does in general not allow explicit solutions and there is no good numerical algorithm for dealing with this. Therefore one is in need of other tools, which allow to compute free convolutions in an more efficient way.

3.7 An Alternative to the R-transform: Subordination

Let x and y be free. Put $w := R_{x+y}(z) + 1/z$, then

$$G_{x+y}(w) = z = G_x[R^x(z) + 1/z] = G_x[w - R_y(z)] = G_x[w - R_y[G_{x+y}(w)]]$$

Thus with $\omega(z) := z - R_y[G_{x+y}(z)]]$ we have the subordination $G_{x+y}(z) = G_x\big(\omega(z)\big)$. It turns out that this subordination function ω is, for selfadjoint x and y, always a nice analytic object and ameanable to robust calculation algorithms.

Theorem 3.9 (Belinschi, Bercovici 2007)**.** *Let $x = x^*$ and $y = y^*$ be free. Put $F(z) := \frac{1}{G(z)}$. Then there exists an analytic $\omega : \mathbb{C}^+ \to \mathbb{C}^+$ such that*

$$F_{x+y}(z) = F_x\big(\omega(z)\big) \qquad \text{and} \qquad G_{x+y}(z) = G_x\big(\omega(z)\big)$$

The subordination function $\omega(z)$ is given as the unique fixed point in the upper half-plane of the map

$$f_z(w) = F_y(F_x(w) - w + z) - (F_x(w) - w).$$

4 Selfadjoint Polynomials in several random matrices

Our motivating problem was the asymptotic eigenvalue distribution of functions (say, selfadjoint polynomials) in several independent random matrices in generic position. A conceptual grasp on this problem was given by

the basic result of Voiculescu (1991): Large classes of independent random matrices (like Wigner or Wishart matrices) become asymptoticially freely independent, with respect to $\varphi = \frac{1}{N}\mathrm{Tr}$, if $N \to \infty$.

As a consequence, this result allows us to reduce our random matrix problem to the problem of polynomials in freely independent variables: If the random matrices $X_1, \ldots, X_k$ are asymptotically freely independent, then the eigenvalue distribution of a polynomial $p(X_1, \ldots, X_k)$ is asymptotically given by the distribution of $p(x_1, \ldots, x_k)$, where $x_1, \ldots, x_k$ are freely independent variables, and the distribution of x_i is the asymptotic distribution of X_i.

We have seen that free convolution gives effective tools for dealing with the simplest polynomial, the sum of two matrices. What can we say for more general polynomials?

4.1 Existing Results for Calculations of the Limit Eigenvalue Distribution

In the random matrix literature there exist quite a bit of results for special cases. Prominent examples are: the work of Marchenko, Pastur 1967 on general Wishart matrices ADA^*; Pastur 1972 on matrices of the form deterministic + Wigner (called "deformed semicircle"); Vasilchuk 2003 on commutator or anti-commutator of random matrices, $X_1X_2 \pm X_2X_1$; more general models in wireless communications (Tulino, Verdu 2004; Couillet, Debbah, Silverstein 2011) of the form $RADA^*R^*$ or $\sum_i R_iA_iD_iA_i^*R_i^*$.

There exist also lots of results showing us that free probability can deal effectively with simple polynomials in free variables, for example: the sum of variables $p(x, y) = x+y$ (Voiculescu 1986, R-transform); the product of variables $p(x, y) = xy$ (or, if we insist on selfadjointness, $\sqrt{x}y\sqrt{x}$) (Voiculescu 1987, S-transform); the commutator of variables $p(x, y) = i(xy - yx)$ (Nica, Speicher 1998).

But there is no hope to calculate effectively more complicated or general polynomials in freely independent variables with usual free probability theory. However there is a possible way to treat more complicated polynomials, by the use of a linearization trick. This trick will be the main topic of the next chapter.

5 The Linearization Trick

The idea of this trick is: instead of understanding general polynomials in non-commuting variables, it suffices to understand matrices of **linear** polynomials in those variables. Such linearization ideas seem to be around in many different communities. In the context of operator algebras, Voiculescu (1987)vused such a linearization philosophy as one motivation for his work on operator-valued free probability. A seminal concrete form is due to Haagerup and Thorbjørnsen (2005), who used such techniques to study the largest eigenvalue of polynomials in independent Gaussian random matrices. In 2012, based on the Schur complement, Anderson developped a self-adjoint version of the linearization trick, which turns out as the right tool in our context.

Definition 5.1. Consider a polynomial p in non-commuting variables x and y. A **linearization** of p is an $N \times N$ matrix (with $N \in \mathbb{N}$) of the form

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

where

- u, v, Q are matrices of the following sizes: u is $1 \times (N-1)$; v is $(N-1) \times N$; and Q is $(N-1) \times (N-1)$
- u, v, Q are polynomials in x and y, each of degree ≤ 1
- Q is invertible and we have

$$p = -uQ^{-1}v.$$

In 2012 Anderson presented the following theorem.

Theorem 5.2 (Anderson 2012)**.** *For each p there exists a linearization $\hat{p}$ (with an explicit algorithm for finding those). Moreover if p is selfadjoint, then this $\hat{p}$ is also selfadjoint.*

Example 5.3. We consider the selfadjoint non-commutative polynomial $p = xy + yx + x^2$. Then a linearization of p is the matrix

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

because we have

$$\begin{pmatrix} x & \frac{1}{2}x + y \end{pmatrix} \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} x \\ \frac{1}{2}x + y \end{pmatrix} = -(xy + yx + x^2)$$

At this point it might not be clear what this linearization trick has to do with our problem. What we are interested in is the distribution of p, which can be recovered from the Cauchy transform of p, which is given by taking expectations of resolvents of p. Thus we need control of inverses of p and of $z - p$. How can the linearization $\hat{p}$ give information on those?

Under the condition that Q is invertible it turns out that p is invertible iff $\hat{p}$ is invertible; this is because we can write $\hat{p}$ as

$$\hat{p} = \begin{pmatrix} 0 & u \\ v & Q \end{pmatrix} = \begin{pmatrix} 1 & uQ^{-1} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} p & 0 \\ 0 & Q \end{pmatrix} \begin{pmatrix} 1 & 0 \\ Q^{-1}v & 1 \end{pmatrix}$$

Remark 5.4. Note that matrices of the form $\begin{pmatrix} 1 & 0 \\ a & 1 \end{pmatrix}$ are always invertible with $\begin{pmatrix} 1 & 0 \\ a & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 \\ -a & 1 \end{pmatrix}$.

More general, for $z \in \mathbb{C}$ we put $b = \begin{pmatrix} z & 0 \\ 0 & 0 \end{pmatrix}$ and then it follows

$$b - \hat{p} = \begin{pmatrix} z & -u \\ -v & -Q \end{pmatrix} = \begin{pmatrix} 1 & uQ^{-1} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} z - p & 0 \\ 0 & -Q \end{pmatrix} \begin{pmatrix} 1 & 0 \\ Q^{-1}v & 1 \end{pmatrix}$$

This calculation shows that $z - p$ is invertible iff $b - \hat{p}$ is invertible. Moreover the inverse can be written as

$$\begin{aligned} (b - \hat{p})^{-1} &= \begin{pmatrix} 1 & 0 \\ -Q^{-1}v & 1 \end{pmatrix} \begin{pmatrix} (z-p)^{-1} & 0 \\ 0 & -Q^{-1} \end{pmatrix} \begin{pmatrix} 1 & -uQ^{-1} \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} (z-p)^{-1} & -(z-p)^{-1}uQ^{-1} \\ -Q^{-1}v(z-p)^{-1} & Q^{-1}v(z-p)^{-1}uQ^{-1} - Q^{-1} \end{pmatrix} \\ &= \begin{pmatrix} (z-p)^{-1} & * \\ * & * \end{pmatrix}, \end{aligned}$$

and so we can get $G_p(z) = \varphi((z-p)^{-1})$ as the (1,1)-entry of the matrix-valued Cauchy-transform

$$G_{\hat{p}}(b) = \mathrm{id} \otimes \varphi((b - \hat{p})^{-1}) = \begin{pmatrix} \varphi((z-p)^{-1}) & \varphi(*) \\ \varphi(*) & \varphi(*) \end{pmatrix}.$$

We consider again the polynomial $p = xy + yx + x^2$ of our last example. Its selfadjoint linearization can be written in the form

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

It is a linear polynomial in free variables, but with matrix-valued coefficients, and we need to calculate its matrix-valued Cauchy transform $G_{\hat{p}}(b) = \mathrm{id} \otimes \varphi((b - \hat{p})^{-1})$. This leads to the question if there exists a suitable matrix-valued version of free probability theory, with respect to the matrix-valued conditional expectation $E = \mathrm{id} \otimes \varphi$.

6 Operator-Valued Extension of Free Probability

6.1 Basic Definitions

We begin with the definition of an operator-valued probability space.

Definition 6.1. Let $\mathcal{B} \subset \mathcal{A}$ be a unital subalgebra. A linear map $E : \mathcal{A} \to \mathcal{B}$ is a **conditional expectation** if $E[b] = b$ for all $b \in \mathcal{B}$ and $E[b_1 a b_2] = b_1 E[a] b_2$ for all $a \in \mathcal{A}$ and $b_1, b_2 \in \mathcal{B}$. An **operator-valued probability space** consists of $\mathcal{B} \subset \mathcal{A}$ and a conditional expectation $E : \mathcal{A} \to \mathcal{B}$.

Example 6.2. Let $(\mathcal{A}, \varphi)$ be a non-commutative probability space. Put

$$M_2(\mathcal{A}) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mid a, b, c, d \in \mathcal{A} \right\}$$

and consider $\psi := \mathrm{tr} \otimes \varphi$ and $E := \mathrm{id} \otimes \varphi$, i.e.:

$$\psi \left[\begin{pmatrix} a & b \\ c & d \end{pmatrix} \right] = \frac{1}{2}(\varphi(a) + \varphi(d)), \qquad E \left[\begin{pmatrix} a & b \\ c & d \end{pmatrix} \right] = \begin{pmatrix} \varphi(a) & \varphi(b) \\ \varphi(c) & \varphi(d) \end{pmatrix}.$$

Then $(M_2(\mathcal{A}), \psi)$ is a non-commutative probability space, and $(M_2(\mathcal{A}), E)$ is an $M_2(\mathbb{C})$-valued probability space.

Of course we should also have a notion of distribution and freeness in the operator-valued sense.

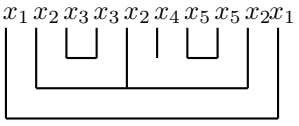
Definition 6.3. Consider an operator-valued probability space $(\mathcal{A}, E : \mathcal{A} \to \mathcal{B})$.

1. The **operator-valued distribution** of $a \in \mathcal{A}$ is given by all operator-valued moments $E[ab_1ab_2 \cdots b_{n-1}a] \in \mathcal{B}$ $(n \in \mathbb{N}, b_1, \ldots, b_{n-1} \in \mathcal{B})$.
2. Random variables $x_i \in \mathcal{A}$ $(i \in I)$ are **free with respect to** E (or **free with amalgamation over** $\mathcal{B}$) if $E[a_1 \cdots a_n] = 0$ whenever $a_i \in \mathcal{B}\langle x_{j(i)} \rangle$ are polynomials in some $x_{j(i)}$ with coefficients from $\mathcal{B}$, $E[a_i] = 0$ for all i, and $j(1) \neq j(2) \neq \cdots \neq j(n)$.

Remark 6.4. Polynomials in x with coefficients from $\mathcal{B}$ are of the form x^2, $b_0 x^2$, $b_1 x b_2 x b_3$, $b_1 x b_2 x b_3 + b_4 x b_5 x b_6 + \cdots$, etc. So we should have in mind that b's and x do not commute in general.

One can see that operator-valued freeness works mostly like ordinary freeness, one only has to take care of the order of the variables. This means in all expressions they have to appear in their original order.

Example 6.5. As in the usual scalar-valued theory one has factorizations of all non-crossing moments in free variables; but now one has to respect the order of the variables, the final expression is of a nested form, corresponding to the nesting of the non-crossing partition. Here is the operator-valued version of the example from Section 2.5.



$$E[x_1x_2x_3x_3x_2x_4x_5x_5x_2x_1] = E\Big[x_1 \cdot E\big[x_2 \cdot E[x_3x_3] \cdot x_2 \cdot E[x_4] \cdot E[x_5x_5] \cdot x_2\big] \cdot x_1\Big]$$

For "crossing" moments one also has analogous formulas as in the scalar-valued case. But again one has to take care to respect the order of the variables. For example, the formula

$$\varphi(x_1x_2x_1x_2) = \varphi(x_1x_1)\varphi(x_2)\varphi(x_2) + \varphi(x_1)\varphi(x_1)\varphi(x_2x_2) - \varphi(x_1)\varphi(x_2)\varphi(x_1)\varphi(x_2)$$

has now to be written as

$$E[x_1x_2x_1x_2] = E\big[x_1E[x_2]x_1\big] \cdot E[x_2] + E[x_1] \cdot E\big[x_2E[x_1]x_2\big] - E[x_1]E[x_2]E[x_1]E[x_2]$$

So we see that in opposite to the scalar-valued theory the freeness property in the operator valued case uses the full nested structure of non-crossing partitions.

6.2 Freeness and Matrices

It is an easy but crucial fact that freeness is compatible with going over to matrices. For example if $\{a_1, b_1, c_1, d_1\}$ and $\{a_2, b_2, c_2, d_2\}$ are free in $(\mathcal{A}, \varphi)$, then $\begin{pmatrix} a_1 & b_1 \\ c_1 & d_1 \end{pmatrix}$ and $\begin{pmatrix} a_2 & b_2 \\ c_2 & d_2 \end{pmatrix}$ are in general, not free in the scalar-valued probability space $(M_2(\mathcal{A}), \mathrm{tr} \otimes \varphi)$, but they are free with amalgamation over $M_2(\mathbb{C})$ in the operator-valued probability space $(M_2(\mathcal{A}), \mathrm{id} \otimes \varphi)$.

Example 6.6. Let $\{a_1, b_1, c_1, d_1\}$ and $\{a_2, b_2, c_2, d_2\}$ be free in $(\mathcal{A}, \varphi)$, consider

$$X_1 := \begin{pmatrix} a_1 & b_1 \\ c_1 & d_1 \end{pmatrix} \qquad \text{and} \qquad X_2 := \begin{pmatrix} a_2 & b_2 \\ c_2 & d_2 \end{pmatrix}$$

Then

$$X_1 X_2 = \begin{pmatrix} a_1 a_2 + b_1 c_2 & a_1 b_2 + b_1 d_2 \\ c_1 a_2 + d_1 c_2 & c_1 b_2 + d_1 d_2 \end{pmatrix}$$

and

$$\begin{aligned} \psi(X_1 X_2) &= \big(\varphi(a_1)\varphi(a_2) + \varphi(b_1)\varphi(c_2) + \varphi(c_1)\varphi(b_2) + \varphi(d_1)\varphi(d_2)\big)/2 \\ &\neq (\varphi(a_1) + \varphi(d_1))(\varphi(a_2) + \varphi(d_2))/4 \\ &= \psi(X_1) \cdot \psi(X_2) \end{aligned}$$

but

$$E(X_1 X_2) = \begin{pmatrix} \varphi(a_1 a_2 + b_1 c_2) & \varphi(a_1 b_2 + b_1 d_2) \\ \varphi(c_1 a_2 + d_1 c_2) & \varphi(c_1 b_2 + d_1 d_2) \end{pmatrix} = \begin{pmatrix} \varphi(a_1) & \varphi(b_1) \\ \varphi(c_1) & \varphi(d_1) \end{pmatrix} \begin{pmatrix} \varphi(a_2) & \varphi(b_2) \\ \varphi(c_2) & \varphi(d_2) \end{pmatrix} = E(X_1) \cdot E(X_2).$$

Note that there is no comparable classical statement. Matrices of independent random variables do not show any reasonable structure, not even in an "operator-valued" or "conditional" sense.

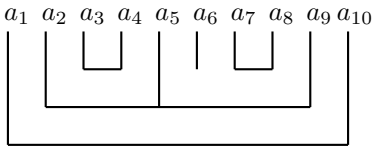
6.3 Operator-Valued Free Cumulants

Definition 6.7. Consider $E : \mathcal{A} \to \mathcal{B}$. We define the **free cumulants** $\kappa_n^{\mathcal{B}} : \mathcal{A}^n \to \mathcal{B}$ by

$$E[a_1 \cdots a_n] = \sum_{\pi \in NC(n)} \kappa_\pi^{\mathcal{B}}[a_1, \ldots, a_n].$$

Remark 6.8. Note that arguments of $\kappa_\pi^{\mathcal{B}}$ are distributed according to the blocks of π. But now the cumulants are also nested inside each other according to nesting of the blocks of π.

Example 6.9. We consider the non-crossing partition $\pi = \big\{\{1, 10\}, \{2, 5, 9\}, \{3, 4\}, \{6\}, \{7, 8\}\big\} \in NC(10)$:



Then we get $\kappa_\pi^{\mathcal{B}}[a_1, \ldots, a_{10}] = \kappa_2^{\mathcal{B}}\Big(a_1 \cdot \kappa_3^{\mathcal{B}}\big(a_2 \cdot \kappa_2^{\mathcal{B}}(a_3, a_4), a_5 \cdot \kappa_1^{\mathcal{B}}(a_6) \cdot \kappa_2^{\mathcal{B}}(a_7, a_8), a_9\big), a_{10}\Big)$.

6.4 Operator-Valued Cauchy and R-transform

Now we are in the position to define the operator-valued analogue of the Cauchy and R-transform.

Definition 6.10. For $a \in \mathcal{A}$ we define its **operator-valued Cauchy transform**

$$G_a(b) := E[\frac{1}{b-a}] = \sum_{n \geq 0} E[b^{-1}(ab^{-1})^n]$$

and **operator-valued R-transform**

$$R_a(b) := \sum_{n \geq 0} \kappa_{n+1}^{\mathcal{B}}(ab, ab, \ldots, ab, a) = \kappa_1^{\mathcal{B}}(a) + \kappa_2^{\mathcal{B}}(ab, a) + \kappa_3^{\mathcal{B}}(ab, ab, a) + \cdots$$

As in the scalar-valued case we get as a relation between those two: $bG(b) = 1 + R(G(b)) \cdot G(b)$ or $G(b) = \frac{1}{b - R(G(b))}$. If one reconsiders our combinatorial proof in the scalar-valued case, one notices that it respects the nesting of the blocks, so it works also in the operator-valued case.

If one treats these concepts on the level of formal power series one gets all the main results as in the scalar-valued case.

Theorem 6.11. *If x and y are free over $\mathcal{B}$, then*

- *mixed $\mathcal{B}$-valued cumulants in x and y vanish,*
- *it holds that $R_{x+y}(b) = R_x(b) + R_y(b)$,*
- *we have the subordination $G_{x+y}(z) = G_x(\omega(z))$.*

6.5 Free Analysis

In the last section we introduced the operator-valued R-transform and Cauchy transform on the level of formal power series. Now in order to use subordination techniques in an efficient way, we want to look at these objects in a more analytical way. This leads to the theory of "Free analysis". This subject aims at developping a non-commutative generalization of holomorphic functions in the setting of operator-valued variables (or in the setting of several variables with the highest degree of non-commutativity). Free Analysis was started by Voiculescu in the context of free probability around 2000; it builds on the seminal work of J.L. Taylor (1976): Functions of several non-commuting variables. Similar ideas are also used in work of Helton, Vinnikov etc around non-commutative convexity, linear matrix inequalities, or descriptor systems in electrical engineering.

6.6 Subordination in the Operator-Valued Case

Similar to the scalar-valued theory it is hard to deal with the R transform in an analytical way. Also, the operator-valued equation $G(b) = \frac{1}{b - R(G(b))}$ has hardly ever explicit solutions and, from the numerical point of view, it becomes quite intractable: instead of one algebraic equation we have now a system of algebraic equations. However there is a subordination version for the operator-valued case which was treated by Biane (1998) and, more conceptually, by Voiculescu (2000).

Remark 6.12. Let us present this conceptual way of looking on the subordination function. First, it is fairly easy to check that, for x and y free in $(\mathcal{A}, \varphi)$, there exists a conditional expectation (which can also be extended to formal power series) $E : \mathbb{C}\langle x, y\rangle \to \mathbb{C}\langle x\rangle$ which is determined by the requirements

$$\varphi(x^n E[f(x, y)]) = \varphi(x^n f(x, y)) \qquad \forall n \in \mathbb{N}.$$

(In the Hilbert space $L^2(\varphi)$ setting this E is just the orthogonal projection from functions in x and y to functions in x.)
Biane and Voiculescu showed now that for each $z \in \mathbb{C}$ there exists $\omega(z) \in \mathbb{C}$ such that

$$E\big[\frac{1}{z - (x+y)}\big] = \frac{1}{\omega(z) - x}.$$

This can be paraphrased by saying that best approximations in x to resolvents in $x + y$ are resolvents in x.

The following theorem shows that the analytic properties of the subordination function in the operator-valued situation are as nice as in the scalar-valued case.

Theorem 6.13 (Belinschi, Mai, Speicher 2013)**.** *Let x and y be selfadjoint operator-valued random variables which are free over B. Then there exists a Fréchet analytic map $\omega\colon \mathbb{H}^+(B) \to \mathbb{H}^+(B)$ so that*

$$G_{x+y}(b) = G_x(\omega(b)) \text{ for all } b \in \mathbb{H}^+(B).$$

Moreover, if $b \in \mathbb{H}^+(B)$, then $\omega(b)$ is the unique fixed point of the map

$$f_b\colon \mathbb{H}^+(B) \to \mathbb{H}^+(B), \quad f_b(w) = h_y(h_x(w) + b) + b,$$

and

$$\omega(b) = \lim_{n\to\infty} f_b^{\circ n}(w) \qquad \text{for any } w \in \mathbb{H}^+(B).$$

Here, $\mathbb{H}^+(B) := \{b \in B \mid (b-b^*)/(2i) > 0\}$ denotes the operator-valued upper halfplane and $h(b) := \frac{1}{G(b)} - b$.

7 Polynomials of Independent Random Matrices and Polynomials in Free Variables

Now we are able to solve the problem of calculating the distribution of a polynomial p in free variables. The idea is to linearize the polynomial and use operator-valued convolution for the linearization $\hat{p}$.

Example 7.1. The linearization of $p = xy + yx + x^2$ is given by

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

This means that the Cauchy transform $G_p(z)$ is given as the (1,1)-entry of the operator-valued (3×3 matrix) Cauchy transform of $\hat{p}$:

$$G_{\hat{p}}(b) = \mathrm{id} \otimes \varphi\left[(b - \hat{p})^{-1}\right] = \begin{pmatrix} G_p(z) & * & * \\ * & * & * \\ * & * & * \end{pmatrix} \qquad \text{for} \quad b = \begin{pmatrix} z & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

But

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

with

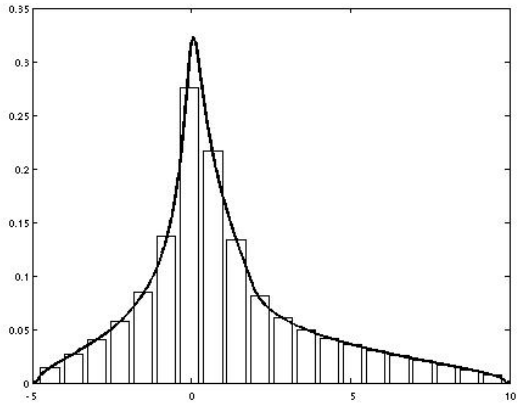
$$\hat{x} = \begin{pmatrix} 0 & x & \frac{x}{2} \\ x & 0 & 0 \\ \frac{x}{2} & 0 & 0 \end{pmatrix} \quad \text{and} \quad \hat{y} = \begin{pmatrix} 0 & 0 & y \\ 0 & 0 & -1 \\ y & -1 & 0 \end{pmatrix}.$$

According to Section 6.2, $\hat{x}$ and $\hat{y}$ are free over $M_3(\mathbb{C})$. Furthermore, the distribution of x determines the operator-valued distribution of $\hat{x}$ and the distribution of y determines the operator-valued distribution of $\hat{y}$. Thus we have the operator-valued Cauchy transforms of $\hat{x}$ and of $\hat{y}$ as input and can use our results on operator-valued free convolution to calculate the operator-valued Cauchy transform of $\hat{x} + \hat{y}$ in the

subordination form $G_{\hat{p}}(b) = G_{\hat{x}}(\omega(b))$, where $\omega(b)$ is the unique fixed point in the upper half plane $\mathbb{H}_+(M_3(\mathbb{C})$ of the iteration

$$w \mapsto G_{\hat{y}}(b + G_{\hat{x}}(w)^{-1} - w)^{-1} - (G_{\hat{x}}(w)^{-1} - w).$$

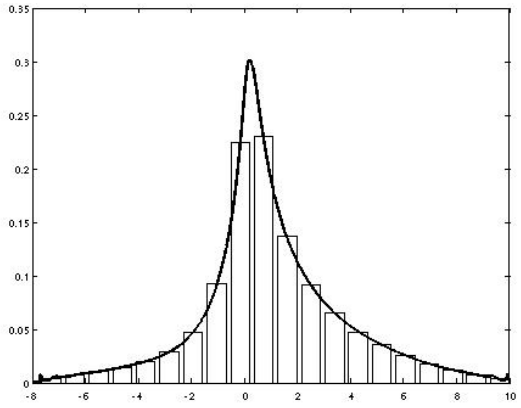
Letting this algorithm run for x semicircular and y having a Marchenko-Pastur distribution gives the following distribution for $p(x, y)$. This is compared with the histogram of a 4000×4000 random matrix $p(X, Y)$, where X and Y are independent, X is Gaussian and Y Wishart.



Example 7.2. We consider the polynomial $p(x_1, x_2, x_3) = x_1x_2x_1 + x_2x_3x_2 + x_3x_1x_3$. Then a linearization of p is given by

$$\hat{p} = \begin{pmatrix} 0 & 0 & x_1 & 0 & x_2 & 0 & x_3 \\ 0 & x_2 & -1 & 0 & 0 & 0 & 0 \\ x_1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_3 & -1 & 0 & 0 \\ x_2 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & x_1 & -1 \\ x_3 & 0 & 0 & 0 & 0 & -1 & 0 \end{pmatrix}$$

If x_1, x_2 are semicircular elements and x_3 has a Marchenko-Pastur distribution, our algorithm yields the following distribution for $p(x_1, x_2, x_3)$. Again, we compare this with the histogram of the corresponding random matrix problem $p(X_1, X_2, X_3)$; where X_1, X_2, X_3 are independent $N \times N$ random matrices; X_1 and X_2 are Gaussian and X_3; for $N = 4000$.



8 Problems (and some Solutions)

Problem 8.1. (**Gaussian Random Matrices**) Consider two Gaussian random matrices A and B. Prove that $\lim_{N\to\infty} \frac{1}{N} E[\mathrm{tr}(A^{n_1}B^{m_1}\ldots)]$ is the number of non-crossing pairings which do not pair A with B.

Problem 8.2. (Wishart Random Matrices) Consider a Wishart matrix A with $\lambda = \frac{M}{N}$. Show that

$$\lim_{N\to\infty} \frac{1}{N} E[\mathrm{tr}(A^n)] = \sum_{\pi\in NC(n)} \lambda^{|\pi|}.$$

Problem 8.3. (Formula for classical cumulants) Recall that classical cumulants c_k are combinatorially defined by $m_n = \sum_{\pi\in\mathcal{P}(n)} c_\pi$ and that we consider the power series $\tilde{M}(z) = 1 + \sum_{n=1}^{\infty} \frac{m_n}{n!} z^n$ and $\tilde{C}(z) = \sum_{n=1}^{\infty} \frac{c_n}{n!} z^n$. Show that $\tilde{C}(z) = \log \tilde{M}(z)$.

Solution: Note that

$$\exp(\tilde{C}(z)) = \sum_{k\geq 0} \sum_{r=0}^{k} \sum_{n_1+\ldots+n_r=k} \frac{1}{k!} \frac{r!}{n_1!\cdots n_r!} c_{n_1}\cdots c_{n_r} z^k$$

and we have by a combinatorial argument that

$$m_k = \sum_{r=0}^{k} \sum_{n_1+\ldots+n_r=k} \frac{r!}{n_1!\cdots n_r!} c_{n_1}\cdots c_{n_r}.$$

Problem 8.4. (R-transform of the semicircle) Consider a random variable x having the R-transform $R(z) = z$. Calculate the distribution of x.

Solution: Using the identity $\frac{1}{G(z)} + R[G(z)] = z$ we get the algebraic equation $1 + G^2(z) = zG(z)$. One solution of this equation is $G(z) = \frac{1}{2}\sqrt{z^2-4} + z$ (we can ignore the other solution by looking at the asymptotic behaviour). Applying the Stieltjes-inversion-formula gives the semicircular law.

Problem 8.5. (The Linearization Trick) We want to calculate the selfadjoint linearization of the non-commutative polynomial $p(x_1, x_2, x_3) = x_1x_2x_1 + x_2x_3x_2 + x_3x_1x_3$. For this consider the following problems.

1. Calculate the linearization of each monomial of p.
2. Given the linearizations of monomials $q_1, \ldots, q_n$, what is the linearization of $q_1 + \ldots + q_n$?
3. Consider a polynomial p of the form $q + q^*$ and let $\hat{q}$ be the linarization of q. Calculate the linearization of p in terms of $\hat{q}$.

Solution:

1. A linearization of $q = x_ix_jx_i$ is $\hat{q} = \begin{pmatrix} 0 & 0 & x_i \\ 0 & x_j & -1 \\ x_i & -1 & 0 \end{pmatrix}$.
2. We consider two linearizations $\hat{q_1} = \begin{pmatrix} 0 & u_1 \\ v_1 & Q_1 \end{pmatrix}$ and $\hat{q_2} = \begin{pmatrix} 0 & u_2 \\ v_2 & Q_2 \end{pmatrix}$. A linearization $\widehat{q_1+q_2}$ of $q_1 + q_2$ is given by $\begin{pmatrix} 0 & u_1 & u_2 \\ v_1 & Q_1 & 0 \\ v_2 & 0 & Q_2 \end{pmatrix}$.
3. If $\hat{q} = \begin{pmatrix} 0 & u \\ v & Q \end{pmatrix}$ then we can choose $\widehat{q+q^*} = \begin{pmatrix} 0 & u & v^* \\ u^* & 0 & Q \\ v & Q^* & 0 \end{pmatrix}$.

Putting these steps together gives

$$\hat{p} = \begin{pmatrix} 0 & 0 & x_1 & 0 & x_2 & 0 & x_3 \\ 0 & x_2 & -1 & 0 & 0 & 0 & 0 \\ x_1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_3 & -1 & 0 & 0 \\ x_2 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & x_1 & -1 \\ x_3 & 0 & 0 & 0 & 0 & -1 & 0 \end{pmatrix}.$$

Problem 8.6. (Freeness over $M_2(\mathbb{C})$) Let $\{a_1, b_1, c_1, d_1\}$ and $\{a_2, b_2, c_2, d_2\}$ be free in $(\mathcal{A}, \varphi)$. Show that $X_1 = \begin{pmatrix} a_1 & b_1 \\ c_1 & d_1 \end{pmatrix}$ and $X_2 = \begin{pmatrix} a_2 & b_2 \\ c_2 & d_2 \end{pmatrix}$ are free with amalgamation over $M_2(\mathbb{C})$ in $(M_2(\mathcal{A}), \mathrm{id} \otimes \varphi)$.

Problem 8.7. (Calculate the conditional expectation for the simplest non-trivial situation!) For x and y free, and E the conditional expectation from polynomials in x and y to polynomials in x, calculate (or guess and verify)

$$E[xyxy] = ?$$

Problem 8.8. (Classical analogue of conditional expectation and subordination) Let x and y be classical, commuting random variables, which are independent. Then there exists a conditional expectation

$$E : \mathbb{C}[x, y] \to \mathbb{C}[x]$$

such that

$$\varphi(x^n E[f(x, y)]) = \varphi(x^n f(x, y)) \qquad \forall n \in \mathbb{N}$$

Determine E and show that for each $z \in \mathbb{C}$ there is an $\omega(z) \in \mathbb{C}$ such that

$$E\left[ze^{x+y}\right] = \omega(z)e^x$$

Thus, in the classical world, the best approximations in x to exponential functions in $x + y$ are exponential functions in x.

Problem 8.9. Why is the exponential functional the analogue of the resolvent?

- Recall that the only solutions of $\frac{d}{dx}f(x) = f(x)$ are given by $f(x) = ze^x$ for some $z \in \mathbb{C}$.
- Show that the only solutions of $\partial_x f(x) = f(x) \otimes f(x)$ are given by $f(x) = \frac{1}{z-x}$ for some $z \in \mathbb{C}$, where $\partial_x : \mathbb{C}\langle x \rangle \to \mathbb{C}\langle x \rangle \otimes \mathbb{C}\langle x \rangle$ is the *non-commutative derivative* with respect to x, given by linear extension of

$$\partial_x 1 = 0, \qquad \partial_x x = 1 \otimes 1, \qquad \partial_x x^n = \sum_{k=0}^{n-1} x^k \otimes x^{n-1-k}$$

Solution: A simple calculation shows that the formal power series $f(x) = \sum_{k=0}^{\infty} \frac{1}{z^{k+1}} x^k$ is a solution of $\partial_x f(x) = f(x) \otimes f(x)$ for given $z \in \mathbb{C}\backslash\{0\}$. On the other hand every solution of this equation, written as a formal power series $\sum_{k=0}^{\infty} \alpha_k x^k$ has to fulfill the recursive relation $\alpha_k = \alpha_{k-1}\alpha_0$ for every $k \in \mathbb{N}$ and some $\alpha_0 \in \mathbb{C}$.

References

[1] Greg W. Anderson, *Convergence of the largest singular value of a polynomial in independent Wigner matrices*, Preprint, 2011, arXiv:1103.4825v2.

[2] Serban Belinschi, Tobias Mai, and Roland Speicher, *Analytic subordination theory of operator-valued free additive convolution and the solution of a general random matrix problem.* Preprint, 2013, arXiv:1303.3196.

[3] Alexandru Nica and Roland Speicher, *Lectures on the combinatorics of free probability.* Cambridge University Press, 2006.

[4] Roland Speicher, *Combinatorial theory of the free product with amalgamation and operator-valued free probability theory.* Mem. Amer. Math. Soc. **132** (1998), no. 627, x+88.

[5] Dan Voiculescu, *Limit laws for random matrices and free products.* Invent. Math.,**104** (1991), 201–220.

[6] Dan Voiculescu, *Operations on certain non-commutative operator-valued random variables*, Astérisque (1995), no. 232, 243–275, Recent advances in operator algebras (Orléans, 1992).

[7] Dan Voiculescu, Kenneth Dykema and Alexandru Nica, *Free random variables.* CRM Monograph Series, Vol. 1, AMS, 1992.