Second Order Freeness and Fluctuations of Random Matrices III. Higher order freeness and free cumulants

Benoît Collins[1], James A. Mingo[2],
Piotr Śniady[3], Roland Speicher[2,4]

Received: January 15, 2007

Communicated by Joachim Cuntz

Abstract. We extend the relation between random matrices and free probability theory from the level of expectations to the level of all correlation functions (which are classical cumulants of traces of products of the matrices). We introduce the notion of "higher order freeness" and develop a theory of corresponding free cumulants. We show that two independent random matrix ensembles are free of arbitrary order if one of them is unitarily invariant. We prove R-transform formulas for second order freeness. Much of the presented theory relies on a detailed study of the properties of "partitioned permutations".

2000 Mathematics Subject Classification: 46L54 (Primary), 15A52, 60F05
Keywords and Phrases: free cumulants, random matrices, planar diagrams

[1]Research supported by JSPS and COE postdoctoral fellowships

[2]Research supported by Discovery Grants and a Leadership Support Initiative Award from the Natural Sciences and Engineering Research Council of Canada

[3]Research supported by MNiSW (project 1 P03A 013 30), EU Research Training Network "QP-Applications", (HPRN-CT-2002-00279) and by European Commission Marie Curie Host Fellowship for the Transfer of Knowledge "Harmonic Analysis, Nonlinear Analysis and Probability" (MTKD-CT-2004-013389)

[4]Research supported by a Premier's Research Excellence Award from the Province of Ontario and a Killam Fellowship from the Canada Council for the Arts

1. Introduction

Random matrix models and their large dimension behavior have been an important subject of study in Mathematical Physics and Statistics since Wishart and Wigner. Global fluctuations of the eigenvalues (that is, linear functionals of the eigenvalues) of random matrices have been widely investigated in the last decade; see, e.g., [Joh98, Dia03, Rad06, AZ06, MN04, MŚS07]. Roughly speaking, the trend of these investigations is that for a wide class of converging random matrix models, the non-normalized trace asymptotically behaves like a Gaussian variable whose variance only depends on macroscopic parameters such as moments. The philosophy of these results, together with the freeness results of Voiculescu served as a motivation for our series of papers on second order freeness.

One of the main achievements of the free probability theory of Voiculescu [Voi91, VDN92] was an abstract description via the notion of "freeness" of the expectation of these Gaussian variables for a large class of non-commuting tuples of random matrices.

In the previous articles of this series [MS06, MŚS07] we showed that for many interesting ensembles of random matrices an analogue of the results of Voiculescu for expectations holds also true on the level of variances as well; thus pointing in the direction that the structure of random matrices and the fine structure of their eigenvalues can be studied in much more detail by using the new concept of "second order freeness". One of the main obstacles for such a detailed study was the absence of an effective machinery for doing concrete calculations in this framework. Within free probability theory of first order, such a machinery was provided by Voiculescu with the concept of the R-transform, and by Speicher with the concept of free cumulants; see, e.g., [VDN92, NSp06].

One of the main achievements of the present article is to develop a theory of second order cumulants (and show that the original definition of second order freeness from Part I of this series [MS06] is equivalent to the vanishing of mixed second order cumulants) and provide the corresponding R-transform machinery.

In Section 2 we will give a more detailed (but still quite condensed) survey of the connection between Voiculescu's free probability theory and random matrix theory. We will there also provide the main motivation, notions and concepts for our extension of this theory to the level of fluctuations (second order), as well as the statement of our main results concerning second order cumulants and R-transforms.

Having first and second order freeness it is, of course, a natural question whether this theory can be generalized to higher orders. It turns out that this is the case, most of the general theory is the same for all orders. So we will in this paper consider freeness of all orders from the very beginning and develop a general theory of higher order freeness and higher order cumulants. Let us, however, emphasize that first and second order freeness seem to be more important than the higher order ones. Actually, we can prove some of the most important results (e.g. the R-transform machinery) only for first and second order, mainly because of the complexity of the underlying combinatorial objects.

The basic combinatorial notion behind the (usual) free cumulants are non-crossing partitions. Basically, passage to higher order free cumulants corresponds to a change to multi-annular non-crossing permutations [MN04], or more general objects which we call "partitioned permutations". For much of the conceptual framework there is no difference between different levels of freeness, however for many concrete questions it seems that increasing the order makes some calculations much harder. This relates to the fact that n-th order freeness is described in terms of planar permutations which connect points on n different circles. Whereas enumeration of all non-crossing permutations in the case of one circle is quite easy, the case of two circles gets more complicated, but is still feasible; for the case of three or more circles, however, the answer does not seem to be of a nice compact form.

In the present paper we develop the notion and combinatorial machinery for freeness of all orders by a careful analysis of the main example: unitarily invariant random matrices. We start with the calculation of mixed correlation functions for random matrices and use the structure which we observe there as a motivation for our combinatorial setup. In this way the concept of partitioned permutations and the moment–cumulant relations appear quite canonically.

We want to point out that even though our notion of second and higher order freeness is modeled on the situation found for correlation functions of random matrices, this notion and theory also have some far-reaching applications. Let us mention in this respect two points.

Firstly, recently one of us [Śni06] developed a quite general theory for fluctuations of characters and shapes of random Young diagrams contributing to many natural representations of symmetric groups. The results presented there are closely (though, not explicitly) related to combinatorics of higher order cumulants. This connection will be studied in detail in the part IV of this series where we prove that under some mild technical conditions Jucys-Murphy elements, which arise naturally in the study of symmetric groups, are examples of free random variables of higher order.

In another direction, the description of subfactors in von Neumann algebras via planar algebras [Jon99] relies very much on the notions of annular non-crossing partitions and thus resembles the combinatorial objects lying at the basis of our theory of second order freeness. This indicates that our results could have some relevance for subfactors.

Overview of the article. In Section 2 we will give a compact survey of the connection between Voiculescu's free probability theory and random matrix theory, provide the main motivation, notions and concepts for our extension of this theory to the level of fluctuations (second order), as well as the statement of our main results concerning second order cumulants and R-transforms. We will also make a few general remarks about higher order freeness.

In Section 3 we will introduce the basic notions and relevant results on permutations, partitions, classical cumulants, Haar unitary random matrices, and the Weingarten function.
In Section 4 we study the correlation functions (classical cumulants of traces) of random matrix models. We will see how those are related to cumulants of entries of the matrices for unitarily invariant random matrices and we will in particular look on the correlation functions for products of two independent ensembles of random matrices, one of which is unitarily invariant. The limit of those formulas if the size N of the matrices goes to infinity will be the essence of what we are going to call "higher order freeness". Also our main combinatorial objects, "partitioned permutations", will arise very naturally in these calculations.
In Section 5 we will forget for a while random variables and just look on the combinatorial essence of our formulas, thus dealing with multiplicative functions on partitioned permutations and their convolution. The Zeta and Möbius functions on partitioned permutations will play an important role in these considerations.
In Section 6 we will derive, for the case of second order, the analogue of the R-transform formulas.
In Section 7 we will finally come back to a (non-commutative) probabilistic context, give the definition and work out the basic properties of "higher order freeness".
In Section 8 we introduce the notion of "asymptotic higher order freeness" and show the relevance of our work for Itzykson-Zuber integrals.
In an appendix, Section 9, we provide a graphical interpretation of partitioned permutations as a special case of "surfaced permutations".

2. Motivation and Statement of our Main Results Concerning Second Order Freeness and Cumulants

In this section we will first recall in a quite compact form the main connection between Voiculescu's free probability theory and questions about random matrices. Then we want to motivate our notion of second order freeness by extending these questions from the level of expectations to the level of fluctuations. We will recall the relevant results from the papers [MS06, MŚS07] and state the main new results of the present paper. Even though in the later parts of the paper our treatment will include freeness of arbitrarily high order, we restrict ourselves in this section mainly to the second order. The reason for this is that (apart from first order) second order freeness seems to be the most important order for applications, so that it seems worthwhile to spell out our general results for this case more explicitly. Furthermore, it is only there that we have an analogue of R-transform formulas. We will make a few general remarks about higher order freeness at the end of this section.

2.1. Moments of random matrices and asymptotic freeness. Assume we know the eigenvalue distribution of two matrices A and B. What can we say about the eigenvalue distribution of the sum $A+B$ of the matrices? Of course, the latter is not just determined by the eigenvalues of A and the eigenvalues of B, but also by the relation between the eigenspaces of A and of B. Actually, it is quite a hard problem (Horn's conjecture) — which was only solved recently — to characterize all possible eigenvalue distributions of $A+B$. However, if one is asking this question in the context of $N \times N$-random matrices, then in many situations the answer becomes deterministic in the limit $N \to \infty$.

Definition 2.1. Let $A = (A_N)_{N\in\mathbb{N}}$ be a sequence of $N \times N$-random matrices. We say that A has a *first order limit distribution* if the limit of all moments

$$\alpha_n := \lim_{N\to\infty} E[\mathrm{tr}(A_N^n)] \qquad (n \in \mathbb{N})$$

exists and for all $r > 1$ and all $n_1, \dots, n_r \in \mathbb{N}$

$$\lim_{N\to\infty} k_r(\mathrm{tr}(A_N^{n_1}), \mathrm{tr}(A_N^{n_2}), \dots, \mathrm{tr}(A_N^{n_r})) = 0,$$

where E denotes the expectation, tr the normalized trace, and k_r the r^{th} classical cumulant.

In this language, our question becomes: Given two random matrix ensembles of $N \times N$-random matrices, $A = (A_N)_{N\in\mathbb{N}}$ and $B = (B_N)_{N\in\mathbb{N}}$, with first order limit distribution, does also their sum $C = (C_N)_{N\in\mathbb{N}}$, with $C_N = A_N + B_N$, have a first order limit distribution, and furthermore, can we calculate the limit moments α_n^C of C out of the limit moments $(\alpha_k^A)_{k\geq 1}$ of A and the limit moments $(\alpha_k^B)_{k\geq 1}$ of B in a deterministic way. It turns out that this is the case if the two ensembles are in generic position, and then the rule for calculating the limit moments of C are given by Voiculescu's concept of "freeness". Let us recall this fundamental result of Voiculescu.

Theorem 2.2 (Voiculescu [Voi91]). *Let A and B be two random matrix ensembles of $N \times N$-random matrices, $A = (A_N)_{N\in\mathbb{N}}$ and $B = (B_N)_{N\in\mathbb{N}}$, each of them with a first order limit distribution. Assume that A and B are independent (i.e., for each $N \in \mathbb{N}$, all entries of A_N are independent from all entries of B_N), and that at least one of them is unitarily invariant (i.e., for each N, the joint distribution of the entries does not change if we conjugate the random matrix with an arbitrary unitary $N \times N$ matrix). Then A and B are asymptotically free in the sense of the following definition.*

Definition 2.3 (Voiculescu [Voi85]). Two random matrix ensembles $A = (A_N)_{N\in\mathbb{N}}$ and $B = (B_N)_{N\in\mathbb{N}}$ with limit eigenvalue distributions are *asymptotically free* if we have for all $p \geq 1$ and all $n(1), m(1), \dots, n(p), m(p) \geq 1$ that

$$\lim_{N\to\infty} E\Big[\mathrm{tr}\big\{(A_N^{n(1)} - \alpha_{n(1)}^A 1)\cdot(B_N^{m(1)} - \alpha_{m(1)}^B 1)\cdots \\ \cdots(A^{n(p)} - \alpha_{n(p)}^A 1)\cdot(B^{m(p)} - \alpha_{m(p)}^B 1)\big\}\Big] = 0$$

One should realize that asymptotic freeness is actually a rule which allows to calculate all mixed moments in A and B, i.e. all expressions

$$\lim_{N\to\infty} E[\mathrm{tr}(A^{n(1)}B^{m(1)}A^{n(2)}B^{m(2)}\cdots A^{n(p)}B^{m(p)})]$$

out of the limit moments of A and the limit moments of B. In particular, this means that all limit moments of $A+B$ (which are sums of mixed moments) exist and are actually determined in terms of the limit moments of A and the limit moments of B. The actual calculation rule is not directly clear from the above definition but a basic result of Voiculescu shows how this can be achieved by going over from the moments α_n to new quantities κ_n. In [Spe94], the combinatorial structure behind these κ_n was revealed and the name "free cumulants" was coined for them. Whereas in the later parts of this paper we will have to rely crucially on the combinatorial description and their extensions to higher orders, as well as on the definition of more general "mixed" cumulants, we will here state the results in the simplest possible form in terms of generating power series, which avoids the use of combinatorial objects.

DEFINITION 2.4 (Voiculescu [Voi86], Speicher [Spe94]). Given the moments $(\alpha_n)_{n\geq 1}$ of some distribution (or limit moments of some random matrix ensemble), we define the corresponding *free cumulants* $(\kappa_n)_{n\geq 1}$ by the following relation between their generating power series: If we put

$$M(x) := 1 + \sum_{n\geq 1} \alpha_n x^n \qquad \text{and} \qquad C(x) := 1 + \sum_{n\geq 1} \kappa_n x^n,$$

then we require as a relation between these formal power series that

$$C(xM(x)) = M(x).$$

Voiculescu actually formulated the relation above in a slightly different way using the so-called R-transform $\mathcal{R}(x)$, which is related to $C(x)$ by the relation

$$C(x) = 1 + x\mathcal{R}(x)$$

and in terms of the Cauchy transform $G(x)$ corresponding to a measure with moments α_n, which is related to $M(x)$ by

$$G(x) = \frac{M(\frac{1}{x})}{x}.$$

In these terms the equation $C(xM(x)) = M(x)$ says that

$$\frac{1}{G(x)} + \mathcal{R}(G(x)) = x, \tag{1}$$

i.e., that $G(x)$ and $K(x) := \frac{1}{x} + \mathcal{R}(x)$ are inverses of each other under composition.

One should also note that the relation $C(xM(x)) = M(x)$ determines the moments uniquely in terms of the cumulants and the other way around. The relevance of the κ_n and the R-transform for our problem comes from the following result of Voiculescu, which provides, together with (1), a very efficient way for calculating eigenvalue distributions of the sum of asymptotically free random matrices.

Theorem 2.5 (Voiculescu [Voi86]). *Let A and B be two random matrix ensembles which are asymptotically free. Denote by κ_n^A, κ_n^B, κ_n^{A+B} the free cumulants of A, B, $A+B$, respectively. Then one has for all $n \geq 1$ that*

$$\kappa_n^{A+B} = \kappa_n^A + \kappa_n^B.$$

Alternatively,

$$\mathcal{R}^{A+B}(x) = \mathcal{R}^A(x) + \mathcal{R}^B(x).$$

This theorem is one reason for calling the κ_n cumulants, but there is also another justification for this, namely they are also the limit of classical cumulants of the entries of our random matrix, in the case that this is unitarily invariant. This description will follow from our formulas (28) and (30). We denote the classical cumulants by k_n, considered as multi-linear functionals in n arguments.

Theorem 2.6. *Let $A = (A_N)_{N\in\mathbb{N}}$ be a unitarily invariant random matrix ensemble of $N \times N$ random matrices A_N whose first order limit distribution exists. Then the free cumulants of this matrix ensemble can also be expressed as the limit of special classical cumulants of the entries of the random matrices: If $A_N = (a_{ij}^{(N)})_{i,j=1}^N$, then*

$$\kappa_n^A = \lim_{N\to\infty} N^{n-1} k_n(a_{i(1)i(2)}^{(N)}, a_{i(2)i(3)}^{(N)}, \dots, a_{i(n),i(1)}^{(N)})$$

for any choice of distinct $i(1), \dots, i(n)$.

2.2. Fluctuations of random matrices and asymptotic second order freeness. There are many more refined questions about the limiting eigenvalue distribution of random matrices. In particular, questions around fluctuations have received a lot of interest in the last decade or so. The main motivation for introducing the concept of "second order freeness" was to understand the global fluctuations of the eigenvalues, which means that we look at the probabilistic behavior of traces of powers of our matrices. The limiting eigenvalue distribution, as considered in the last section, gives us the limit of the average of this traces. However, one can make more refined statements about their distributions. Consider a random matrix $A = (A_N)_{N\in\mathbb{N}}$ and look on the normalized traces $\mathrm{tr}(A_N^k)$. Our assumption of a limit eigenvalue distribution means that the limits $\alpha_k := \lim_{N\to\infty} E[\mathrm{tr}(A_N^k)]$ exist. It turned out that in many cases the fluctuation around this limit,

$$\mathrm{tr}(A_N^k) - \alpha_k$$

is asymptotically Gaussian of order $1/N$; i.e., the random variable

$$N \cdot (\mathrm{tr}(A_N^k) - \alpha_k) = \mathrm{Tr}(A_N^k) - N\alpha_k = \mathrm{Tr}(A_N^k - \alpha_k 1)$$

(where Tr denotes the unnormalized trace) converges for $N \to \infty$ to a normal variable. Actually, the whole family of centered unnormalized traces $(\mathrm{Tr}(A_N^k) -$

$N\alpha_k)_{k\geq 1}$ converges to a centered Gaussian family. (One should note that we restrict all our considerations to complex random matrices; in the case of real random matrices there are additional complications, which will be addressed in some future investigations.) Thus the main information about fluctuations of our considered ensemble is contained in the covariance matrix of the limiting Gaussian family, i.e., in the quantities

$$\alpha_{m,n} := \lim_{N\to\infty} \operatorname{cov}(\operatorname{Tr}(A_N^m), \operatorname{Tr}(A_N^n)).$$

Let us emphasize that the α_n and the $\alpha_{m,n}$ are actually limits of classical cumulants of traces; for the first and second order, with expectation as first and variance as second cumulant, this might not be so visible, but it will become evident when we go over to higher orders. Nevertheless, the α's will behave and will also be treated like moments; accordingly we will call the $\alpha_{m,n}$ 'fluctuation moments'. We will later define some other quantities $\kappa_{m,n}$, which take the role of cumulants in this context.
This kind of convergence to a Gaussian family was formalized in [MS06] as follows. Note that convergence to Gaussian means that all higher order classical cumulants converge to zero. As before, we denote the classical cumulants by k_n; so k_1 is just the expectation, and k_2 the covariance.

DEFINITION 2.7. Let $A = (A_N)_{N\in\mathbb{N}}$ be an ensemble of $N \times N$ random matrices A_N. We say that it has a *second order limit distribution* if for all $m, n \geq 1$ the limits

$$\alpha_n := \lim_{N\to\infty} k_1(\operatorname{tr}(A_N^n))$$

and

$$\alpha_{m,n} := \lim_{N\to\infty} k_2(\operatorname{Tr}(A_N^m), \operatorname{Tr}(A_N^n))$$

exist and if

$$\lim_{N\to\infty} k_r\big(\operatorname{Tr}(A_N^{n(1)}), \dots, \operatorname{Tr}(A_N^{n(r)})\big) = 0$$

for all $r \geq 3$ and all $n(1), \dots, n(r) \geq 1$.

We can now ask the same kind of question for the limit fluctuations as for the limit moments; namely, if we have two random matrix ensembles A and B and we know the second order limit distribution of A and the second order limit distribution of B, does this imply that we have a second order limit distribution for $A+B$, and, if so, is there an effective way for calculating it. Again, we can only hope for a positive solution to this if A and B are in a kind of generic position. As it turned out, the same requirements as before are sufficient for this. The rule for calculating mixed fluctuations constitutes the essence of the definition of the concept of second order freeness.

THEOREM 2.8 (Mingo, Śniady, Speicher [MŚS07]). *Let A and B be two random matrix ensembles of $N \times N$-random matrices, $A = (A_N)_{N\in\mathbb{N}}$ and $B = (B_N)_{N\in\mathbb{N}}$, each of them having a second order limit distribution. Assume that A and B are independent and that at least one of them is unitarily* *invariant. Then A and B are asymptotically free of second order in the sense of the following definition.*

Definition 2.9 (Mingo, Speicher [MS06]). Consider two random matrix ensembles $A = (A_N)_{N\in\mathbb{N}}$ and $B = (B_N)_{N\in\mathbb{N}}$, each of them with a second order limit distribution. Denote by

$$Y_N\big(n(1), m(1), \dots, n(p), m(p)\big)$$

the random variable

$$\mathrm{Tr}\big((A_N^{n(1)} - \alpha^A_{n(1)}1)(B_N^{m(1)} - \alpha^B_{m(1)}1)\cdots(A_N^{n(p)} - \alpha^A_{n(p)}1)(B_N^{m(p)} - \alpha^B_{m(p)}1)\big).$$

The random matrices $A = (A_N)_{N\in\mathbb{N}}$ and $B = (B_N)_{N\in\mathbb{N}}$ are *asymptotically free of second order* if for all $n, m \geq 1$

$$\lim_{N\to\infty} k_2\big(\mathrm{Tr}(A_N^n - \alpha^A_n 1), \mathrm{Tr}(B_N^m - \alpha^B_m 1)\big) = 0$$

and for all $p, q \geq 1$ and $n(1), \dots, n(p)$,$m(1), \dots, m(p)$,$\tilde{n}(1), \dots, \tilde{n}(q)$, $\tilde{m}(1), \dots, \tilde{m}(q) \geq 1$ we have

$$\lim_{N\to\infty} k_2\Big(Y_N\big(n(1), m(1), \dots, n(p), m(p)\big), Y_N\big(\tilde{n}(1), \tilde{m}(2), \dots, \tilde{n}(q), \tilde{m}(q)\big)\Big) = 0$$

if $p \neq q$, and otherwise (where we count modulo p for the arguments of the indices, i.e., $n(i + p) = n(i)$)

$$\lim_{N\to\infty} k_2\Big(Y_N\big(n(1), m(1), \dots, n(p), m(p)\big), Y_N\big(\tilde{n}(p), \tilde{m}(p), \dots, \tilde{n}(1), \tilde{m}(1)\big)\Big)$$
$$= \sum_{k=1}^{p}\prod_{i=1}^{p}\big(\alpha^A_{n(i+k)+\tilde{n}(i)} - \alpha^A_{n(i+k)}\alpha^A_{\tilde{n}(i)}\big)\big(\alpha^B_{m(i+k)+\tilde{m}(i+1)} - \alpha^B_{m(i+k)}\alpha^B_{\tilde{m}(i+1)}\big).$$

Again, it is crucial to realize that this definition allows one (albeit in a complicated way) to express every second order mixed moment, i.e., a limit of the form

$$\lim_{N\to\infty} k_2\big(\mathrm{Tr}(A_N^{n(1)}B_N^{m(1)}\cdots A_N^{n(p)}B_N^{m(p)}), \mathrm{Tr}(A_N^{\tilde{n}(1)}B_N^{\tilde{m}(1)}\cdots A_N^{\tilde{n}(q)}B_N^{\tilde{m}(q)})\big)$$

in terms of the second order limits of A and the second order limits of B. In particular, asymptotic freeness of second order also implies that the sum $A + B$ of our random matrix ensembles has a second order limit distribution and allows one to express them in principle in terms of the second order limit distribution of A and the second order limit distribution of B. As in the case of first order freeness, it is not clear at all how this calculation of the fluctuations of $A+B$ out of the fluctuations of A and the fluctuations of B can be performed effectively. It is one of the main results of the present paper to achieve such an effective description. We are able to solve this problem by providing a second order cumulant machinery, similar to the first order case. Again, the idea is to go over to quantities which behave like cumulants in this setting. The actual description of those relies on combinatorial objects (annular non-crossing permutations), but as before this can be reformulated in terms of formal power series. Let us spell out the definition here in this form. (That this is equivalent to our actual definition of the cumulants will follow from Theorem 6.3.)

DEFINITION 2.10. Let $(\alpha_n)_{n\geq 1}$ and $(\alpha_{m,n})_{m,n\geq 1}$ describe the first and second order limit moments of a random matrix ensemble. We define the corresponding *first and second order free cumulants* $(\kappa_n)_{n\geq 1}$ and $(\kappa_{m,n})_{m,n\geq 1}$ by the following requirement in terms of the corresponding generating power series. Put

$$C(x) := 1 + \sum_{n\geq 1} \kappa_n x^n, \qquad C(x,y) := \sum_{m,n\geq 1} \kappa_{m,n} x^m y^n$$

and

$$M(x) := 1 + \sum_{n\geq 1} \alpha_n x^n, \qquad M(x,y) := \sum_{m,n\geq 1} \alpha_{m,n} x^m y^n.$$

Then we require as relations between these formal power series that

$$C(xM(x)) = M(x) \tag{2}$$

and for the second order

$$M(x,y) = H\big(xM(x), yM(y)\big) \cdot \frac{\frac{d}{dx}(xM(x))}{M(x)} \cdot \frac{\frac{d}{dy}(yM(y))}{M(y)}, \tag{3}$$

where

$$H(x,y) := C(x,y) - xy\frac{\partial^2}{\partial x \partial y} \log\Big(\frac{xC(y) - yC(x)}{x-y}\Big), \tag{4}$$

or equivalently,

$$\begin{aligned} M(x,y) = C\big(xM(x), yM(y)\big) \cdot \frac{\frac{d}{dx}(xM(x))}{M(x)} \cdot \frac{\frac{d}{dy}(yM(y))}{M(y)} \\ + xy\Big(\frac{\frac{d}{dx}(xM(x)) \cdot \frac{d}{dy}(yM(y))}{(xM(x) - yM(y))^2} - \frac{1}{(x-y)^2}\Big). \end{aligned} \tag{5}$$

From equation (5) one can calculate the second order version of *moment-cumulant* relations.

$$\begin{aligned}
\alpha_{1,1} &= \kappa_{1,1} + \kappa_2 \\
\alpha_{2,1} &= \kappa_{1,2} + 2\kappa_1\kappa_{1,1} + 2\kappa_3 + 2\kappa_1\kappa_2 \\
\alpha_{2,2} &= \kappa_{2,2} + 4\kappa_1\kappa_{2,1} + 4\kappa_1^2\kappa_{1,1} + 4\kappa_4 + 8\kappa_1\kappa_3 + 2\kappa_2^2 + 4\kappa_1^2\kappa_2 \\
\alpha_{1,3} &= \kappa_{1,3} + 3\kappa_1\kappa_{2,1} + 3\kappa_2\kappa_{1,1} + 3\kappa_1^2\kappa_{1,1} + 3\kappa_4 + 6\kappa_1\kappa_3 + 3\kappa_2^2 + 3\kappa_1^2\kappa_2 \\
\alpha_{2,3} &= \kappa_{2,3} + 2\kappa_1\kappa_{1,3} + 3\kappa_1\kappa_{2,2} + 3\kappa_2\kappa_{1,2} + 9\kappa_1^2\kappa_{1,2} + 6\kappa_1\kappa_2\kappa_{1,1} + 6\kappa_1^3\kappa_{1,1} \\
&\quad + 6\kappa_5 + 18\kappa_1\kappa_4 + 12\kappa_2\kappa_3 + 18\kappa_1^2\kappa_3 + 12\kappa_1\kappa_2^2 + 6\kappa_1^3\kappa_2 \\
\alpha_{3,3} &= \kappa_{3,3} + 6\kappa_1\kappa_{2,3} + 6\kappa_2\kappa_{1,3} + 6\kappa_1^2\kappa_{1,3} + 9\kappa_1^2\kappa_{2,2} + 18\kappa_1\kappa_2\kappa_{1,2} + 18\kappa_1^3\kappa_{1,2} \\
&\quad + 9\kappa_2^2\kappa_{1,1} + 18\kappa_1^2\kappa_2\kappa_{1,1} + 9\kappa_1^4\kappa_{1,1} + 9\kappa_6 + 36\kappa_1\kappa_5 + 27\kappa_2\kappa_4 + 54\kappa_1^2\kappa_4 \\
&\quad + 9\kappa_3^2 + 72\kappa_1\kappa_2\kappa_3 + 36\kappa_1^3\kappa_3 + 12\kappa_2^3 + 36\kappa_1^2\kappa_2^2 + 9\kappa_1^4\kappa_2
\end{aligned}$$

$$\begin{aligned}
\kappa_{1,1} &= \alpha_1^2 - \alpha_2 + \alpha_{1,1} \\
\kappa_{1,2} &= -4\alpha_1^3 + 6\alpha_1\alpha_2 - 2\alpha_3 - 2\alpha_1\alpha_{1,1} + \alpha_{1,2}
\end{aligned}$$

$$\kappa_{2,2} = 18\alpha_1^4 - 36\alpha_1^2\alpha_2 + 6\alpha_2^2 + 16\alpha_1\alpha_3 - 4\alpha_4 + 4\alpha_1^2\alpha_{1,1} - 4\alpha_1\alpha_{1,2} + \alpha_{2,2}$$
$$\kappa_{1,3} = 15\alpha_1^4 - 30\alpha_1^2\alpha_2 + 6\alpha_2^2 + 12\alpha_1\alpha_3 - 3\alpha_4 + 6\alpha_1^2\alpha_{1,1} - 3\alpha_2\alpha_{1,1} - 3\alpha_1\alpha_{1,2} + \alpha_{1,3}$$
$$\kappa_{2,3} = -72\alpha_1^5 + 180\ \alpha_1^3\ \alpha_2 - 72\alpha_1\alpha_2^2 - 84\ \alpha_1^2\alpha_3 + 24\alpha_2\alpha_3 + 30\alpha_1\alpha_4 - 6\alpha_5$$
$$- 12\alpha_1^3\alpha_{1,1} + 6\alpha_1\alpha_2\alpha_{1,1} + 12\alpha_1^2\alpha_{1,2} - 3\alpha_2\alpha_{1,2} - 2\alpha_1\alpha_{1,3} - 3\alpha_1\alpha_{2,2} + \alpha_{2,3}$$
$$\kappa_{3,3} = 300\alpha_1^6 - 900\alpha_1^4\alpha_2 + 576\alpha_1^2\alpha_2^2 - 48\alpha_2^3 + 432\alpha_1^3\alpha_3 - 288\alpha_1\alpha_2\alpha_3 + 18\alpha_3^2$$
$$- 180\alpha_1^2\alpha_4 + 45\alpha_2\alpha_4 + 54\alpha_1\alpha_5 - 9\alpha_6 + 36\alpha_1^4\alpha_{1,1} - 36\alpha_1^2\alpha_2\alpha_{1,1} + 9\alpha_2^2\alpha_{1,1}$$
$$- 36\alpha_1^3\alpha_{1,2} + 18\alpha_1\alpha_2\alpha_{1,2} + 12\alpha_1^2\alpha_{1,3} - 6\alpha_2\alpha_{1,3} + 9\alpha_1^2\alpha_{2,2} - 6\alpha_1\alpha_{2,3} + \alpha_{3,3}$$

As in the first order case, instead of the moment power series $M(x,y)$ one can consider a kind of second order Cauchy transform, defined by
$$G(x,y) := \frac{M(\frac{1}{x},\frac{1}{y})}{xy}.$$
If we also define a kind of second order R transform $\mathcal{R}(x,y)$ by
$$\mathcal{R}(x,y) := \frac{1}{xy}C(x,y),$$
then the formula (5) takes on a particularly nice form:
$$G(x,y) = G'(x)G'(y)\Big\{\mathcal{R}(G(x),G(y)) + \frac{1}{(G(x)-G(y))^2}\Big\} - \frac{1}{(x-y)^2}. \tag{6}$$
$G(x)$ is here, as before, the first order Cauchy transform, $G(x) = \frac{1}{x}M(1/x)$. The $\kappa_{m,n}$ defined above deserve the name "cumulants" as they linearize the problem of adding random matrices which are asymptotically free of second order. Namely, as will follow from our Theorem 7.15, we have the following theorem, which provides, together with (6), an effective machinery for calculating the fluctuations of the sum of asymptotically free random matrices.

Theorem 2.11. *Let A and B be two random matrix ensembles which are asymptotically free. Then one has for all $m,n \geq 1$ that*
$$\kappa_n^{A+B} = \kappa_n^A + \kappa_n^B \qquad \textit{and} \qquad \kappa_{m,n}^{A+B} = \kappa_{m,n}^A + \kappa_{m,n}^B.$$
Alternatively,
$$\mathcal{R}^{A+B}(x) = \mathcal{R}^A(x) + \mathcal{R}^B(x)$$
and
$$\mathcal{R}^{A+B}(x,y) = \mathcal{R}^A(x,y) + \mathcal{R}^B(x,y).$$

Again, one can express the second order cumulants as limits of classical cumulants of entries of a unitarily invariant matrix. In contrast to the first order case, we have now to run over two disjoint cycles in the indices of the matrix entries. This theorem will follow from our formulas (28) and (30).

Theorem 2.12. *Let $A = (A_N)_{N\in\mathbb{N}}$ be a unitarily invariant random matrix ensemble which has a second order limit distribution. Then the second order free cumulants of this matrix ensemble can also be expressed as the limit of classical cumulants of the entries of the random matrices: If $A_N = (a_{ij}^{(N)})_{i,j=1}^N$, then*

$$\kappa^A_{m,n} = \lim_{N\to\infty} N^{m+n} k_{m+n}(a^{(N)}_{i(1)i(2)}, a^{(N)}_{i(2)i(3)}, \dots, a^{(N)}_{i(m),i(1)}, \\ a^{(N)}_{j(1)j(2)}, a^{(N)}_{j(2)j(3)}, \dots, a^{(N)}_{j(n),j(1)})$$

for any choice of distinct $i(1),\dots,i(m),j(1),\dots,j(n)$.

This latter theorem makes it quite obvious that the second order cumulants for Gaussian as well as for Wishart matrices vanish identically, i.e., $\mathcal{R}(x,y) = 0$ and thus we obtain in these cases that the second order Cauchy transform is totally determined in terms of the first order Cauchy transform (i.e., in terms of the limiting eigenvalue distribution) via

$$G(x,y) = \frac{G'(x)G'(y)}{(G(x)-G(y))^2} - \frac{1}{(x-y)^2}. \tag{7}$$

This formula for fluctuations of Wishart matrices was also derived by Bai and Silverstein in [BS04].

2.3. Higher order freeness. The idea for higher order freeness is the same as for second order one. For a random matrix ensemble $A = (A_N)_{N\in\mathbb{N}}$ we define r-th order limit moments as the scaled limit of classical cumulants of r traces of powers of our matrices,

$$\alpha_{n_1,\dots,n_r} := \lim_{N\to\infty} N^{r-2} k_r\big(\mathrm{Tr}(A_N^{n(1)}),\dots,\mathrm{Tr}(A_N^{n(r)})\big).$$

(The choice of N^{r-2} is motivated by the fact that this is the leading order for many interesting random matrix ensembles, e.g. Gaussian or Wishart. Thus our theory of higher order freeness captures the features of random matrix ensembles whose cumulants of traces decay in the same way as Gaussian random matrices.) Then we look at two random matrix ensembles A and B which are independent, and one of them unitarily invariant. The mixed moments in A and B of order r are, in leading order in the limit $N\to\infty$, determined by the limit moments of A up to order r and the limit moments of B up to order r. The structure of these formulas motivates directly the definition of cumulants of the considered order. The definition of those is in terms of a moment-cumulant formula, which gives a moment in terms of cumulants by summing over special combinatorial objects, which we call "partitioned permutations". Most of the theory we develop relies on an in depth analysis of properties of these partitioned permutations and the corresponding convolution of multiplicative functions on partitioned permutations. Our definition of "higher order freeness" is then in terms of the vanishing of mixed cumulants. It follows quite easily that in the first and second order case this gives the same as the relations

in Definitions 2.3 and 2.9, respectively. For higher orders, however, we are not able to find an explicit relation of that type.
This reflects somehow the observation that our general formulas in terms of sums over partitioned permutations are the same for all orders, but that evaluating or simplifying these sums (by doing partial summations) is beyond our abilities for orders greater than 2. Reformulating the combinatorial relation between moments and cumulants in terms of generating power series is one prominent example for this. Whereas this is quite easy for first order, the complexity of the arguments and the solution (given in Definition 2.10) is much higher for second order, and out of reach for higher order.
One should note that an effective (analytic or symbolic) calculation of higher order moments of a sum $A+B$ for A and B free of higher order relies usually on the presence of such generating power series formulas. In this sense, we have succeeded in providing an effective machinery for dealing with fluctuations (second order), but we were not able to do so for higher order.
Our results for higher orders are more of a theoretical nature. One of the main problems we have to address there is the associativity of the notion of higher order freeness. Namely, in order to be an interesting concept, our definition that A and B are free of higher order should of course imply that any function of A is also free of higher order from any function of B. Whereas for first and second order this follows quite easily from the equivalent characterization of freeness in terms of moments as in Definitions 2.3 and 2.9, the absence of such a characterization for higher orders makes this a more complicated matter. Namely, what we have to see is that the vanishing of mixed cumulants in random variables implies also the vanishing of mixed cumulants in elements from the generated algebras. This is quite a non-trivial fact and requires a careful analysis, see section 7.

3. Preliminaries

3.1. Some general notation. For natural numbers $m, n \in \mathbb{N}$ with $m < n$, we denote by $[m, n]$ the interval of natural numbers between m and n, i.e.,

$$[m, n] := \{m, m+1, m+2, \ldots, n-1, n\}.$$

For a matrix $A = (a_{ij})_{i,j=1}^N$, we denote by Tr the unnormalized and by tr the normalized trace,

$$\mathrm{Tr}(A) := \sum_{i=1}^N a_{ii}, \qquad \mathrm{tr}(A) := \frac{1}{N}\mathrm{Tr}(A).$$

3.2. Permutations. We will denote the set of permutations on n elements by S_n. We will quite often use the cycle notation for such permutations, i.e., $\pi = (i_1, i_2, \ldots, i_r)$ is a cycle which sends i_k to i_{k+1} $(k = 1, \ldots, r)$, where $i_{r+1} = i_1$.

3.2.1. *Length function.* For a permutation $\pi \in S_n$ we denote by $\#\pi$ the number of cycles of π and by $|\pi|$ the minimal number of transpositions needed to write π as a product of transpositions. Note that one has

$$|\pi| + \#\pi = n \qquad \text{for all } \pi \in S_n.$$

3.2.2. *Non-crossing permutations.* Let us denote by $\gamma_n \in S_n$ the cycle

$$\gamma_n = (1, 2, \ldots, n).$$

For all $\pi \in S_n$ one has that

$$|\pi| + |\gamma_n \pi^{-1}| \leq n - 1.$$

If we have equality then we call π *non-crossing.* Note that this is equivalent to

$$\#\pi + \#(\gamma_n \pi^{-1}) = n + 1.$$

If π is non-crossing, then so are $\gamma_n \pi^{-1}$ and $\pi^{-1}\gamma_n$; the latter is called the *(Kreweras) complement* of π.
We will denote the set of non-crossing permutations in S_n by $NC(n)$. Note that such a non-crossing permutation can be identified with a non-crossing partition, by forgetting the order on the cycles. There is exactly one cyclic order on the blocks of a non-crossing partition which makes it into a non-crossing permutation.

3.2.3. *Annular non-crossing permutations.* Fix $m, n \in \mathbb{N}$ and denote by $\gamma_{m,n}$ the product of the two cycles

$$\gamma_{m,n} = (1, 2, \ldots, m)(m + 1, m + 2, \ldots, m + n).$$

More generally, we shall denote by $\gamma_{m_1,\ldots,m_k}$ the product of the corresponding k cycles.
We call a $\pi \in S_{m+n}$ *connected* if the pair π and $\gamma_{m,n}$ generates a transitive subgroup in S_{m+n}. A connected permutation $\pi \in S_{m+n}$ always satisfies

$$|\pi| + |\gamma_{m,n} \pi^{-1}| \leq m + n. \tag{8}$$

If π is connected and if we have equality in that equation then we call π *annular non-crossing.* Note that if π is annular non-crossing then $\gamma_{m,n}\pi^{-1}$ is also annular non-crossing. Again, we call the latter the *complement* of π. Of course, all the above notations depend on the pair (m, n); if we want to emphasize this dependency we will also speak about (m, n)-connected permutations and (m, n)-annular non-crossing permutations.
We will denote the set of (m, n)-annular non-crossing permutations by $S_{NC}(m, n)$. A cycle of a $\pi \in S_{NC}(m, n)$ is called a *through-cycle* if it contains points on both cycles. Each $\pi \in S_{NC}(m, n)$ is connected and must thus have at least one through-cycle. The subset of $S_{NC}(m, n)$ where all cycles are through-cycles will be denoted by $S_{NC}^{all}(m, n)$.
Again one can go over from annular non-crossing permutations to annular non-crossing partitions by forgetting the cyclic orders on cycles; however, in the annular case, the relation between non-crossing permutation and non-crossing

partition is not one-to-one. Since we will not use the language of annular partitions in the present paper, this is of no relevance here.
Annular non-crossing permutations and partitions were introduced in [MN04]; there, many different characterizations—in particular, the one (8) above in terms of the length function—were given.

3.3. PARTITIONS. We say that $\mathcal{V} = \{V_1, \dots, V_k\}$ is a partition of a set $[1, n]$ if the sets V_i are disjoint and non–empty and their union is equal to $[1, n]$. We call $V_1, \dots, V_k$ the blocks of partition $\mathcal{V}$.
If $\mathcal{V} = \{V_1, \dots, V_k\}$ and $\mathcal{W} = \{W_1, \dots, W_l\}$ are partitions of the same set, we say that $\mathcal{V} \le \mathcal{W}$ if for every block V_i there exists some block W_j such that $V_i \subseteq W_j$. For a pair of partitions $\mathcal{V}, \mathcal{W}$ we denote by $\mathcal{V} \vee \mathcal{W}$ the smallest partition $\mathcal{U}$ such that $\mathcal{V} \le \mathcal{U}$ and $\mathcal{W} \le \mathcal{U}$. We denote by $1_n = \big\{[1, n]\big\}$ the biggest partition of the set $[1, n]$.
If $\pi \in S_n$ is a permutation, then we can associate to π in a natural way a partition whose blocks consist exactly of the cycles of π; we will denote this partition either by $0_\pi \in \mathcal{P}(n)$ or, if the context makes the meaning clear, just by $\pi \in \mathcal{P}(n)$.
For a permutation $\pi \in S_n$ we say that a partition $\mathcal{V}$ is π-invariant if π preserves each block of $\mathcal{V}$. This means that $0_\pi \le \mathcal{V}$ (which we will usually write just as $\pi \le \mathcal{V}$).
If $\mathcal{V} = \{V_1, \dots, V_k\}$ is a partition of the set $[1, n]$ and if, for $1 \le i \le k$, π_i is a permutation of the set V_i we denote by $\pi_1 \times \cdots \times \pi_k \in S_n$ the concatenation of these permutations. We say that $\pi = \pi_1 \times \cdots \times \pi_k$ is a cycle decomposition if additionally every factor π_i is a cycle.

3.4. CLASSICAL CUMULANTS. Given some classical probability space (Ω, P) we denote by E the expectation with respect to the corresponding probability measure,

$$\mathrm{E}(a) := \int_\Omega a(\omega) dP(\omega)$$

and by $L^{\infty-}(\Omega, P)$ the algebra of random variables for which all moments exist. Let us for the following put $\mathcal{A} := L^{\infty-}(\Omega, P)$.
We extend the linear functional $\mathrm{E} : \mathcal{A} \to \mathbb{C}$ to a corresponding multiplicative functional on all partitions by ($\mathcal{V} \in \mathcal{P}(n)$, $a_1, \dots, a_n \in \mathcal{A}$)

$$\mathrm{E}_\mathcal{V}[a_1, \dots, a_n] := \prod_{V \in \mathcal{V}} \mathrm{E}[a_1, \dots, a_n|_V], \tag{9}$$

where we use the notation

$$\mathrm{E}[a_1, \dots, a_n|_V] := \mathrm{E}(a_{i_1} \cdots a_{i_s}) \qquad \text{for} \qquad V = (i_1 < \cdots < i_s) \in \mathcal{V}.$$

Then, for $\mathcal{V} \in \mathcal{P}(n)$, we define the *classical cumulants* $k_\mathcal{V}$ as multilinear functionals on $\mathcal{A}$ by

$$k_\mathcal{V}[a_1, \dots, a_n] = \sum_{\substack{\mathcal{W} \in \mathcal{P}(n) \\ \mathcal{W} \le \mathcal{V}}} \mathrm{E}_\mathcal{W}[a_1, \dots, a_n] \cdot \mathrm{Möb}_{\mathcal{P}(n)}(\mathcal{W}, \mathcal{V}), \tag{10}$$

where $\text{Möb}_{\mathcal{P}(n)}$ denotes the Möbius function on $\mathcal{P}(n)$ (see [Rot64]).
The above definition is, by Möbius inversion on $\mathcal{P}(n)$, equivalent to

$$\mathrm{E}(a_1 \cdots a_n) = \sum_{\pi \in \mathcal{P}(n)} k_\pi[a_1, \dots, a_n].$$

The k_π are also multiplicative with respect to the blocks of $\mathcal{V}$ and thus determined by the values of

$$k_n(a_1, \dots, a_n) := k_{1_n}[a_1, \dots, a_n].$$

Note that we have in particular

$$k_1(a) = \mathrm{E}(a) \qquad \text{and} \qquad k_2(a_1, a_2) = \mathrm{E}(a_1 a_2) - \mathrm{E}(a_1)\mathrm{E}(a_2).$$

An important property of classical cumulants is the following formula of Leonov and Shiryaev [LS59] for cumulants with products as arguments.
Let $m, n \in \mathbb{N}$ and $1 \le i(1) < i(2) < \cdots < i(m) = n$. Define $\mathcal{U} \in \mathcal{P}(n)$ by

$$\mathcal{U} = \big\{\big(1, \dots, i(1)\big), \big(i(1)+1, \dots, i(2)\big), \dots, \big(i(m-1)+1, \dots, i(m)\big)\big\}.$$

Consider now random variables $a_1, \dots, a_n \in \mathcal{A}$ and define

$$\begin{aligned} A_1 &:= a_1 \cdots a_{i(1)} \\ A_2 &:= a_{i(1)+1} \cdots a_{i(2)} \\ &\vdots \\ A_m &:= a_{i(m-1)+1} \cdots a_{i(m)}. \end{aligned}$$

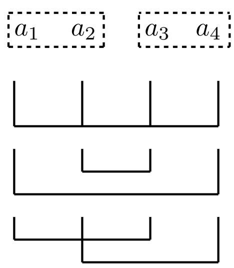
Then we have

$$k_m(A_1, A_2, \dots, A_m) = \sum_{\substack{\mathcal{V} \in \mathcal{P}(n) \\ \mathcal{V} \vee \mathcal{U} = 1_n}} k_{\mathcal{V}}[a_1, \dots, a_n]. \tag{11}$$

The sum on the right-hand side is running over those partitions of n elements which satisfy $\mathcal{V} \vee \mathcal{U} = 1_n$, which are, informally speaking, those partitions which connect all the arguments of the cumulant k_m, when written in terms of the a_i.
Here is an example for this formula; for $k_2(a_1 a_2, a_3 a_4)$. In order to reduce the number of involved terms we will restrict to the special case where $\mathrm{E}(a_i) = 0$ (and thus also $k_1(a_i) = 0$) for all $i = 1, 2, 3, 4$. There are three partitions $\pi \in \mathcal{P}(4)$ without singletons which satisfy

$$\pi \vee \{(1,2),(3,4)\} = 1_4,$$

namely



and thus formula (11) gives in this case

$$k_2(a_1a_2, a_3a_4) = k_4(a_1, a_2, a_3, a_4) \\ + k_2(a_1, a_4)k_2(a_2, a_3) + k_2(a_1, a_3)k_2(a_2, a_4).$$

As a consequence of (11) one has the following important corollary: If $\{a_1, \dots, a_n\}$ and $\{b_1, \dots, b_n\}$ are independent then

$$k_{\mathcal{W}}[a_1b_1, \dots, a_nb_n] = \sum_{\substack{\mathcal{V},\mathcal{V}' \in \mathcal{P}(n) \\ \mathcal{V}\vee\mathcal{V}'=\mathcal{W}}} k_{\mathcal{V}}[a_1, \dots, a_n] \cdot k_{\mathcal{V}'}[b_1, \dots, b_n]. \tag{12}$$

3.5. Haar distributed unitary random matrices and the Weingarten function. In the following we will be interested in the asymptotics of special matrix integrals over the group $\mathcal{U}(N)$ of unitary $N \times N$-matrices. We always equip the compact group $\mathcal{U}(N)$ with its Haar probability measure. A random matrix whose distribution is this measure will be called a *Haar distributed unitary random matrix*. Thus the expectation E over this ensemble is given by integrating with respect to the Haar measure.

The expectation of products of entries of Haar distributed unitary random matrices can be described in terms of a special function on the permutation group. Since such considerations go back to Weingarten [Wei78], Collins [Col03] calls this function the *Weingarten function* and denotes it by Wg. We will follow his notation. In the following we just recall the relevant information about this Weingarten function, for more details we refer to [Col03, CŚ06].

We use the following definition of the Weingarten function. For $\pi \in S_n$ and $N \geq n$ we put

$$\mathrm{Wg}(N, \pi) = \mathrm{E}[u_{11} \cdots u_{nn}\overline{u_{1\pi(1)}} \cdots \overline{u_{n\pi(n)}}],$$

where $U = (u_{ij})_{i,j=1}^N$ is an $N \times N$ Haar distributed unitary random matrix. Sometimes we will suppress the dependency on N and just write $\mathrm{Wg}(\pi)$. This $\mathrm{Wg}(N, \pi)$ depends only on the conjugacy class of π. General matrix integrals over the unitary group can be calculated as follows:

$$\mathrm{E}[u_{i'_1j'_1} \cdots u_{i'_nj'_n}\overline{u_{i_1j_1}} \cdots \overline{u_{i_nj_n}}] \\ = \sum_{\alpha,\beta\in S_n} \delta_{i_1i'_{\alpha(1)}} \cdots \delta_{i_ni'_{\alpha(n)}}\delta_{j_1j'_{\beta(1)}} \cdots \delta_{j_nj'_{\beta(n)}}\mathrm{Wg}(\beta\alpha^{-1}). \tag{13}$$

This formula for the calculation of moments of the entries of a Haar unitary random matrix bears some resemblance to the Wick formula for the joint moments of the entries of Gaussian random matrices; thus we will call (13) the *Wick formula for Haar unitary matrices*.

The Weingarten function is quite a complicated object, and its full understanding is at the basis of questions around Itzykson-Zuber integrals. One knows (see, e.g., [Col03, CŚ06]) that the leading order in $1/N$ is given by $|\pi| + n$ and increases in steps of 2.

3.6. Cumulants of the Weingarten function. We will also need some (classical) relative cumulants of the Weingarten function, which were introduced in [Col03, §2.3]. As before, let $\text{Möb}_{\mathcal{P}(n)}$ be the Möbius function on the partially ordered set of partitions of $[1, n]$ ordered by inclusion.

Let us first extend the Weingarten function by multiplicative extension, for $\mathcal{V} \geq \pi$, by

$$\text{Wg}(\mathcal{V}, \pi) := \prod_{V \in \mathcal{V}} \text{Wg}(\pi|_V),$$

where $\pi|_V$ denotes the restriction of π to the block $V \in \mathcal{V}$ (which is invariant under π since $\pi \leq \mathcal{V}$).

The *relative cumulant of the Weingarten function* is now, for $\sigma \leq \mathcal{V} \leq \mathcal{W}$, defined by

$$C_{\mathcal{V},\mathcal{W}}(\sigma) = \sum_{\substack{\mathcal{U} \in \mathcal{P}(n) \\ \mathcal{V} \leq \mathcal{U} \leq \mathcal{W}}} \text{Möb}(\mathcal{U}, \mathcal{W}) \cdot \text{Wg}(\mathcal{U}, \sigma). \tag{14}$$

Note that, by Möbius inversion, this is, for any $\sigma \leq \mathcal{V} \leq \mathcal{W}$, equivalent to

$$\text{Wg}(\mathcal{W}, \sigma) = \sum_{\substack{\mathcal{U} \in \mathcal{P}(n) \\ \mathcal{V} \leq \mathcal{U} \leq \mathcal{W}}} C_{\mathcal{V},\mathcal{U}}(\sigma). \tag{15}$$

In [Col03, Cor. 2.9] it was shown that the order of $C_{\mathcal{V},\mathcal{W}}(\sigma)$ is at most

$$N^{-2n + \#\sigma + 2\#\mathcal{W} - 2\#\mathcal{V}}. \tag{16}$$

4. Correlation functions for random matrices

4.1. Correlation functions and partitioned permutations. Let us consider $N \times N$-random matrices $B_1, \dots, B_n : \Omega \to M_N(\mathbb{C})$. The main information we are interested in are the "correlation functions" φ_n of these matrices, given by classical cumulants of their traces, i.e.,

$$\varphi_n(B_1, \dots, B_n) := k_n(\text{Tr}(B_1), \dots, \text{Tr}(B_n)).$$

Even though these correlation functions are cumulants, it is more adequate to consider them as a kind of moments for our random matrices. Thus, we will also call them sometimes *correlation moments*.

We will also need to consider traces of products which are best encoded via permutations. Thus, for $\pi \in S_n$, $\varphi(\pi)[B_1, \dots, B_n]$ shall mean that we take cumulants of traces of products along the cycles of π. For an n-tuple $B = (B_1, \dots, B_n)$ of random matrices and a cycle $c = (i_1, i_2, \dots, i_k)$ with $k \leq n$ we denote

$$B|_c := B_{i_1} B_{i_2} \cdots B_{i_k}.$$

(We do not distinguish between products which differ by a cyclic rotation of the factors; however, in order to make this definition well-defined we could normalize our cycle $c = (i_1, i_2, \dots, i_k)$ by the requirement that i_1 is the smallest among the appearing numbers.) For any $\pi \in S(n)$ and any n-tuple $B = (B_1, \dots, B_n)$ of random matrices we put

$$\varphi(\pi)[B_1, \dots, B_n] := \varphi_r(B|_{c_1}, \dots, B|_{c_r}),$$

where π consists of the cycles $c_1, \dots, c_r$.
Example:

$$\begin{aligned}\varphi((1,3)(2,5,4))[B_1,B_2,B_3,B_4,B_5] &= \varphi_2(B_1B_3, B_2B_5B_4)\\ &= k_2(\mathrm{Tr}(B_1B_3), \mathrm{Tr}(B_2B_5B_4))\end{aligned}$$

Furthermore, we also need to consider more general products of such $\varphi(\pi)$'s. In order to index such products we will use pairs $(\mathcal{V}, \pi)$ where π is, as above, an element in S_n, and $\mathcal{V} \in \mathcal{P}(n)$ is a partition which is compatible with the cycle structure of π, i.e., each block of $\mathcal{V}$ is fixed under π, or to put it another way, $\mathcal{V} \geq \pi$. In the latter inequality we use the convention that we identify a permutation with the partition corresponding to its cycles if this identification is obvious from the structure of the formula; we will write this partition 0_π or just 0 if no confusion will result.

NOTATION 4.1. A *partitioned permutation* is a pair $(\mathcal{V}, \pi)$ consisting of $\pi \in S_n$ and $\mathcal{V} \in \mathcal{P}(n)$ with $\mathcal{V} \geq \pi$. We will denote the set of partitioned permutations of n elements by $\mathcal{PS}(n)$. We will also put

$$\mathcal{PS} := \bigcup_{n \in \mathbb{N}} \mathcal{PS}(n).$$

For such a $(\mathcal{V}, \pi) \in \mathcal{PS}$ we denote finally

$$\varphi(\mathcal{V}, \pi)[B_1, \dots, B_n] := \prod_{V \in \mathcal{V}} \varphi(\pi|_V)[B_1, \dots, B_n|_V].$$

Example:

$$\begin{aligned}\varphi\big(\{1,3,4\}\{2\}, (1,3)(2)(4)\big)&[B_1,B_2,B_3,B_4]\\ &= \varphi_2(B_1B_3, B_4) \cdot \varphi_1(B_2)\\ &= k_2(\mathrm{Tr}(B_1B_3), \mathrm{Tr}(B_4)) \cdot k_1(\mathrm{Tr}(B_2))\end{aligned}$$

Let us denote by Tr_σ as usual a product of traces along the cycles of σ. Then we have the relation

$$\mathrm{E}\{\mathrm{Tr}_\sigma[A_1, \dots, A_n]\} = \sum_{\substack{\mathcal{W} \in \mathcal{P}(n)\\ \mathcal{W} \geq \sigma}} \varphi(\mathcal{W}, \sigma)[A_1, \dots, A_n].$$

By using the formula (11) of Leonov and Shiryaev one sees that in terms of the entries of our matrices $B_k = (b_{ij}^{(k)})_{i,j=1}^N$ our $\varphi(\mathcal{U}, \gamma)$ can also be written as

$$\varphi(\mathcal{U}, \gamma)[B_1, \dots, B_n] = \sum_{\substack{\mathcal{V} \leq \mathcal{U}\\ \mathcal{V} \vee \gamma = \mathcal{U}}} \sum_{i(1),\dots,i(n)=1}^{N} k_{\mathcal{V}}[b^{(1)}_{i(1)i(\gamma(1))}, \dots, b^{(n)}_{i(n)i(\gamma(n))}]. \tag{17}$$

4.2. MOMENTS OF UNITARILY INVARIANT RANDOM MATRICES. For unitarily invariant random matrices there exists a definite relation between cumulants of traces and cumulants of entries. We want to work out this connection in this section. Related considerations were presented by Capitaine and Casalis in [CC06].

Definition 4.2. Random matrices $A_1, \ldots, A_n$ are called *unitarily invariant* if the joint distribution of all their entries does not change by global conjugation with any unitary matrix, i.e., if, for any unitary matrix U, the matrix-valued random variables $A_1, \ldots, A_n : \Omega \to M_N(\mathbb{C})$ have the same joint distribution as the matrix-valued random variables $UA_1U^*, \ldots, UA_nU^* : \Omega \to M_N(\mathbb{C})$.

Let $A_1, \ldots, A_n$ be unitarily invariant random matrices. We will now try expressing the microscopic quantities "cumulants of entries of the A_i" in terms of the macroscopic quantities "cumulants of traces of products of the A_i".
In order to make this connection we have to use the unitary invariance of our ensemble. By definition, this means that $A_1, \ldots, A_n$ has the same distribution as $\tilde{A}_1, \ldots, \tilde{A}_n$ where $\tilde{A}_i := UA_iU^*$. Since this holds for any unitary U, the same is true after averaging over such U, i.e., we can take in the definition of the $\tilde{A}_i$ the U as Haar distributed unitary random matrices, independent from $A_1, \ldots, A_n$. This reduces calculations for unitarily invariant ensembles essentially to properties of Haar unitary random matrices; in particular, the Wick formula for the U's implies that we have an analogous Wick formula for joint moments in the entries of the A_i. Let us write $A_k = (a_{ij}^{(k)})_{i,j=1}^N$ and $\tilde{A}_k = (\tilde{a}_{ij}^{(k)})_{i,j=1}^N$. Then we can calculate:

$$\begin{aligned}
\mathrm{E}\{a^{(1)}_{p_1r_1} \cdots\, a^{(n)}_{p_nr_n}\} &= \mathrm{E}\{\tilde{a}^{(1)}_{p_1r_1} \cdots\, \tilde{a}^{(n)}_{p_nr_n}\} \\
&= \sum_{i,j} \mathrm{E}\{u_{p_1i_1} a^{(1)}_{i_1j_1} \overline{u_{r_1j_1}} \cdots u_{p_ni_n} a^{(n)}_{i_nj_n} \overline{u_{r_nj_n}}\} \\
&= \sum_{i,j} \mathrm{E}\{u_{p_1i_1} \overline{u_{r_1j_1}} \cdots u_{p_ni_n} \overline{u_{r_nj_n}}\} \cdot \mathrm{E}\{a^{(1)}_{i_1j_1} \cdots a^{(n)}_{i_nj_n}\} \\
&= \sum_{i,j} \sum_{\pi,\sigma\in S_n} \delta_{r,p\circ\pi}\delta_{j,i\circ\sigma} \mathrm{Wg}(\sigma\pi^{-1}) \cdot \mathrm{E}\{a^{(1)}_{i_1j_1} \cdots a^{(n)}_{i_nj_n}\} \\
&= \sum_{\pi\in S_n} \delta_{r,p\circ\pi} \cdot \mathcal{G}(\pi)[A_1, \ldots, A_n],
\end{aligned}$$

where

$$\begin{aligned}
(18) \qquad \mathcal{G}(\pi)[A_1, \ldots, A_n] :&= \sum_{\sigma\in S_n} \mathrm{Wg}(\sigma\pi^{-1}) \cdot \sum_i \mathrm{E}\{a^{(1)}_{i_1 i_{\sigma(1)}} \cdots a^{(n)}_{i_n i_{\sigma(n)}}\} \\
&= \sum_{\sigma\in S_n} \mathrm{Wg}(\sigma\pi^{-1}) \cdot \mathrm{E}\{\mathrm{Tr}_\sigma[A_1, \ldots, A_n]\}. \\
&= \sum_{\sigma\in S_n} \mathrm{Wg}(\sigma\pi^{-1}) \cdot \sum_{\substack{\mathcal{W}\in\mathcal{P}(n)\\ \mathcal{W}\geq\sigma}} \varphi(\mathcal{W},\sigma)[A_1, \ldots, A_n] \\
&= \sum_{(\mathcal{W},\sigma)\in\mathcal{PS}(n)} \mathrm{Wg}(\sigma\pi^{-1}) \cdot \varphi(\mathcal{W},\sigma)[A_1, \ldots, A_n].
\end{aligned}$$

The important point here is that $\mathcal{G}(\pi)[A_1, \ldots, A_n]$ depends only on the macroscopic correlation moments of A.

We can extend the above to products of expectations by

$$\mathrm{E}_{\mathcal{V}}[a_{p_1r_1},\dots a_{p_nr_n}]=\sum_{\substack{\pi\in S_n\\ \pi\leq\mathcal{V}}}\delta_{r,p\circ\pi}\cdot\mathcal{G}(\mathcal{V},\pi)[A_1,\dots,A_n],$$

where $\mathcal{G}(\mathcal{V},\pi)$ is given by multiplicative extension:

$$\begin{aligned}\mathcal{G}(\mathcal{V},\pi)[A_1,\dots,A_n]:&=\prod_{V\in\mathcal{V}}\mathcal{G}(\pi|_V)[A_1,\dots,A_n|_V]\\ &=\sum_{\substack{(\mathcal{W},\sigma)\in\mathcal{PS}(n)\\ \mathcal{W}\leq\mathcal{V}}}\mathrm{Wg}(\mathcal{V},\sigma\pi^{-1})\cdot\varphi(\mathcal{W},\sigma)[A_1,\dots,A_n].\end{aligned}\tag{19}$$

Now we can look on the cumulants of the entries of our unitarily invariant matrices A_i; they are given by

$$\begin{aligned}k_{\mathcal{V}}\{a^{(1)}_{p_1r_1},\dots,a^{(n)}_{p_nr_n}\}&=\sum_{\substack{\mathcal{U}\in\mathcal{P}(n)\\ \mathcal{U}\leq\mathcal{V}}}\mathrm{M\ddot{o}b}_{\mathcal{P}(n)}(\mathcal{U},\mathcal{V})\cdot\mathrm{E}_{\mathcal{U}}[a^{(1)}_{p_1r_1},\dots a^{(n)}_{p_nr_n}]\\ &=\sum_{\mathcal{U}\leq\mathcal{V}}\sum_{\substack{\pi\in S_n\\ \pi\leq\mathcal{U}}}\delta_{r,p\circ\pi}\cdot\mathrm{M\ddot{o}b}_{\mathcal{P}(n)}(\mathcal{U},\mathcal{V})\cdot\mathcal{G}(\mathcal{U},\pi)[A_1,\dots,A_n]\\ &=\sum_{\substack{\pi\in S_n\\ \pi\leq\mathcal{V}}}\delta_{r,p\circ\pi}\sum_{\substack{\mathcal{U}\in\mathcal{P}(n)\\ \mathcal{V}\geq\mathcal{U}\geq\pi}}\mathrm{M\ddot{o}b}_{\mathcal{P}(n)}(\mathcal{U},\mathcal{V})\cdot\mathcal{G}(\mathcal{U},\pi)[A_1,\dots,A_n].\end{aligned}$$

With the definition

$$\kappa(\mathcal{V},\pi)[A_1,\dots,A_n]:=\sum_{\substack{\mathcal{U}\in\mathcal{P}(n)\\ \mathcal{V}\geq\mathcal{U}\geq\pi}}\mathrm{M\ddot{o}b}_{\mathcal{P}(n)}(\mathcal{U},\mathcal{V})\cdot\mathcal{G}(\mathcal{U},\pi)[A_1,\dots,A_n].\tag{20}$$

we thereby get

$$k_{\mathcal{V}}\{a^{(1)}_{p_1r_1},\dots,a^{(n)}_{p_nr_n}\}=\sum_{\substack{\pi\in S_n\\ \pi\leq\mathcal{V}}}\delta_{r,p\circ\pi}\cdot\kappa(\mathcal{V},\pi)[A_1,\dots,A_n].\tag{21}$$

It follows that

$$\begin{aligned}\varphi(\mathcal{U},\gamma)[A_1,\dots,A_n]&=\sum_{\substack{\mathcal{V}\leq\mathcal{U}\\ \mathcal{V}\vee\gamma=\mathcal{U}}}\sum_{i(1),\dots,i(n)=1}^{N}k_{\mathcal{V}}[a^{(1)}_{i(1)i(\gamma(1))},\dots,a^{(n)}_{i(n)i(\gamma(n))}]\\ &=\sum_{\substack{\mathcal{V}\leq\mathcal{U}\\ \mathcal{V}\vee\gamma=\mathcal{U}}}\sum_{i(1),\dots,i(n)=1}^{N}\sum_{\substack{\pi\in S_n\\ \pi\leq\mathcal{V}}}\delta_{i\circ\gamma,i\circ\pi}\cdot\kappa(\mathcal{V},\pi)[A_1,\dots,A_n]\\ &=\sum_{\substack{\mathcal{V}\leq\mathcal{U}\\ \mathcal{V}\vee\gamma=\mathcal{U}}}\sum_{\substack{\pi\in S_n\\ \pi\leq\mathcal{V}}}\kappa(\mathcal{V},\pi)[A_1,\dots,A_n]\cdot N^{\#(\gamma\pi^{-1})}.\end{aligned}$$

Since $\mathcal{V}\vee\gamma=\mathcal{U}$ is, under the assumption $\pi\leq\mathcal{V}$, equivalent to $\mathcal{V}\vee\gamma\pi^{-1}=\mathcal{U}$ we can write this also as

$$\varphi(\mathcal{U},\gamma)[A_1,\dots,A_n]=\sum_{\substack{(\mathcal{V},\pi)\in\mathcal{PS}(n)\\ \mathcal{V}\vee\gamma\pi^{-1}=\mathcal{U}}}\kappa(\mathcal{V},\pi)[A_1,\dots,A_n]\cdot N^{\#(\gamma\pi^{-1})}.\tag{22}$$

Remark 4.3. 1) Note that although the quantity κ is defined by (20) in terms of the macroscopic moments of the A_i, they have also a very concrete meaning in terms of cumulants of entries of the A_i. Namely, if we choose $\pi \in S_n$ and distinct $1 \leq i(1), \ldots, i(n) \leq N$ then equation (21) becomes, when we set $\mathcal{V} = 1_n$,

$$\kappa(1_n, \pi)[A_1, \ldots, A_n] = k_n\big(a^{(1)}_{i(1)i(\pi(1))}, \ldots, a^{(n)}_{i(n)i(\pi(n))}\big) \tag{23}$$

as the the only term in the sum that survives is the one for π.
2) Equation (22) should be considered as a kind of moment-cumulant formula in our context, thus it should contain all information for defining the "cumulants" κ in terms of the moments φ. Actually, we can solve this linear system of equations for κ in terms of φ, by using equation (20) to define κ and equation (19) for $\mathcal{G}$.

$$\begin{aligned}
&\kappa(\mathcal{V}, \pi)[A_1, \ldots, A_n] \\
&= \sum_{\substack{\mathcal{U} \in \mathcal{P}(n) \\ \mathcal{V} \geq \mathcal{U} \geq \pi}} \text{Möb}_{\mathcal{P}(n)}(\mathcal{U}, \mathcal{V}) \cdot \sum_{\substack{(\mathcal{W}, \sigma) \in \mathcal{PS}(n) \\ \mathcal{W} \leq \mathcal{U}}} \text{Wg}(\mathcal{U}, \sigma\pi^{-1}) \cdot \varphi(\mathcal{W}, \sigma)[A_1, \ldots, A_n] \\
&= \sum_{(\mathcal{W}, \sigma) \in \mathcal{PS}(n)} \varphi(\mathcal{W}, \sigma)[A_1, \ldots, A_n] \cdot \sum_{\substack{\mathcal{U} \in \mathcal{P}(n) \\ \mathcal{V} \geq \mathcal{U} \geq \pi \vee \mathcal{W}}} \text{Möb}_{\mathcal{P}(n)}(\mathcal{U}, \mathcal{V}) \cdot \text{Wg}(\mathcal{U}, \sigma\pi^{-1}).
\end{aligned}$$

Thus, by using the relative cumulants of the Weingarten function from (14), we get finally

$$\kappa(\mathcal{V}, \pi)[A_1, \ldots, A_n] = \sum_{\substack{(\mathcal{W}, \sigma) \in \mathcal{PS}(n) \\ \mathcal{W} \leq \mathcal{V}}} \varphi(\mathcal{W}, \sigma)[A_1, \ldots, A_n] \cdot C_{\pi \vee \mathcal{W}, \mathcal{V}}(\sigma\pi^{-1}). \tag{24}$$

3) One should also note that we have defined the Weingarten function only for $N \geq n$; thus in the above formulas we should always consider sufficiently large N. This is consistent with the observation that the system of equations (22) might not be invertible for N too small; the matrix $\big(N^{\#(\sigma\pi^{-1})}\big)_{\sigma,\pi \in S_n}$ is invertible for $N \geq n$, however, in general not for all $N < n$ (e.g, clearly not for $N = 1$). One can make sense of some formulas involving the Weingarten function also for $N < n$ (see [CŚ06]). However, since we are mainly interested in the asymptotic behavior of our formulas for $N \to \infty$, we will not elaborate on this.

4.3. PRODUCT OF TWO INDEPENDENT ENSEMBLES. Let us now calculate the correlation functions for a product of two independent ensembles $A_1, \ldots, A_n$ and $B_1, \ldots, B_n$ of random matrices, where we assume that one of them, let's say the B_i's, is unitarily invariant. We have, by using (17) and the special version (12) of the formula of Leonov and Shiryaev, the following:

$$
\begin{aligned}
&\varphi(\mathcal{U},\gamma)[A_1B_1,\dots,A_nB_n]\\
&=\sum_{\substack{i(1),\cdots,i(n)\\ j(1),\cdots,j(n)}}\ \sum_{\substack{\mathcal{V},\mathcal{V}'\le\mathcal{U}\\ \mathcal{V}\vee\mathcal{V}'\vee\gamma=\mathcal{U}}} k_{\mathcal{V}}[a^{(1)}_{j(1)i(1)},\dots,a^{(n)}_{j(n)i(n)}]\cdot k_{\mathcal{V}'}[b^{(1)}_{i(1)j(\gamma(1))},\dots,b^{(n)}_{i(n)j(\gamma(n))}]\\
&\overset{(20)}{=}\sum_{i,j}\ \sum_{\substack{\mathcal{V},\mathcal{V}'\le\mathcal{U}\\ \mathcal{V}\vee\mathcal{V}'\vee\gamma=\mathcal{U}}}\ \sum_{\substack{\pi\in S_n\\ \pi\le\mathcal{V}}}\delta_{i,j\circ\pi}\cdot\kappa(\mathcal{V},\pi)[A_1,\dots,A_n]\cdot k_{\mathcal{V}'}[b^{(1)}_{i(1)j(\gamma(1))},\dots,b^{(n)}_{i(n)j(\gamma(n))}]\\
&=\sum_{\pi\in S_n}\ \sum_{\substack{\mathcal{V}\in\mathcal{P}(n)\\ \mathcal{U}\ge\mathcal{V}\ge\pi}}\kappa(\mathcal{V},\pi)[A_1,\dots,A_n]\cdot\\
&\qquad\qquad\cdot\Big(\sum_{\substack{\mathcal{V}'\le\mathcal{U}\\ \mathcal{V}'\vee\mathcal{V}\vee\gamma=\mathcal{U}}}\ \sum_i k_{\mathcal{V}'}[b^{(1)}_{i(1)i(\pi^{-1}\gamma(1))},\dots,b^{(n)}_{i(n)i(\pi^{-1}\gamma(n))}]\Big).
\end{aligned}
$$

In order to evaluate the second factor we note first that, under the assumption $\pi\le\mathcal{V}$, the condition $\mathcal{V}'\vee\mathcal{V}\vee\gamma=\mathcal{U}$ is equivalent to $\mathcal{V}'\vee\mathcal{V}\vee\pi^{-1}\gamma=\mathcal{U}$. Next, we rewrite the sum over all $\mathcal{V}'\in\mathcal{P}(n)$ with $\mathcal{V}'\le\mathcal{U}$ and $\mathcal{V}'\vee\mathcal{V}\vee\pi^{-1}\gamma=\mathcal{U}$ as a double sum over all $\mathcal{W}\in\mathcal{P}(n)$ with $\mathcal{V}\vee\mathcal{W}=\mathcal{U}$ and all $\mathcal{V}'\in\mathcal{P}(n)$ with $\mathcal{V}'\le\mathcal{W}$ and $\mathcal{V}'\vee\pi^{-1}\gamma=\mathcal{W}$.

$$
\begin{aligned}
&\sum_{\substack{\mathcal{V}'\in\mathcal{P}(n)\\ \mathcal{V}'\le\mathcal{U},\mathcal{V}'\vee\mathcal{V}\vee\gamma=\mathcal{U}}}\ \sum_i k_{\mathcal{V}'}[b^{(1)}_{i(1)i(\pi^{-1}\gamma(1))},\dots,b^{(n)}_{i(n)i(\pi^{-1}\gamma(n))}]\\
&\qquad=\sum_{\substack{\mathcal{W}\in\mathcal{P}(n)\\ \mathcal{V}\vee\mathcal{W}=\mathcal{U}}}\ \sum_{\substack{\mathcal{V}'\le\mathcal{W}\\ \mathcal{V}'\vee\pi^{-1}\gamma=\mathcal{W}}}\ \sum_i k_{\mathcal{V}'}[b^{(1)}_{i(1)i(\pi^{-1}\gamma(1))},\dots,b^{(n)}_{i(n)i(\pi^{-1}\gamma(n))}]\\
&\qquad=\sum_{\substack{\mathcal{W}\in\mathcal{P}(n)\\ \mathcal{W}\ge\pi^{-1}\gamma,\mathcal{V}\vee\mathcal{W}=\mathcal{U}}}\varphi(\mathcal{W},\pi^{-1}\gamma)[B_1,\dots,B_n].
\end{aligned}
$$

Thus we finally get

$$
\begin{aligned}
&\varphi(\mathcal{U},\gamma)[A_1B_1,\dots,A_nB_n]\\
&=\sum_{\pi\in S_n}\ \sum_{\substack{\mathcal{V}\in\mathcal{P}(n)\\ \mathcal{U}\ge\mathcal{V}\ge\pi}}\ \sum_{\substack{\mathcal{W}\in\mathcal{P}(n)\\ \mathcal{W}\ge\pi^{-1}\gamma,\mathcal{V}\vee\mathcal{W}=\mathcal{U}}}\kappa(\mathcal{V},\pi)[A_1,\dots,A_n]\cdot\varphi(\mathcal{W},\pi^{-1}\gamma)[B_1,\dots,B_n]\\
&=\sum_{\substack{(\mathcal{V},\pi),(\mathcal{W},\sigma)\in\mathcal{PS}(n)\\ \mathcal{V}\vee\mathcal{W}=\mathcal{U},\pi\sigma=\gamma}}\kappa(\mathcal{V},\pi)[A_1,\dots,A_n]\cdot\varphi(\mathcal{W},\pi^{-1}\gamma)[B_1,\dots,B_n].
\end{aligned}
$$

Let us summarize the result of our calculations in the following theorem. In order to indicate that our main formulas are valid for any fixed N, we will decorate the relevant quantities with a superscript $^{(N)}$. Note that up to now we have not made any asymptotic consideration.

Theorem 4.4. *Let $\mathcal{M}_N := M_N\otimes L^\infty(\Omega)$ be an ensemble of $N\times N$-random matrices. Define correlation functions $\varphi_n^{(N)}$ on $\mathcal{M}_N$ by ($n\in\mathbb{N}$, $D_1,\dots,D_n\in$*

$\mathcal{M}_N$)

$$\varphi_n^{(N)}(D_1,\dots,D_n) := k_n(\mathrm{Tr}(D_1),\dots,\mathrm{Tr}(D_n)) \tag{25}$$

and corresponding "cumulant functions" $\kappa^{(N)}$ (for $n\le N$) by

$$\kappa^{(N)}(\mathcal{V},\pi)[D_1,\dots,D_n] = \sum_{\substack{\mathcal{W}\in\mathcal{P}(n),\,\sigma\in S_n\\ W\le\mathcal{V}}} \varphi^{(N)}(\mathcal{W},\sigma)[D_1,\dots,D_n]\cdot C^{(N)}_{\pi\vee\mathcal{W},\mathcal{V}}(\sigma\pi^{-1}), \tag{26}$$

or equivalently by the implicit system of equations

$$\varphi^{(N)}(\mathcal{U},\gamma)[D_1,\dots,D_n] = \sum_{\mathcal{V},\pi} \kappa^{(N)}(\mathcal{V},\pi)[D_1,\dots,D_n]\cdot N^{\#(\gamma\pi^{-1})}, \tag{27}$$

where the sum is over all $\mathcal{V}\in\mathcal{P}(n)$ all $\pi\in S_n$ such that $\pi\le\mathcal{V}$ and $\mathcal{V}\vee\gamma\pi^{-1}=\mathcal{U}$.
1) Let $\mathcal{A}_N$ be an algebra of unitarily invariant random matrices in $\mathcal{M}_N$. Then we have for all $n\le N$, all distinct $i(1),\dots,i(n)$, all $A_k=\big(a_{ij}^{(k)}\big)_{i,j=1}^N\in\mathcal{A}_N$, and all $\pi\in S_n$ that

$$\kappa^{(N)}(1_n,\pi)[A_1,\dots,A_n] = k_n\big(a^{(1)}_{i(1)i(\pi(1))},\dots,a^{(n)}_{i(n)i(\pi(n))}\big). \tag{28}$$

2) Assume that we have two subalgebras $\mathcal{A}_N$ and $\mathcal{B}_N$ of $\mathcal{M}_N$ such that
- ⋄ *$\mathcal{A}_N$ is a unitarily invariant ensemble,*
- ⋄ *$\mathcal{A}_N$ and $\mathcal{B}_N$ are independent.*

Then we have for all $n\in\mathbb{N}$ with $n\le N$ and all $A_1,\dots,A_n\in\mathcal{A}_N$ and $B_1,\dots,B_n\in\mathcal{B}_N$:

$$\begin{aligned}\varphi^{(N)}(\mathcal{U},\gamma)&[A_1B_1,\dots,A_nB_n]\\ &= \sum_{\mathcal{V},\pi,\mathcal{W},\sigma} \kappa^{(N)}(\mathcal{V},\pi)[A_1,\dots,A_n]\cdot\varphi^{(N)}(\mathcal{W},\sigma)[B_1,\dots,B_n],\end{aligned} \tag{29}$$

where the sum is over all $\mathcal{V},\mathcal{W}\in\mathcal{P}(n)$ and all $\pi,\sigma\in S_n$ such that $\pi\le\mathcal{V}$, $\sigma\le\mathcal{W}$, $\mathcal{V}\vee\mathcal{W}=\mathcal{U}$, and $\gamma=\pi\sigma$.

4.4. Large N asymptotics for moments and cumulants. Our main interest in this paper will be the large N limit of formula (29). This structure in leading order between independent ensembles of random matrices which are randomly rotated against each other will be captured in our abstract notion of higher order freeness.

Of course, now we must make an assumption about the asymptotic behavior in N of our correlation functions. We will require that the cumulants of traces of our random matrices decays in N with the same order as in the case of Gaussian or Wishart random matrices. In these cases one has very detailed "genus expansions" for those cumulants; see, e.g. [Oko00, MN04] and one knows that the n-th cumulant of unnormalized traces in polynomials of those random matrices decays like N^{2-n} (see e.g. [MS06, Thm. 3.1 and Thm. 3.5]).

Definition 4.5. Let, for each $N \in \mathbb{N}$, $B_1^{(N)}, \dots, B_r^{(N)} \subset M_N \otimes L^{\infty-}(\Omega)$ be $N \times N$-random matrices. Suppose that the leading term of the correlation moments of $B_1^{(N)}, \dots, B_r^{(N)}$ are of order $2-n$, i.e., that for all $n \in \mathbb{N}$ and all polynomials $p_1, \dots, p_t$ in r non-commuting variables the limits

$$\lim_{N\to\infty} \varphi_n^{(N)}(p_1(B_1^{(N)}, \dots, B_r^{(N)}), \dots, p_t(B_1^{(N)}, \dots, B_r^{(N)})) \cdot N^{n-2}$$

exist. Then we will say that $\{B_1^{(N)}, \dots, B_r^{(N)}\}$ has *limit distributions of all orders*. Let $\mathcal{B}$ be the free algebra generated by generators $b_1, \dots, b_r$. Then we define the *limit correlation functions* of $\mathcal{B}$ by

$$\begin{aligned}\varphi_n(p_1(b_1, \dots, b_r), \dots, p_t(b_1, \dots, b_r)) \\ = \lim_{N\to\infty} \varphi_n^{(N)}(p_1(B_1^{(N)}, \dots, B_r^{(N)}), \dots, p_t(B_1^{(N)}, \dots, B_r^{(N)})) \cdot N^{n-2}\end{aligned}$$

Note that this assumption implies that the leading term for the quantities $\varphi^{(N)}(\mathcal{V}, \pi)$ is of order $2\#(\mathcal{V}) - \#(\pi)$. Indeed, if $\mathcal{V}$ has k blocks and the i^{th} block of $\mathcal{V}$ contains r_i cycles of π then $\varphi^{(N)}(\mathcal{V}, \pi) = \varphi_{r_1} \cdots \varphi_{r_k}$ and each φ_{r_i} has order $2 - r_i$. Then the order of $\varphi^{(N)}(\mathcal{V}, \pi)$ is $(2 - r_1) + \cdots + (2 - r_k) = 2k - (r_1 + \cdots + r_k) = 2\#(\mathcal{V}) - \#(\pi)$. Thus

$$\begin{aligned}\varphi(\mathcal{V}, \pi)(p_1(b_1, \dots, b_r), \dots, p_t(b_1, \dots, b_r)) \\ = \lim_{N\to\infty} \varphi^{(N)}(\mathcal{V}, \pi)(p_1(B_1^{(N)}, \dots, B_r^{(N)}), \dots, p_t(B_1^{(N)}, \dots, B_r^{(N)})) \\ \cdot N^{-2\#(\mathcal{V})+\#(\pi)}\end{aligned}$$

From formula (27) one can deduce that the leading order of $\kappa^{(N)}(\mathcal{V}, \pi)$ is given by the term $(\mathcal{U}, \gamma) = (\mathcal{V}, \pi)$ and thus must be of order

$$N^{-n+2\#\mathcal{V}-\#\pi}.$$

(Indeed, this also follows from equation (24) and the leading order of the relative cumulant of the Weingarten function given in equation (16).)
Thus we can define the *limiting cumulant functions* to be the limit of the leading order of the cumulants by the equation

$$\kappa(\mathcal{V}, \pi)[b_1, \dots, b_n] := \lim_{N\to\infty} N^{n-2\#\mathcal{V}+\#\pi} \cdot \kappa^{(N)}(\mathcal{V}, \pi)[B_1^{(N)}, \dots, B_n^{(N)}] \tag{30}$$

When $(\mathcal{V}, \pi) = (1_n, \gamma_n)$ and $B_1 = B_2 = \cdots = B_n = B$ equation (28) becomes

$$\kappa^{(N)}(1_n, \gamma_n)[B, \dots, B] = k_n\big(b_{i(1)i(2)}^{(N)}, \dots, b_{i(n)i(1)}^{(N)}\big)$$

Thus to prove Theorem 2.6 we must show that $\kappa^{(N)}(1_n, \gamma_n)[B, \dots, B] \cdot N^{n-1}$ converges to κ_n^b the n^{th} free cumulant of the limiting eigenvalue distribution of $B^{(N)}$.
When $(\mathcal{V}, \pi) = (1_{m+n}, \gamma_{m,n})$ equation (28) becomes

$$\kappa^{(N)}(1_{m+n}, \gamma_{m,n})[B, \dots, B] = k_{m+n}\big(b_{i(1)i(2)}^{(N)}, \dots, b_{i(m)i(1)}^{(N)}, b_{j(1)j(2)}^{(N)}, \dots, b_{j(n),j(1)}^{(N)}\big)$$

Thus to prove Theorem 2.12 we must show that $\kappa^{(N)}(1_{m+n}, \gamma_{m,n})[B, \ldots, B] \cdot N^{m+n}$ converges to $\kappa^b_{m,n}$ the $(m,n)^{th}$ free cumulant of second order of the limiting second order distribution of $B^{(N)}$.

4.5. Length functions. We want to understand the asymptotic behavior of formula (29). The leading order in N of the right hand side is given by

$$-n + 2\#\mathcal{V} - \#\pi + 2\#\mathcal{W} - \#\sigma = n + (|\pi| - 2|\mathcal{V}|) + (|\sigma| - 2|\mathcal{W}|),$$

whereas the leading order of the left hand side is given by

$$2\#\mathcal{U} - \#\gamma = 2\#(\mathcal{V} \vee \mathcal{W}) - \#(\sigma\pi) = n + (|\pi\sigma| - 2|\mathcal{V} \vee \mathcal{W}|).$$

This suggests the introducing of the following "length functions" for permutations, partitions, and partitioned permutations.

Notation 4.6.

(1) For $\mathcal{V} \in \mathcal{P}(n)$ and $\pi \in S_n$ we put

$$|\mathcal{V}| := n - \#\mathcal{V}$$
$$|\pi| := n - \#\pi.$$

(2) For any $(\mathcal{V}, \pi) \in \mathcal{PS}(n)$ we put

$$|(\mathcal{V}, \pi)| := 2|\mathcal{V}| - |\pi| = n - (2\#\mathcal{V} - \#\pi).$$

Let us first observe that these quantities behave actually like a length. It is clear from the definition that they are always non-negative; that they also obey a triangle inequality is the content of the next lemma.

Lemma 4.7.

(1) *For all* $\pi, \sigma \in S_n$ *we have*

$$|\pi\sigma| \leq |\pi| + |\sigma|.$$

(2) *For all* $\mathcal{V}, \mathcal{W} \in \mathcal{P}(n)$ *we have*

$$|\mathcal{V} \vee \mathcal{W}| \leq |\mathcal{V}| + |\mathcal{W}|.$$

(3) *For all partitioned permutations* $(\mathcal{V}, \pi), (\mathcal{W}, \sigma) \in \mathcal{PS}(n)$ *we have*

$$|(\mathcal{V} \vee \mathcal{W}, \pi\sigma)| \leq |(\mathcal{V}, \pi)| + |(\mathcal{W}, \sigma)|.$$

Proof. (1) This is well-known, since $|\pi|$ is the minimal number of factors needed to write π as a product of transpositions.
(2) Each block B of $\mathcal{W}$ can glue at most $\#B - 1$ many blocks of $\mathcal{V}$ together, i.e., $\mathcal{W}$ can glue at most $n - \#\mathcal{W}$ many blocks of $\mathcal{V}$ together, thus the difference between $|\mathcal{V}|$ and $|\mathcal{V} \vee \mathcal{W}|$ cannot exceed $n - \#\mathcal{W}$ and hence

$$\#\mathcal{V} - \#(\mathcal{V} \vee \mathcal{W}) \leq n - \#\mathcal{W}.$$

This is equivalent to our assertion.
(3) We prove this, for fixed π and σ by induction over $|\mathcal{V}| + |\mathcal{W}|$. The smallest possible value of the latter appears for $|\mathcal{V}| = |\pi|$ and $|\mathcal{W}| = |\sigma|$ (i.e., $\mathcal{V} = 0_\pi$ and $\mathcal{W} = 0_\sigma$). But then we have (since $\mathcal{V} \vee \mathcal{W} \geq \pi\sigma$)

$$2|\mathcal{V} \vee \mathcal{W}| - |\pi\sigma| \leq |\mathcal{V} \vee \mathcal{W}| \leq |\mathcal{V}| + |\mathcal{W}|,$$

which is exactly our assertion for this case. For the induction step, on the other side, one only has to observe that if one increases $|\mathcal{V}|$ (or $|\mathcal{W}|$) by one then $|\mathcal{V} \vee \mathcal{W}|$ can also increase by at most 1. $\square$

Remark 4.8. 1) Note that the triangle inequality for partitioned permutations together with (29) implies the following. Given random matrices $A = (A_N)_{N\in\mathbb{N}}$ and $B = (B_N)_{N\in\mathbb{N}}$ which have limit distributions of all orders. If A and B are independent and at least one of them is unitarily invariant, then $C = (C_N)_{N\in\mathbb{N}}$ with $C_N := A_N B_N$ also has limit distributions of all orders.
2) Since we know that Gaussian and Wishart random matrices have limit distributions of all orders (see e.g. [MS06, Thm. 3.1 and Thm. 3.5]), and since they are unitarily invariant, it follows by induction from the previous part that any polynomial in independent Gaussian and Wishart matrices has limit distributions of all orders.

4.6. Multiplication of partitioned permutations. Suppose $\{B_1^{(N)}, \dots, B_n^{(N)}\}$ has limit distributions of all orders. Then the left hand side of equation (27) has order $N^{2\#(\mathcal{U})-\#(\gamma)}$ and the right hand side of equation (27) has order $N^{-n+2\#(\mathcal{V})-\#(\pi)+|\gamma\pi^{-1}|}$. Thus the only terms of the right hand side that have order $N^{2\#(\mathcal{U})-\#(\gamma)}$ are those for which

$$2\#(\mathcal{U}) - \#(\gamma) = -n + 2\#(\mathcal{V}) - \#(\pi) + |\gamma\pi^{-1}|$$

i.e. for which $|(\mathcal{U},\gamma)| = |(\mathcal{V},\pi)| + |\gamma\pi^{-1}|$. Hence

$$\begin{aligned}\varphi^{(N)}(\mathcal{U},\gamma)[B_1^{(N)},\dots,B_n^{(N)}]\\ &= \sum_{\substack{(\mathcal{V},\pi)\in\mathcal{PS}(n)\\ \mathcal{V}\vee\gamma\pi^{-1}=\mathcal{U}\\ |(\mathcal{U},\gamma)|=|(\mathcal{V},\pi)|+|\gamma\pi^{-1}|}} \kappa^{(N)}(\mathcal{V},\pi)[B_1^{(N)},\dots,B_n^{(N)}]\cdot N^{|\gamma\pi^{-1}|}\\ &\qquad\qquad + O(N^{2\#(\mathcal{U})-\#(\gamma)-2})\end{aligned}$$

Thus after taking limits we have

$$\varphi(\mathcal{U},\gamma)[b_1,\dots,b_n] = \sum_{(\mathcal{V},\pi)\in\mathcal{PS}(n)} \kappa(\mathcal{V},\pi)[b_1,\dots,b_n] \tag{31}$$

where the sum is over all $(\mathcal{V},\pi)$ in $\mathcal{PS}(n)$ such that $\mathcal{V}\vee\gamma\pi^{-1} = \mathcal{U}$ and $|(\mathcal{U},\gamma)| = |(\mathcal{V},\pi)| + |\gamma\pi^{-1}|$.
A similar analysis of equation (29) gives that for independent $\{A_1^{(N)},\dots,A_n^{(N)}\}$ and $\{B_1^{(N)},\dots,B_n^{(N)}\}$ with the $A_i^{(N)}$'s unitarily invariant and both having limit distributions of all orders we have

$$
\begin{aligned}
&\varphi^{(N)}(\mathcal{U},\gamma)[A_1^{(N)}B_1^{(N)},\dots,A_n^{(N)}B_n^{(N)}] \\
&\quad = \sum_{\substack{(\mathcal{V},\pi),(\mathcal{W},\sigma)\in\mathcal{PS}(n)\\ \mathcal{V}\vee\mathcal{W}=\mathcal{U},\ \pi\sigma=\gamma \\ |(\mathcal{V},\pi)|+|(\mathcal{W},\sigma)|=|(\mathcal{V}\vee\mathcal{W},\pi\sigma)|}} \kappa^{(N)}(\mathcal{V},\pi)[A_1^{(N)},\dots,A_n^{(N)}]\cdot\varphi^{(N)}(\mathcal{W},\sigma)[B_1^{(N)},\dots,B_n^{(N)}] \\
&\qquad\qquad + O(N^{2\#(\mathcal{U})-\#(\gamma)-2})
\end{aligned}
$$

and again after taking limits

$$
\begin{aligned}
&\varphi(\mathcal{U},\gamma)[a_1b_1,\dots,a_nb_n] \\
&\quad = \sum_{(\mathcal{V},\pi),(\mathcal{W},\sigma)\in\mathcal{PS}(n)} \kappa(\mathcal{V},\pi)[a_1,\dots,a_n]\cdot\varphi(\mathcal{W},\sigma)[b_1,\dots,b_n]
\end{aligned}
\tag{32}
$$

where the sum is over all $(\mathcal{V},\pi),(\mathcal{W},\sigma)\in\mathcal{PS}(n)$ such that

- ⋄ $\mathcal{V}\vee\mathcal{W}=\mathcal{U}$
- ⋄ $\pi\sigma=\gamma$
- ⋄ $|(\mathcal{V},\pi)|+|(\mathcal{W},\sigma)|=|(\mathcal{U},\gamma)|$

In order to write this in a more compact form it is convenient to define a multiplication for partitioned permutations (in $\mathbb{C}\mathcal{PS}(n)$) as follows.

Definition 4.9. For $(\mathcal{V},\pi),(\mathcal{W},\sigma)\in\mathcal{PS}(n)$ we define their product as follows.

$$
\begin{aligned}
&(\mathcal{V},\pi)\cdot(\mathcal{W},\sigma) := \\
&\quad = \begin{cases} (\mathcal{V}\vee\mathcal{W},\pi\sigma) & \text{if } |(\mathcal{V},\pi)|+|(\mathcal{W},\sigma)|=|(\mathcal{V}\vee\mathcal{W},\pi\sigma)|, \\ 0 & \text{otherwise.}\end{cases}
\end{aligned}
\tag{33}
$$

Proposition 4.10. *The multiplication defined in Definition 4.9 is associative.*

Proof. We have to check that

$$
\big((\mathcal{V},\pi)\cdot(\mathcal{W},\sigma)\big)\cdot(\mathcal{U},\tau)=(\mathcal{V},\pi)\cdot\big((\mathcal{W},\sigma)\cdot(\mathcal{U},\tau)\big). \tag{34}
$$

Since both sides are equal to $(\mathcal{V}\vee\mathcal{W}\vee\mathcal{U},\pi\sigma\tau)$ in case they do not vanish, we have to see that the conditions for non-vanishing are for both sides the same. The conditions for the left hand side are

$$|(\mathcal{V},\pi)|+|(\mathcal{W},\sigma)|=|(\mathcal{V}\vee\mathcal{W},\pi\sigma)|$$

and

$$|(\mathcal{V}\vee\mathcal{W},\pi\sigma)|+|(\mathcal{U},\tau)|=|(\mathcal{V}\vee\mathcal{W}\vee\mathcal{U},\pi\sigma\tau)|.$$

These imply

$$
\begin{aligned}
|(\mathcal{V},\pi)|+|(\mathcal{W},\sigma)|+|(\mathcal{U},\tau)| &= |(\mathcal{U}\vee\mathcal{W}\vee\mathcal{U},\pi\sigma\tau)| \\
&\le |(\mathcal{V},\pi)|+|(\mathcal{W}\vee\mathcal{U},\sigma\tau)|,
\end{aligned}
$$

However, the triangle inequality

$$|(\mathcal{W}\vee\mathcal{U},\sigma\tau)|\le|(\mathcal{W},\sigma)|+|(\mathcal{U},\tau)|$$

yields that we have actually equality in the above inequality, thus leading to

$$|(\mathcal{W},\sigma)| + |(\mathcal{U},\tau)| = |(\mathcal{W}\vee\mathcal{U},\sigma\tau)|$$

and

$$|(\mathcal{V},\pi)| + |(\mathcal{W}\vee\mathcal{U},\sigma\tau)| = |(\mathcal{V}\vee\mathcal{W}\vee\mathcal{U},\pi\sigma\tau)|.$$

These are exactly the two conditions for the vanishing of the right hand side of (34). The other direction goes analogously. $\square$

Now we can write formulas (31) and (32) in convolution form

$$\varphi(\mathcal{U},\gamma)[b_1,\dots,b_n] = \sum_{\substack{(\mathcal{V},\pi)\in\mathcal{PS}(n)\\ (\mathcal{V},\pi)\cdot(0,\gamma\pi^{-1})=(\mathcal{U},\gamma)}} \kappa(\mathcal{V},\pi)[b_1,\dots,b_n] \tag{35}$$

and

$$\begin{aligned}\varphi(\mathcal{U},\gamma)&[a_1b_1,\dots,a_nb_n]\\ &= \sum_{\substack{(\mathcal{V},\pi),(\mathcal{W},\sigma)\in\mathcal{PS}(n)\\ (\mathcal{V},\pi)\cdot(\mathcal{W},\sigma)=(\mathcal{U},\gamma)}} \kappa(\mathcal{V},\pi)[a_1,\dots,a_n]\cdot\varphi(\mathcal{W},\sigma)[b_1,\dots,b_n].\end{aligned} \tag{36}$$

Note that both $\varphi(\mathcal{V},\pi)$ and $\kappa(\mathcal{V},\pi)$ are multiplicative in the sense that they factor according to the decomposition of $\mathcal{V}$ into blocks.
The philosophy for our definition of higher order freeness will be that equation (35) is the analogue of the moment-cumulant formula and shall be used to define the quantities κ, which will thus take on the role of cumulants in our theory – whereas the φ are the moments (see Definition 7.4). We shall define higher order freeness by requiring the vanishing of mixed cumulants, see Definition 7.6. On the other hand, equation (36) would be another way of expressing the fact that the a's are free from the b's. Of course, we will have to prove that those two possibilities are actually equivalent (see Theorem 7.9).

5. Multiplicative functions on partitioned permutations and their convolution

5.1. Convolution of multiplicative functions. Formulas (35) and (36) above are a generalization of the formulas describing first order freeness in terms of cumulants and convolution of multiplicative functions on non-crossing partitions. Since the dependence on the random matrices is irrelevant for this structure we will free ourselves in this section from the random matrices and look on the combinatorial heart of the observed formulas. In Section 7, we will return to the more general situation involving multiplicative functions which depend also on random matrices or more generally elements from an algebra.

Definition 5.1.

(1) We denote by $\mathcal{PS}$ the set of partitioned permutations on an arbitrary number of elements, i.e.,

$$\mathcal{PS} = \bigcup_{n\in\mathbb{N}} \mathcal{PS}(n).$$

(2) For two functions
$$f, g : \mathcal{PS} \to \mathbb{C}$$
we define their convolution
$$f * g : \mathcal{PS} \to \mathbb{C}$$
by
$$(f * g)(\mathcal{U}, \gamma) := \sum_{\substack{(\mathcal{V},\pi),(\mathcal{W},\sigma)\in\mathcal{PS}(n)\\(\mathcal{V},\pi)\cdot(\mathcal{W},\sigma)=(\mathcal{U},\gamma)}} f(\mathcal{V}, \pi)\; g(\mathcal{W}, \sigma)$$
for any $(\mathcal{U}, \gamma) \in \mathcal{PS}(n)$.

Definition 5.2. A function $f : \mathcal{PS} \to \mathbb{C}$ is called *multiplicative* if $f(1_n, \pi)$ depends only on the conjugacy class of π and we have
$$f(\mathcal{V}, \pi) = \prod_{V \in \mathcal{V}} f(1_V, \pi|_V).$$
Our main interest will be in multiplicative functions. It is easy to see that the convolution of two multiplicative functions is again multiplicative. It is clear that a multiplicative function is determined by the values of $f(1_n, \pi)$ for all $n \in \mathbb{N}$ and all $\pi \in S_n$.
An important example of a multiplicative function is the δ-function presented below.

Notation 5.3. The δ-function on $\mathcal{PS}$ is the multiplicative function determined by
$$\delta(1_n, \pi) = \begin{cases} 1 & \text{if } n = 1, \\ 0 & \text{otherwise.} \end{cases}$$
Thus for $(\mathcal{U}, \pi) \in \mathcal{PS}(n)$
$$\delta(\mathcal{U}, \pi) = \begin{cases} 1 & \text{if } (\mathcal{U}, \pi) = \big(0_n, (1)(2)\dots(n)\big) \text{ for some } n, \\ 0 & \text{otherwise.} \end{cases}$$

Proposition 5.4. *The convolution of multiplicative functions on $\mathcal{PS}$ is commutative and δ is the unit element.*

Proof. It is clear that δ is the unit element. For commutativity, we note that for multiplicative functions we have
$$f(\mathcal{V}, \pi) = f(\mathcal{V}, \pi^{-1}),$$
and thus
$$(g * f)(\mathcal{U}, \gamma) = (g * f)(\mathcal{U}, \gamma^{-1}) = \sum_{\substack{(\mathcal{V},\pi),(\mathcal{W},\sigma)\in\mathcal{PS}(n)\\(\mathcal{V},\pi)\cdot(\mathcal{W},\sigma)=(\mathcal{U},\gamma^{-1})}} g(\mathcal{V}, \pi) f(\mathcal{W}, \sigma).$$
Since the condition $(\mathcal{V}, \pi) \cdot (\mathcal{W}, \sigma) = (\mathcal{U}, \gamma^{-1})$ is equivalent to the condition $(\mathcal{W}, \sigma^{-1}) \cdot (\mathcal{V}, \pi^{-1}) = (\mathcal{U}, \gamma)$ we can continue with
$$(g * f)(\mathcal{U}, \gamma) = \sum_{\substack{(\mathcal{V},\pi),(\mathcal{W},\sigma)\in\mathcal{PS}(n)\\(\mathcal{W},\sigma^{-1})\cdot(\mathcal{V},\pi^{-1})=(\mathcal{U},\gamma)}} f(\mathcal{W}, \sigma^{-1}) g(\mathcal{V}, \pi^{-1}) = (f * g)(\mathcal{U}, \gamma).$$

□

5.2. Factorizations. Let us now try to characterize the non-trivial factorizations $(\mathcal{U},\gamma) = (\mathcal{V},\pi)\cdot(\mathcal{W},\sigma)$ appearing in the definition of our convolution. Let us first observe some simple general inequalities.

Lemma 5.5.

(1) *For permutations* $\pi,\sigma\in S(n)$ *we have*

$$|\pi| + |\sigma| + |\pi\sigma| \geq 2|\pi\vee\sigma|.$$

(2) *For partitions* $\mathcal{V}_2\leq\mathcal{V}_1$ *and* $\mathcal{W}_2\leq\mathcal{W}_1$ *we have*

$$|\mathcal{W}_1| + |\mathcal{V}_1| + |\mathcal{V}_2\vee\mathcal{W}_2| \geq |\mathcal{V}_1\vee\mathcal{W}_1| + |\mathcal{W}_2| + |\mathcal{V}_2|$$

and

$$|\mathcal{V}_1\vee\mathcal{W}_2| + |\mathcal{V}_2\vee\mathcal{W}_1| \geq |\mathcal{V}_1\vee\mathcal{W}_1| + |\mathcal{V}_2\vee\mathcal{W}_2|.$$

Proof. (1) By the triangle inequality for partitioned permutations we have

$$|(0_\pi\vee 0_\sigma,\pi\sigma)| \leq |(0_\pi,\pi)| + |(0_\sigma,\sigma)|,$$

i.e.,

$$2|\pi\vee\sigma| - |\pi\sigma| \leq |\pi| + |\sigma|. \tag{37}$$

(2) Consider first the special case $\mathcal{W}_1=\mathcal{W}_2=\mathcal{W}$. Then we clearly have

$$\#(\mathcal{V}_2\vee\mathcal{W}) - \#(\mathcal{V}_1\vee\mathcal{W}) \leq \#\mathcal{V}_2 - \#\mathcal{V}_1,$$

which leads to

$$|\mathcal{V}_1\vee\mathcal{W}| - |\mathcal{V}_2\vee\mathcal{W}| \leq |\mathcal{V}_1| - |\mathcal{V}_2|.$$

From this the general case follows by

$$\begin{aligned}|\mathcal{V}_1\vee\mathcal{W}_1| - |\mathcal{V}_2\vee\mathcal{W}_2| &= |\mathcal{V}_1\vee\mathcal{W}_1| - |\mathcal{V}_1\vee\mathcal{W}_2| + |\mathcal{V}_1\vee\mathcal{W}_2| - |\mathcal{V}_2\vee\mathcal{W}_2|\\ &\leq |\mathcal{W}_1| - |\mathcal{W}_2| + |\mathcal{V}_1| - |\mathcal{V}_2|.\end{aligned}$$

The second inequality follows from this as follows:

$$\begin{aligned}|\mathcal{V}_1\vee\mathcal{W}_1| - |\mathcal{V}_1\vee\mathcal{W}_2| &= |\mathcal{V}_1\vee(\mathcal{V}_2\vee\mathcal{W}_1)| - |\mathcal{V}_1\vee(\mathcal{V}_2\vee\mathcal{W}_2)|\\ &\leq |\mathcal{V}_2\vee\mathcal{W}_1| - |\mathcal{V}_2\vee\mathcal{W}_2|.\end{aligned}$$

□

Theorem 5.6. *For* $(\mathcal{V},\pi),(\mathcal{W},\sigma)\in\mathcal{PS}(n)$ *the equation*

$$(\mathcal{V},\pi)\cdot(\mathcal{W},\sigma) = (\mathcal{V}\vee\mathcal{W},\pi\sigma)$$

is equivalent to the conjunction of the following four conditions:

$$\begin{aligned}|\pi| + |\sigma| + |\pi\sigma| &= 2|\pi\vee\sigma|,\\ |\mathcal{V}| + |\pi\vee\sigma| &= |\pi| + |\mathcal{V}\vee\sigma|,\\ |\mathcal{W}| + |\pi\vee\sigma| &= |\sigma| + |\pi\vee\mathcal{W}|,\\ |\mathcal{V}\vee\sigma| + |\pi\vee\mathcal{W}| &= |\mathcal{V}\vee\mathcal{W}| + |\pi\vee\sigma|.\end{aligned}$$

Proof. Adding the four inequalities given by Lemma 5.5

$$\begin{aligned}|\pi| + |\sigma| + |\pi\sigma| &\geq 2|\pi \vee \sigma|,\\ 2|\mathcal{V}| + 2|\pi \vee \sigma| &\geq 2|\pi| + 2|\mathcal{V} \vee \sigma|,\\ 2|\mathcal{W}| + 2|\pi \vee \sigma| &\geq 2|\sigma| + 2|\pi \vee \mathcal{W}|,\\ 2|\mathcal{V} \vee \sigma| + 2|\pi \vee \mathcal{W}| &\geq 2|\mathcal{V} \vee \mathcal{W}| + 2|\pi \vee \sigma|\end{aligned}$$

gives

$$2|\mathcal{V}| - |\pi| + 2|\mathcal{W}| - |\sigma| \geq 2|\mathcal{V} \vee \mathcal{W}| - |\pi\sigma|,$$

i.e.,

$$|(\mathcal{V}, \pi)| + |(\mathcal{W}, \sigma)| \geq |(\mathcal{V} \vee \mathcal{W}, \pi\sigma).$$

Since $(\mathcal{V}, \pi) \cdot (\mathcal{W}, \sigma) = (\mathcal{V} \vee \mathcal{W}, \pi\sigma)$ means that we require equality in the last inequality, this is equivalent to having equality in all the four inequalities. □

The conditions describing our factorizations have a quite geometrical meaning. Let us elaborate on this in the following.

DEFINITION 5.7. Let $\gamma \in S(n)$ be a fixed permutation.

(1) A permutation $\pi \in S(n)$ is called γ-*planar* if
$$|\pi| + |\pi^{-1}\gamma| + |\gamma| = 2|\pi \vee \gamma|.$$

(2) A partitioned permutation $(\mathcal{V}, \pi) \in \mathcal{PS}(n)$ is called γ-*minimal* if
$$|\mathcal{V} \vee \gamma| - |\pi \vee \gamma| = |\mathcal{V}| - |\pi|.$$

Remark 5.8. *i*) It is easy to check (for example, by calculating the Euler characteristic) that γ-planarity of π corresponds indeed to a planar diagram, i.e. one can draw a planar graph representing permutations γ and π without any crossings. The most important cases are when γ consists of a single cycle [Bia97] and when γ consists of two cycles [MN04].
ii) The notion of γ-minimality of $(\mathcal{V}, \pi)$ means that $\mathcal{V}$ connects only blocks of π which are not already connected by γ.
iii) If $(\mathcal{V}, \pi)$ satisfies both (1) and (2) of Definition 5.7 then $(\mathcal{V}, \pi)(0, \pi^{-1}\gamma) = (1, \gamma)$, by Theorem 5.6.

COROLLARY 5.9. *Assume that we have the equation*
$$(\mathcal{U}, \gamma) = (\mathcal{V}, \pi) \cdot (\mathcal{W}, \sigma).$$
Then π *and* σ *must be* γ*-planar and* $(\mathcal{V}, \pi)$ *and* $(\mathcal{W}, \sigma)$ *must be* γ*-minimal.*

5.3. FACTORIZATIONS OF DISC AND TUNNEL PERMUTATIONS.

NOTATION 5.10. *i*) We call $(\mathcal{V}, \pi) \in \mathcal{PS}_n$ a *disc permutation* if $\mathcal{V} = 0_\pi$; the latter is equivalent to the condition $|\mathcal{V}| = |\sigma|$. For $\pi \in S_n$, by $(0, \pi)$ we will always mean the disc permutation
$$(0, \pi) := (0_\pi, \pi) \in \mathcal{PS}(n).$$
ii) We call $(\mathcal{V}, \pi) \in \mathcal{PS}_n$ a *tunnel permutation* if $|\mathcal{V}| = |\pi| + 1$. This means that $\mathcal{V}$ is obtained from π by joining a pair of cycles; i.e. one block of $\mathcal{V}$ contains exactly two cycles of π and all other blocks contain only one cycle of π.

A motivation for those names comes from the identification between partitioned permutations and so-called surfaced permutations; see the Appendix for more information on this.
Our goal is now to understand more explicitly the factorizations of disc and tunnel permutations. (It will turn out that those are the relevant ones for first and second order freeness). For this, note that we can rewrite the crucial condition for our product of partitioned permutations,

$$2|\mathcal{V}| - |\pi| + 2|\mathcal{W}| - |\sigma| = 2|\mathcal{V} \vee \mathcal{W}| - |\pi\sigma|,$$

in the form

$$\big(|\mathcal{V}| - |\pi|\big) + \big(|\mathcal{W}| - |\sigma|\big) + \big(|\mathcal{V}| + |\mathcal{W}| - |\mathcal{V} \vee \mathcal{W}|\big) = \big(|\mathcal{V} \vee \mathcal{W}| - |\pi\sigma|\big).$$

Since all terms in brackets are non-negative integers this formula can be used to obtain explicit solutions to our factorization problem for small values of the right hand side. Essentially, this tells us that factorizations of a disc permutation can only be of the form disc $\times$ disc; and factorizations of a tunnel permutation can only be of the form disc$\times$disc, disc$\times$tunnel, and tunnel$\times$disc. Of course, one can generalize the following arguments to higher order type permutations, however, the number of possibilities grows quite quickly.

PROPOSITION 5.11.

(1) *The solutions to the equation*

$$(1_n, \gamma_n) = (0, \gamma_n) = (\mathcal{V}, \pi) \cdot (\mathcal{W}, \sigma)$$

are exactly of the form

$$(1_n, \gamma_n) = (0, \pi) \cdot (0, \pi^{-1}\gamma_n),$$

for some $\pi \in NC(n)$.

(2) *The solutions to the equation*

$$(1_{m+n}, \gamma_{m,n}) = (\mathcal{V}, \pi) \cdot (\mathcal{W}, \sigma)$$

are exactly of the following three forms:

(a)

$$(1_{m+n}, \gamma_{m,n}) = (0, \pi) \cdot (0, \pi^{-1}\gamma_{m,n}),$$

where $\pi \in S_{NC}(m, n)$*;*

(b)

$$(1_{m+n}, \gamma_{m,n}) = (0, \pi) \cdot (\mathcal{W}, \pi^{-1}\gamma_{m,n}),$$

where $\pi \in NC(m) \times NC(n)$*,* $|\mathcal{W}| = |\pi^{-1}\gamma_{m,n}| + 1$*, and* $\mathcal{W}$ *connects a cycle of* $\pi^{-1}\gamma_{m,n}$ *in* $NC(m)$ *with a cycle in* $NC(n)$*;*

(c)

$$(1_{m+n}, \gamma_{m,n}) = (\mathcal{V}, \pi) \cdot (0, \pi^{-1}\gamma_{m,n}),$$

where $\pi \in NC(m) \times NC(n)$*,* $|\mathcal{V}| = |\pi| + 1$*, and* $|\mathcal{W}| = |\pi^{-1}\gamma_{m,n}| + 1$*, and* $\mathcal{V}$ *connects a cycle of* π *in* $NC(m)$ *with a cycle in* $NC(n)$*.*

Proof. (1) The correspondence between non-crossing partitions and permutations was studied in detail by Biane [Bia97]. In this case we have

$$\big(|\mathcal{V}| - |\pi|\big) + \big(|\mathcal{W}| - |\sigma|\big) + \big(|\mathcal{V}| + |\mathcal{W}| - |\mathcal{V} \vee \mathcal{W}|\big) = |1_n| - |\gamma_n| = 0.$$

Since all three terms in brackets are greater or equal to zero, all of them must vanish, i.e.,

$$|\mathcal{V}| = |\pi|, \qquad \text{thus} \qquad \mathcal{V} = 0_\pi$$

$$|\mathcal{W}| = |\sigma|, \qquad \text{thus} \qquad \mathcal{W} = 0_\sigma$$

and

$$|\pi| + |\sigma| = |\mathcal{V}| + |\mathcal{W}| = |\mathcal{V} \vee \mathcal{W}| = |\gamma| = n - 1.$$

(2) Now we have

$$\big(|\mathcal{V}| - |\pi|\big) + \big(|\mathcal{W}| - |\sigma|\big) + \big(|\mathcal{V}| + |\mathcal{W}| - |\mathcal{V} \vee \mathcal{W}|\big) = \big(|\mathcal{V} \vee \mathcal{W}| - |\pi\sigma|\big) = 1,$$

which means that two of the terms on the left-hand side must be equal to 0, and the other term must be equal to 1. Thus we have the following three possibilities.

(a)

$$|\mathcal{V}| = |\pi|, \qquad \text{thus} \qquad \mathcal{V} = 0_\pi,$$
$$|\mathcal{W}| = |\sigma|, \qquad \text{thus} \qquad \mathcal{W} = 0_\sigma$$

and

$$|\pi| + |\sigma| = |\mathcal{V}| + |\mathcal{W}| = |\mathcal{V} \vee \mathcal{W}| + 1 = m + n.$$

Note that

$$\pi \vee \sigma = \mathcal{V} \vee \mathcal{W} = 1_{m+n},$$

and thus π connects the two cycles of $\gamma_{m,n}$. This means that π is a non-crossing (m,n)-permutation.

(b)

$$|\mathcal{V}| = |\pi|, \qquad \text{thus} \qquad \mathcal{V} = 0_\pi,$$

$$|\mathcal{W}| = |\sigma| + 1,$$

and

$$|\mathcal{V}| + |\mathcal{W}| = |\mathcal{V} \vee \mathcal{W}| = m + n - 1.$$

This implies

$$|\pi| + |\gamma_{m,n}\pi^{-1}| = m + n - 2,$$

which means that π must be a disconnected non-crossing (m,n)-annular permutation, i.e.,

$$\pi = \pi_1 \times \pi_2 \qquad \text{with} \qquad \pi_1 \in NC(m), \pi_2 \in NC(n).$$

(c)

$$|\mathcal{V}| = |\pi| + 1,$$

$$|\mathcal{W}| = |\sigma| + 1,, \qquad \text{thus} \qquad \mathcal{W} = 0_\sigma$$

and

$$|\mathcal{V}| + |\mathcal{W}| = |\mathcal{V} \vee \mathcal{W}| = m + n - 1.$$

This implies

$$|\pi| + |\gamma_{m,n}\pi^{-1}| = m + n - 2,$$

which means that π must be a disconnected non-crossing (m,n)-annular permutation, i.e.,

$$\pi = \pi_1 \times \pi_2 \qquad \text{with} \qquad \pi_1 \in NC(m), \pi_2 \in NC(n).$$

□

Example 5.12. We can now use the previous description of factorizations of disc and tunnel permutations to write down explicit first and second order formulas for our convolution of multiplicative functions.
1) In the first order case we have

$$(f * g)(1_n, \gamma_n) = (f * g)(0, \gamma_n) = \sum_{\pi \in NC(n)} f(0,\pi) g(0, \pi^{-1}\gamma_n). \tag{38}$$

This equation is exactly the formula for the convolution of multiplicative functions on non-crossing partitions, which is the cornerstone of the combinatorial description of first order freeness [NSp97]. (Note that $\pi^{-1}\gamma_n$ is in this case the Kreweras complement of π.)
2) In the second order case we have

$$\begin{aligned}(f * g)(1_{m+n}, \gamma_{m,n}) &= \sum_{\pi \in S_{NC}(m,n)} f(0,\pi) g(0, \pi^{-1}\gamma_{m,n}) \\ &+ \sum_{\substack{\pi \in NC(m) \times NC(n) \\ |\mathcal{V}| = |\pi| + 1}} \big(f(0, \gamma_{m,n}\pi^{-1}) g(\mathcal{V}, \pi) + f(\mathcal{V}, \pi) g(0, \pi^{-1}\gamma_{m,n}) \big).\end{aligned}$$

We should expect that this formula is the combinatorial key for the understanding of second order freeness. However, in this form it does not match exactly the formulas appearing in [MŚS07]. Let us, however, for a multiplicative function f put, for $\pi \in NC(n)$,

$$\tilde{f}_1(\pi) := f(1_n, \pi) \qquad (\pi \in NC(n)) \tag{39}$$

and, for $\pi_1 \in NC(m)$ and $\pi_2 \in NC(n)$,

$$\tilde{f}_2(\pi_1, \pi_2) = \sum_{\substack{\mathcal{V} \geq \pi_1 \times \pi_2, \\ |\mathcal{V}| = |\pi| + 1, \\ \mathcal{V} \vee (\pi_1 \times \pi_2) = 1_{m+n}}} f(\mathcal{V}, \pi_1 \times \pi_2). \tag{40}$$

Note that in the definition of $\tilde{f}_2$ the sum is running over all $\mathcal{V}$ which connect exactly one cycle of π_1 with one cycle of π_2.

Then, with $h = f * g$, we have

$$\begin{aligned}\tilde{h}_2(1_m, 1_n) = &\sum_{\pi \in S_{NC}(m,n)} \tilde{f}_1(\pi)\tilde{g}_1(\pi^{-1}\gamma_{m,n}) \\ &+ \sum_{\pi_1,\pi_2 \in NC(m)\times NC(n)} \big(\tilde{f}_2(\pi_1,\pi_2)\tilde{g}_1(\pi_1^{-1} \times \pi_2^{-1}\gamma_{m.n}) \\ &\qquad + \tilde{f}_1(\pi_1 \times \pi_2)\tilde{g}_2(\pi_1^{-1}\gamma_m, \pi_2^{-1}\gamma_n)\big).\end{aligned}$$

In this form we recover exactly the structure of the formula (10) from [MŚS07], which describes second order freeness. The descriptions in terms of f and in terms of $\tilde{f}_2$ are equivalent. Whereas f is multiplicative, $\tilde{f}_2$ satisfies a kind of cocycle property. From our present perspective the description of second (and higher) order freeness in terms of multiplicative functions seems more natural. In any case, we see that our convolution of multiplicative functions on partitioned permutations is a generalization of the structure underlying first and second order freeness.

5.4. ZETA AND MÖBIUS FUNCTION. In the definition of our convolution we are running over factorizations of $(\mathcal{U}, \gamma)$ into products $(\mathcal{V}, \pi) \cdot (\mathcal{W}, \sigma)$. In the first order case the second factor is determined if the first factor is given. In the general case, however, we do not have such a uniqueness of the decomposition; if we fix $(\mathcal{V}, \pi)$ there might be different choices for $(\mathcal{W}, \sigma)$. For example, this situation was considered in Proposition 5.11 in the case (2b). However, in the case when $(\mathcal{W}, \sigma)$ is a disc permutation, it must be of the form $(0_{\pi^{-1}\gamma}, \pi^{-1}\gamma)$ and is thus uniquely determined. Note that factorizations of such a special form appear in our formula (35) and thus deserve special attention.

NOTATION 5.13. Let $(\mathcal{U}, \gamma) \in \mathcal{PS}$ be a fixed partitioned permutation. We say that $(\mathcal{V}, \pi) \in \mathcal{PS}$ is $(\mathcal{U}, \gamma)$*–non-crossing* if

$$(\mathcal{V}, \pi) \cdot (0_{\pi^{-1}\gamma}, \pi^{-1}\gamma) = (\mathcal{U}, \gamma).$$

The set of $(\mathcal{U}, \gamma)$–non-crossing partitioned permutations will be denoted by $\mathcal{PS}_{NC}(\mathcal{U}, \gamma)$, see Remark 5.8.

To justify this notation we point out that $(1_n, \gamma_n)$–non-crossing partitioned permutations can be identified with non-crossing permutations; to be precise

$$\mathcal{PS}_{NC}(1_n, \gamma_n) = \{(0_\pi, \pi) \mid \pi \in NC(n)\}.$$

Furthermore,

$$\begin{aligned}\mathcal{PS}_{NC}(1_{m+n}, \gamma_{m,n}) = &\{(0_\pi, \pi) \mid \pi \in S_{NC}(m,n)\} \cup \\ &\cup \{(\mathcal{V}, \pi_1 \times \pi_2) \mid \pi_1 \in NC(m), \pi_2 \in NC(n), \mathcal{V} \geq \pi, |\mathcal{V}| = |\pi| + 1 \\ &\qquad \text{and } \mathcal{V} \text{ connects one cycle of } \pi_1 \text{ to a cycle of } \pi_2\}.\end{aligned}$$

We can now also use a special multiplicative function, which we will call Zeta-function ζ, to single out such factorizations. It will be useful to be able to invert formula (35), which means we need also the inverse of ζ under our convolution.

This inverse, called the Möbius-function μ, is a key object in the theory and contains a lot of important information.

NOTATION 5.14.

(1) The *Zeta-function* ζ is the multiplicative function on $\mathcal{PS}$ which is determined by

$$\zeta(1_n, \pi) = \begin{cases} 1 & \text{if } (1_n, \pi) \text{ is a disc permutation, i.e., if } 1_n = 0_\pi, \\ 0 & \text{otherwise.} \end{cases}$$

(2) The *Möbius function* μ is the inverse of ζ under convolution, i.e., it is determined by

$$\zeta * \mu = \delta = \mu * \zeta.$$

Note that in general

$$\zeta(\mathcal{V}, \pi) = \begin{cases} 1, & \text{if } \mathcal{V} = 0_\pi \\ 0, & \text{if } \mathcal{V} > 0_\pi. \end{cases}$$

It is also quite easy to see that the Möbius function exists and is uniquely determined as the inverse of the Zeta-function — the determining equations can be solved recursively. Indeed letting $\mu_n = \mu(1_n, \gamma_n)$ and $\mu_{m,n} = \mu(1_{m+n}, \gamma_{m,n})$ we have

$$\begin{aligned}
0 &= \mu_{1,1} + \mu_2 \\
0 &= \mu_{1,2} + 2\mu_1\mu_{1,1} + 2\mu_3 + 2\mu_1\mu_2 \\
0 &= \mu_{2,2} + 4\mu_1\mu_{2,1} + 4\mu_1^2\mu_{1,1} + 4\mu_4 + 8\mu_1\mu_3 + 2\mu_2^2 + 4\mu_1^2\mu_2 \\
0 &= \mu_{1,3} + 3\mu_1\mu_{2,1} + 3\mu_2\mu_{1,1} + 3\mu_4 + 6\mu_1\mu_3 + 3\mu_2^2 + 3\mu_1^2\mu_2 \\
0 &= \mu_{2,3} + 2\mu_1\mu_{1,3} + 3\mu_1\mu_{2,2} + 3\mu_2\mu_{1,2} + 9\mu_1^2\mu_{1,2} + 6\mu_1\mu_2\mu_{1,1} + 6\mu_1^3\mu_{1,1} \\
&\quad + 6\mu_5 + 18\mu_1\mu_4 + 12\mu_2\mu_3 + 18\mu_1^2\mu_3 + 12\mu_1\mu_2^2 + 6\mu_1^3\mu_2 \\
0 &= \mu_{3,3} + 6\mu_1\mu_{2,3} + 6\mu_2\mu_{1,3} + 6\mu_1^2\mu_{1,3} + 9\mu_1^2\mu_{2,2} + 18\mu_1\mu_2\mu_{1,2} + 18\mu_1^3\mu_{1,2} \\
&\quad + 9\mu_2^2\mu_{1,1} + 18\mu_1^2\mu_2\mu_{1,1} + 9\mu_1^4\mu_{1,1} + 9\mu_6 + 36\mu_1\mu_5 + 27\mu_2\mu_4 + 54\mu_1^2\mu_4 \\
&\quad + 9\mu_3^2 + 72\mu_1\mu_2\mu_3 + 36\mu_1^3\mu_3 + 12\mu_2^3 + 36\mu_1^2\mu_2^2 + 9\mu_1^4\mu_2
\end{aligned}$$

This shows how, knowing the first order Möbius function μ_n, the second order Möbius function $\mu_{m,n}$ can be calculated recursively.

One should observe that with these notations we have

$$(f * \zeta)(\mathcal{U}, \gamma) = \sum_{(\mathcal{V},\pi) \in \mathcal{PS}_{NC}(\mathcal{U},\gamma)} f(\mathcal{V}, \pi).$$

In the following we will use the notation

$$\zeta^{*p} = \underbrace{\zeta * \cdots * \zeta}_{p\text{-times}}.$$

It is clear, by definition, that ζ^{*p} counts factorizations into the product of p disc permutations, thus we have the following result.

PROPOSITION 5.15. *For $(\mathcal{U}, \gamma) \in \mathcal{PS}$ and $p \geq 1$ we have*

$$\zeta^{*p}(\mathcal{U}, \gamma) := \#\{(\pi_1, \dots, \pi_p) \mid (\mathcal{U}, \gamma) = (0, \pi_1) \cdots (0, \pi_p)\}.$$

Of special interest for us is the case $p = 2$.

PROPOSITION 5.16. *We have for all $r \geq 1$ and $n(1), \dots, n(r) \in \mathbb{N}$, $n := n(1) + \cdots + n(r)$ that*

$$(\zeta * \zeta)(1_n, \gamma_{n(1),\dots,n(r)}) = \# S_{NC}(n(1), \dots, n(r)).$$

Proof. As noted above, $(\zeta * \zeta)(1_n, \gamma_{n(1),\cdots,n(r)})$ counts the number of factorizations of $(1_n, \gamma_{n(1),\dots,n(r)})$ into a product of two disc permutations, i.e., the number of factorizations of the form

$$(1_n, \gamma_{n(1),\dots,n(r)}) = (0, \pi) \cdot (0, \pi^{-1}\gamma_{n(1),\dots,n(r)}),$$

with

$$|\pi| + |\pi^{-1}\gamma| = |\gamma| = n - r$$

and $\pi \vee \gamma = 1_n$. But this describes exactly connected $(n(1), \dots, n(r))$-annular permutations $\pi \in S_{NC}(n(1), \dots, n(r))$. □

NOTATION 5.17. We put

$$c_{n(1),\dots,n(r)} := \# S_{NC}(n(1), \dots, n(r)).$$

Note in particular that c_n counts the number of non-crossing partitions of n elements and thus is the Catalan number

$$c_n = \frac{1}{n+1}\binom{2n}{n},$$

and that $c_{m,n}$ counts the number of non-crossing (m, n)-annular permutations, and thus [MN04]

$$c_{m,n} = \frac{2mn}{m+n}\binom{2m-1}{m}\binom{2n-1}{n}.$$

More generally, an explicit formula for the number of factorizations into p factors was derived by Bousquet-Mélou and Schaeffer [BMS00], namely one has (with $n := n(1) + \cdots + n(r)$)

$$\zeta^{*p}(1_n, \gamma_{n(1),\cdots,n(r)}) = p\frac{[(p-1)n-1]!}{[(p-1)n-r+2]!}\prod_{i=1}^{r}\left[n(i)\binom{pn(i)-1}{n(i)}\right],$$

and thus in particular

$$c_{n(1),\dots,n(r)} = 2\frac{(n-1)!}{(n-r+2)!}\prod_{i=1}^{r}\left[n(i)\binom{2n(i)-1}{n(i)}\right].$$

For our purposes, however, the following recursive formula for the number of factorizations is more interesting.
In the next theorem we will show how to reduce the problem of counting the number of disc factorizations on $[n]$ to counting the factorizations on $[n-1]$. This will enable of to obtain a recursive formula for $c_{n_1,\dots,n_r}$.

Notation 5.18. Let $(\mathcal{U}, \gamma)$ be a partitioned permutation of $[n]$ with $\gamma(1) \neq 1$. Let $\hat{\gamma}_k$ be the restriction of $(1,k)\gamma(1,\gamma^{-1}(k))$ to the invariant subset $[2,n] := \{2,3,4,\ldots,n\}$. Then

$$|\hat{\gamma}_k| = \begin{cases} |\gamma| & \text{if } 1 \text{ and } k \text{ are in different cycles of } \gamma, \\ |\gamma|-1 & \text{if } k=1 \text{ or } \gamma(1) \\ |\gamma|-2 & \text{if } 1 \text{ and } k \text{ are in the same cycle of } \gamma, \\ & \text{but } k \neq 1 \text{ and } k \neq \gamma(1), \end{cases}$$

Let $\overline{\mathcal{U}} = \mathcal{U}|_{[2,n]}$ be the restriction of $\mathcal{U}$ to $[2,n]$, *i.e.* if the blocks of $\mathcal{U}$ are $U_1, \ldots, U_r$ and $1 \in U_1$, then the blocks of $\overline{\mathcal{U}}$ are $\overline{U}_1, U_2, \ldots, U_r$ where $\overline{U}_1 = U_1 \cap [2,n]$. In the theorem below we sum over a set of partitions $\mathcal{P}_k$ of $[2,n]$ described as follows.

For $k=1$, $\gamma(1)$ or k not in the γ-orbit of 1, $\mathcal{P}_k = \{\,\overline{\mathcal{U}}\,\}$ *i.e.* $\mathcal{P}_k$ consists of the single partition $\overline{\mathcal{U}}$.

For k in the γ-orbit of 1 but $k \neq 1$, $\gamma(1)$, $\mathcal{P}_k = \{\widehat{\mathcal{U}} \mid \hat{\gamma}_k \leq \widehat{\mathcal{U}}, |\widehat{\mathcal{U}}| = |\mathcal{U}| - 2$, and $\overline{\mathcal{U}} = \widehat{\mathcal{U}} \vee (k, \gamma^{-1}(k))\}$. In words this means $\overline{U}_1$ is split into two blocks:

- the first containing the cycle of $\hat{\gamma}_k$ containing $\gamma^{-1}(k)$ and some (possibly none) of the other cycles of γ contained in U_1
- the second containing the cycle of $\hat{\gamma}_k$ containing k and the remaining (possibly none) cycles of γ contained in U_1.

More explicitly, in the case k is in the γ-orbit of 1 but $k \neq 1$, $\gamma(1)$, let us write γ as as a product of cycles $d_1 \cdots d_s$ where $d_1 = (1, \gamma(1), \ldots, \gamma^t(1))$ is the cycle that contains 1. Let $d_1' = (\gamma(1), \gamma^2(1), \ldots, \gamma^{-1}(k))$ and $d_1'' = (k, \ldots, \gamma^t(1))$. Then $\hat{\gamma}_k = d_1' d_1'' d_2 \cdots d_s$. $\mathcal{P}_k$ consists of all partitions $\widehat{\mathcal{U}}$ of $[2,n]$ such that $\widehat{\mathcal{U}} = \{U_1', U_1'', U_2, \ldots, U_r\}$ where $U_1' \cup U_1'' = \overline{U}_1$, $U_1' \cap U_1'' = \emptyset$, U_1' contains d_1', U_1'' contains d_1'', and each cycle of γ that was in U_1 is now in either U_1' or U_1'', *i.e.* $\hat{\gamma}_k \leq \widehat{\mathcal{U}}$ and $|\widehat{\mathcal{U}}| = |\mathcal{U}| - 2$.

Theorem 5.19.

$$\zeta^{*^2}(U, \gamma) = \sum_{k=1}^{n} \sum_{\widehat{\mathcal{U}} \in \mathcal{P}_k} \zeta^{*^2}(\widehat{\mathcal{U}}, \widehat{\gamma_k}) \tag{41}$$

Proof. We must show that for each factorization $(0,\pi)\cdot(0,\sigma)$ of $(\mathcal{U},\gamma)$ there are $k := \pi(1)$, $\widehat{\mathcal{U}} \in \mathcal{P}_k$, and permutations of $[2,n]$, $\hat{\pi}$ and $\hat{\sigma}$ such that $(0,\hat{\pi})\cdot(0,\hat{\sigma}) = (\widehat{\mathcal{U}}, \hat{\gamma}_k)$. Conversely we must show that given k, $\widehat{\mathcal{U}} \in \mathcal{P}_k$ and a factorization $(0,\hat{\pi}) \cdot (0,\hat{\sigma})$ of $(\widehat{\mathcal{U}}, \hat{\gamma}_k)$ there are π and σ such that $(0,\pi)\cdot(0,\sigma) = (\mathcal{U},\gamma)$ and $\pi(1) = k$. Moreover we must show that these two maps are inverses of each other. The relation between π, σ and $\hat{\pi}$, $\hat{\sigma}$ is given by $\hat{\pi} = (1,k)\pi|_{[2,n]}$, $\hat{\sigma} = \sigma(1,\gamma^{-1}(k))|_{[2,n]}$. So on the level of permutations we have a bijection. The main work of the proof is to show that starting with π and σ we have $\widehat{\mathcal{U}} := \hat{\pi} \vee \hat{\sigma} \in \mathcal{P}_k$ and $2|\widehat{\mathcal{U}}| - |\hat{\gamma}_k| = |\hat{\pi}| + |\hat{\sigma}|$; and then conversely starting with $\widehat{\mathcal{U}} \in \mathcal{P}_k$ and a factorization $(0,\hat{\pi}) \cdot (0,\hat{\sigma})$ of $(\widehat{\mathcal{U}}, \hat{\gamma}_k)$ then $2|\mathcal{U}| - |\gamma| = |\pi| + |\sigma|$ and $\pi \vee \sigma = \mathcal{U}$.

Note that we have for all k

$$|\hat{\pi}| = \begin{cases} |\pi| - 1 & k \neq 1 \\ |\pi| & k = 1 \end{cases}$$

$$|\hat{\sigma}| = \begin{cases} |\sigma| - 1 & k \neq \gamma(1) \\ |\sigma| & k = \gamma(1) \end{cases}$$

It is necessary to break the proof into four cases: k is not in the γ-orbit of 1; k is in the γ-orbit of 1 but $k \neq 1, \gamma(1)$; $k = 1$; and $k = \gamma(1)$.
Suppose we have a factorization

$$(\mathcal{U}, \gamma) = (0, \pi) \cdot (0, \sigma),$$

i.e., $\gamma = \pi\sigma$, $\mathcal{U} = \pi \vee \sigma$, and

$$2|\mathcal{U}| - |\gamma| = |\pi| + |\sigma|$$

with $k := \pi(1)$ not in the γ-orbit of 1. Then$|\hat{\gamma}_k| = |\gamma|$ and $\mathcal{P}_k$ contains only the partition of $[2, n]$ which results from $\mathcal{U}$ by removing 1, *i.e.* $\widehat{\mathcal{U}} = \overline{\mathcal{U}}$. Then we have $|\hat{\mathcal{U}}| = |\mathcal{U}| - 1$. Hence $|\hat{\pi}| + |\hat{\sigma}| = |\pi| + |\sigma| - 2 = 2|\mathcal{U}| - |\gamma| - 2 = |\widehat{\mathcal{U}}| - |\gamma| = |\widehat{\mathcal{U}}| - |\hat{\gamma}_k|$.
Also $0_\pi|_{[2,n]} = 0_{\hat{\pi}}$ and $0_\gamma|_{[2,n]} \leq 0_{\hat{\gamma}_k}$. Thus $\widehat{\mathcal{U}} = (\pi \vee \gamma)|_{[2,n]} \leq \hat{\pi} \vee \hat{\gamma}_k$. On the other hand the difference between $0_\gamma|_{[2,n]}$ and $0_{\hat{\gamma}_k}$ is that the blocks containing 1 and k have been joined. However these points were already connected by π. Thus $\hat{\pi} \vee \hat{\gamma}_k \leq \widehat{\mathcal{U}}$, and so $\hat{\mathcal{U}} = \hat{\pi} \vee \hat{\sigma}$, and thus

$$(\hat{\mathcal{U}}, \hat{\gamma}) = (0, \hat{\pi}) \cdot (0, \hat{\sigma}).$$

Conversely, given a factorization $(0, \hat{\pi}) \cdot (0, \hat{\sigma})$ of $(\widehat{\mathcal{U}}, \hat{\gamma}_k)$, let $\pi = (1, k)\hat{\pi}$ and $\sigma = \hat{\sigma}(1, \gamma^{-1}(k))$. Then $\pi \vee \sigma = \mathcal{U}$ because 1 has been connected to the block of $\widehat{\mathcal{U}}$ containing k. Also $\#(\pi) = \#(\hat{\pi})$ and $\#(\sigma) = \#(\hat{\sigma})$; thus $|\pi| = |\hat{\pi}| - 1$ and $|\sigma| = |\hat{\sigma}| - 1$, and so $|\pi| + |\sigma| = 2|\mathcal{U}| - |\gamma|$. This establishes the bijection when k is not in the γ-orbit of 1.
Let us now consider the case that 1 and k are in the same cycle of γ, but $k \neq 1, \gamma(1)$. Again suppose that $(0, \pi) \cdot (0, \sigma)$ is a factorization of $(\mathcal{U}, \gamma)$ with $\pi(1) = k$. In this case we have that $|\hat{\gamma}_k| = |\gamma| - 2$ and so by the triangle inequality, Lemma 4.7

$$\begin{aligned} 2|\hat{\pi} \vee \hat{\sigma}| - |\gamma| + 2 &= 2|\hat{\pi} \vee \hat{\sigma}| - |\hat{\gamma}| \\ &= |(\hat{\pi} \vee \hat{\sigma}, \hat{\pi}\hat{\sigma})| \\ &\leq |(0, \hat{\pi})| + |(0, \hat{\sigma})| \\ &= |\hat{\pi}| + |\hat{\sigma}| \\ &= |\pi| + |\sigma| - 2 \\ &= 2|\mathcal{U}| - |\gamma| - 2, \end{aligned}$$

and thus

$$|\hat{\pi} \vee \hat{\sigma}| \leq |\mathcal{U}| - 2.$$

On the other hand, let us compare

$$\hat{\pi} \vee \hat{\sigma} = \hat{\pi} \vee \hat{\gamma} \qquad \text{with} \qquad \mathcal{U} = \pi \vee \gamma.$$

Note that all our changes of the permutations affected only what happens on the first cycle of γ. Since the transition from γ to $\hat{\gamma}$ consists in removing the point 1 and splitting the first cycle of γ into two cycles, we can lose at most one block by going over from $\hat{\pi} \vee \hat{\gamma}$ to $\pi \vee \gamma$. Thus

$$|\hat{\pi} \vee \hat{\sigma}| = (n-1) - \#(\hat{\pi} \vee \hat{\sigma}) \geq (n-1) - (\#\mathcal{U} + 1) = |\mathcal{U}| - 2,$$

so that we necessarily have the equality

$$|\hat{\pi} \vee \hat{\sigma}| = |\mathcal{U}| - 2.$$

Thus $\widehat{\mathcal{U}} := \hat{\pi} \vee \hat{\sigma} \in \mathcal{P}_k$ and $2|\hat{\pi} \vee \hat{\sigma}| - |\hat{\gamma}_k| = |\hat{\pi}| + |\hat{\sigma}|$. Hence $(0, \hat{\pi}) \cdot (0, \hat{\sigma})$ is a factorization of $(\widehat{\mathcal{U}}, \hat{\gamma}_k)$.
Conversely let us suppose that k is in the γ-orbit of 1 but $k \neq 1$ or $\gamma(1)$ and $\widehat{\mathcal{U}} \in \mathcal{P}_k$ and $(0, \hat{\pi}) \cdot (0, \hat{\sigma})$ is a factorization of $(\widehat{\mathcal{U}}, \hat{\gamma}_k)$. We must show that $\pi \vee \sigma = \mathcal{U}$ and that $|\pi| + |\sigma| = 2|\mathcal{U}| - |\gamma|$. 1 and k are in the same orbit of π and 1 and $\gamma^{-1}(k)$ are in the same orbit of σ. So the blocks of $\widehat{\mathcal{U}}$ containing d_1' and d_1'' are joined in $\pi \vee \sigma$. Thus $\pi \vee \sigma = \mathcal{U}$. Also $|\widehat{\mathcal{U}}| = |\mathcal{U}| - 2$, so $|\pi| + |\sigma| = |\hat{\pi}| + |\hat{\sigma}| + 2 = 2|\widehat{\mathcal{U}}| - |\hat{\gamma}_k| + 2 = |\mathcal{U}| - |\hat{\gamma}_k| - 2 = 2|\mathcal{U}| - |\gamma|$. Thus $(0, \pi) \cdot (0, \sigma)$ is a factorization of $(\mathcal{U}, \gamma)$. This establishes the bijection in the case k is in the γ-orbit of 1 but $k \neq 1$ or $\gamma(1)$.
Next suppose that $k = 1$ and $(0, \pi) \cdot (0, \sigma)$ is a factorization of $(\mathcal{U}, \gamma)$ with $\pi(1) = 1$. Then $|\hat{\pi}| + |\hat{\sigma}| = |\pi| + |\sigma| - 1 = 2|\mathcal{U}| - |\gamma| - 1 = 2|\widehat{\mathcal{U}}| - |\gamma| + 1 = 2|\widehat{\mathcal{U}}| - |\hat{\gamma}_k|$. Let U_1 be the block of $\mathcal{U}$ containing 1 and $\overline{U}_1 = U_1 \cap [2, n]$. We must show that $\overline{U}_1$ is a block of $\hat{\pi} \vee \hat{\gamma}_k$. Since $\pi \vee \gamma = \mathcal{U}$ we know that if d_i and d_j are cycles of γ contained in U_1 then π must connect them. Since $\pi|_{\overline{U}_1} = \hat{\pi}|_{\overline{U}_1}$ we see that $\hat{\pi}$ connects the corresponding cycles of $\hat{\gamma}_k$ (which are unchanged except for the cycle containing 1). Similarly if f_1 and f_2 are cycles of π contained in U_1 and neither is a singleton then they are connected by γ and thus by $\hat{\gamma}_k$. Thus $(0, \hat{\pi}) \cdot (0, \hat{\sigma})$ is a factorization of $(\widehat{\mathcal{U}}, \hat{\gamma}_k)$.
Conversely suppose that $k = 1$, $\widehat{\mathcal{U}} \in \mathcal{P}_k$, and $(0, \hat{\pi}) \cdot (0, \hat{\sigma})$ is a factorization of $(\widehat{\mathcal{U}}, \hat{\gamma}_k)$. We must show that $\pi(1) = 1$ and $(0, \pi) \cdot (0, \sigma)$ is a factorization of $(\mathcal{U}, \gamma)$. Since $\hat{\pi} \vee \hat{\gamma}_k = \widehat{\mathcal{U}}$ and γ connects 1 to $\gamma(1) \in \overline{U}_1$, we have that $\pi \vee \gamma = \mathcal{U}$. Also $|\pi| + |\sigma| = |\hat{\pi}| + |\hat{\sigma}| + 1 = 2|\widehat{\mathcal{U}}| - |\hat{\gamma}_k| + 1 = 2|\mathcal{U}| - |\hat{\gamma}_1| - 1 = 2|\mathcal{U}| - |\gamma|$. Thus $(0, \pi) \cdot (0, \sigma)$ is a factorization of $(\mathcal{U}, \gamma)$. This completes the case when $k = 1$. The proof in the case $k = \gamma(1)$ is exactly the same except that the roles of π and σ are reversed. $\square$

Let us take a closer look at the meaning of Theorem 5.19 for the case $(\mathcal{U}, \gamma) = (1_n, \gamma_{n(1),\ldots,n(r)})$. To reduce the depth of subscripts we shall write $c(n_1, \ldots, n_r)$ for $c_{n(1),\ldots,n(r)}$.

Proposition 5.20. *We have for all $r, n_1, \dots, n_r \in \mathbb{N}$ the recursion*

$$(42)\quad c(n_1, \dots, n_r) = \sum_{l=2}^{r} n_l \cdot c(n_1 + n_l - 1, n_2, \dots, n_{l-1}, n_{l+1}, \dots, n_r)$$
$$+ \sum_{k=1}^{n_1} \sum_{\substack{A=\{i_1,\dots,i_s\} \\ B=\{j_1,\dots,j_t\}}} c(k-1, n_{i_1}, \dots, n_{i_s}) c(n_1 - k, n_{j_1}, \dots, n_{j_t})$$

where the sum is over all pairs of subsets $A, B \subset [2, r]$ such that $A \cap B = \emptyset$ and $A \cup B = [2, r]$ including the possibility that either A or B could be empty. We have for all $m, n \geq 1$

$$c_n = \sum_{1 \leq k \leq n} c_{k-1} c_{n-k},$$

and

$$(43)\qquad c_{m,n} = \sum_{1 \leq k \leq n} \big(c_{k-1} c_{m,n-k} + c_{m,k-1} c_{n-k}\big) + m c_{m+n-1},$$

where we use the convention that $c_0 = 1$ but $c(n_1, \dots, n_r) = 0$ if $r > 1$ and for some i, $n_i = 0$.

Proof. Let $n = n_1 + \cdots + n_r$. By Proposition 5.16 $c(n_1, \dots, n_r) = \zeta^{*^2}(1_n, \gamma_{n_1,\dots,n_r})$. So we must give the correspondence between the terms on the right hand side of (41) and the right hand side of (42). In this case $\mathcal{U} = 1_n$ and $\overline{\mathcal{U}} = 1_{n-1}$ (in the notation of 5.18). Thus $\mathcal{P}_k = \{1_{n-1}\}$. Also for $n_1 + \cdots + n_{l-1} < k \leq n_1 + \cdots + n_l$, $\zeta^{*^2}(1_{n-1}, \hat{\gamma}_k) = c(n_1 + n_l - 1, n_2, \dots, n_{l-1}, n_{l+1}, \dots, n_r)$. Thus

$$(44)\quad \sum_{k=n_1+1}^{n} \sum_{\widehat{\mathcal{U}} \in \mathcal{P}_k} \zeta^{*^2}(\widehat{\mathcal{U}}, \hat{\gamma}_k) = \sum_{k=n_1+1}^{n} \zeta^{*^2}(1_{n-1}, \hat{\gamma}_k)$$
$$= \sum_{l=2}^{r} \sum_{k=n_1+\cdots+n_{l-1}+1}^{n_1+\cdots+n_l} c(n_1 + n_l - 1, n_2, \dots, n_{l-1}, n_{l+1}, \dots, n_r)$$
$$= \sum_{l=2}^{r} n_l \cdot c(n_1 + n_l - 1, n_2, \dots, n_{l-1}, n_{l+1}, \dots, n_r)$$

For $k \leq n_1$, $\hat{\gamma}_k = d_1' d_1'' d_2 \cdots d_r$, with d_1' a cycle of length $k-1$ and d_1'' a cycle of length $n_1 - k$. $\mathcal{P}_k$ is the set of all partitions of the cycles of $\hat{\gamma}_k$ into two blocks such that d_1' and d_1'' are in different blocks. Hence

$$\sum_{\widehat{\mathcal{U}} \in \mathcal{P}_k} \zeta^{*^2}(\widehat{\mathcal{U}}, \hat{\gamma}_k) = \sum_{\substack{A=\{i_1,\dots,i_s\} \\ B=\{j_1,\dots,j_t\}}} c(n_1 - k, n_{i_1}, \dots, n_{i_s}) c(k-1, n_{j_1}, \dots, n_{j_t})$$

where the sum is over all pairs of subsets $A, B \subset [2, r]$ such that $A \cap B = \emptyset$ and $A \cup B = [2, r]$ including the possibility that either A or B could be empty.

Thus

$$\sum_{k=1}^{n_1}\sum_{\widehat{\mathcal{U}}\in\mathcal{P}_k}\zeta^{*^2}(\widehat{\mathcal{U}},\hat{\gamma}_k) = \sum_{k=1}^{n_1}\sum_{\substack{A=\{i_1,\dots,i_s\}\\ B=\{j_1,\dots,j_t\}}} c(n_1-k,n_{i_1},\dots,n_{i_s})c(k-1,n_{j_1},\dots,n_{j_t}) \tag{45}$$

Assembling equations (44) and (45) gives the result. □

In [OZ84], O'Brien and Zuber used a similar formula of this kind in order to compute the asymptotics of, so called, external field matrix integral. See also [BMS00] and Theorem 5.22.
Clearly, our notions around the convolution of functions on $\mathcal{PS}$ are analogous to (and motivated by) the convolution of functions on posets. Even though we are not able to put the above theory into the framework of posets, it seems that this analogy goes quite far. The following description of the Möbius functions is an instance of this—its poset analogue is due to Hall (see [Rot64]). It is essentially the simple observation that one can expand the Möbius function in terms of a geometric series as

$$\mu=\zeta^{*-1}=\big(\delta+(\zeta-\delta)\big)^{*-1}=\sum_{k=0}^{\infty}(-1)^k(\zeta-\delta)^{*k}.$$

Proposition 5.21. *We have for any* $(\mathcal{U},\gamma)\in\mathcal{PS}$ *that*

$$\mu(\mathcal{U},\gamma)=\delta(\mathcal{U},\gamma)+\sum_{k=1}^{\infty}\sum_{\substack{(\mathcal{U},\gamma)=(0,\pi_1)\cdots(0,\pi_k)\\ \pi_i\neq e\ \forall i}}(-1)^k.$$

Proof. As noted above this is just the geometric series for

$$(\delta+(\zeta-\delta))^{*-1}.$$

(Note that we are working for this in the algebra of functions on $\mathcal{PS}$ with the pointwise sum and the convolution as sum and product—we are not bothering about multiplicativity.) The only thing to check is that the sum is finite, and this is the case because the number of factors k is bounded by $|(\mathcal{U},\gamma)|$, since $|(0,\pi)|\geq 1$ for any $\pi\neq e$. □

This description of the Möbius function allows us now to derive a recursive formula for μ.

Theorem 5.22. *Consider* $(\mathcal{U},\gamma)\in\mathcal{PS}$ *such that* $\gamma(1)\neq 1$. *Then we have*

$$\mu(\mathcal{U},\gamma)=(-1)\sum_{\substack{(0,(1,k))\cdot(\mathcal{V},\pi)=(\mathcal{U},\gamma)\\ k\neq 1}}\mu(\mathcal{V},\pi), \tag{46}$$

where the sum runs over all decompositions of $(\mathcal{U},\gamma)$ *into a product of a disc transposition* $(0,(1,k))$ *(with* $k\geq 2$*) and a* $(\mathcal{V},\pi)\in\mathcal{PS}$.

The proof of this theorem will rely on the following lemma.

Lemma 5.23. *Let $(\mathcal{U},\gamma)\in\mathcal{PS}$ such that $\gamma(1)\neq 1$. For $p\in\mathbb{N}$, we denote by $\mathcal{S}_p$ the set consisting of all tuples $(\pi_1,\dots,\pi_p)$ of permutations such that $\pi_i\neq e$ for all $i=1,\dots,p$ and*

$$(0,\pi_1)\cdots(0,\pi_p)=(\mathcal{U},\gamma).$$

We consider now the two sums

$$S_1:=\sum_{p=1}^{\infty}\sum_{(\pi_1,\dots,\pi_p)\in\mathcal{S}_p}(-1)^p \tag{47}$$

and

$$S_2:=\sum_{p=1}^{\infty}\sum_{\substack{(\pi_1,\dots,\pi_p)\in\mathcal{S}_p\\ \pi_1=(1,k)\ for\ k\neq 1}}(-1)^p \tag{48}$$

where the second sum S_2 is over all tuples $(\pi_1,\dots,\pi_p)$ as for the first sum S_1, but now with the additional property that π_1 is a transposition interchanging the element 1 with some other element.
Then the two sums (47) *and* (48) *are equal,*

$$S_1=S_2.$$

Proof. Let $\pi=(\pi_1,\dots,\pi_p)\in\mathcal{S}_p$. Let $1\leq q\leq p$ denote the smallest index for which 1 is not a fixed point of π_q; note that such a q necessarily exists since $\gamma(1)\neq 1$. We shall group all factorizations into three classes: 1a), 1b) and 2). Class 1) consists of factorizations for which π_q is a transposition interchanging 1 with some other element. The subclass 1a) consists of factorizations for which $q=1$ and subclass 1b) of those for which $q\geq 2$. Class 2) consists of all other factorizations.
Let $\Pi=(\pi_1,\dots,\pi_p)$ be a factorization from the class 1b). We define

$$\Pi'=(\pi'_1,\dots,\pi'_{p-1})=(\pi_1,\dots,\pi_{q-2},\pi_{q-1}\pi_q,\pi_{q+1},\dots,\pi_p).$$

In the following we shall prove that $f:\Pi\mapsto\Pi'$ is a bijection between factorizations of class 1*b*) and factorizations of class 2).
Firstly, we prove that $\Pi'\in\mathcal{S}_p$ and is of class 2). Clearly, $\pi'_{q-1}=\pi_{q-1}\pi_q$ is a permutation which does not fix 1, it is not a transposition interchanging 1 with some other element, and we have

$$(0,\pi_{q-1})\cdot(0,\pi_q)=(0,\pi'_{q-1}).$$

In order to show that f is a bijection we shall describe its inverse. If $\Pi'=(\pi'_1,\dots,\pi'_{p-1})\in\mathcal{S}_p$ and is of class 2), we define $1\leq q\leq p-1$ to be the smallest number for which π'_{q-1} does not fix 1. There is a unique decomposition $\pi'_{q-1}=\pi_{q-1}\pi_q$ such that 1 is a fixed point of π_{q-1} and π_q is a transposition interchanging 1 with some other element. Thus $|\pi_{q-1}|+|\pi_q|=|\pi'_{q-1}|$. The assumption that the factorization Π' is of class 2) implies that $\pi_{q-1}\neq e$. For $1\leq i\leq q-2$ we set $\pi_i=\pi'_i$ and for $q+1\leq i\leq p$ we set $\pi_i=\pi'_{i-1}$. In this

way we defined $\Pi = (\pi_1, \ldots, \pi_p)$. Now it is easy to check that $g : \Pi' \mapsto \Pi$ is a left and right inverse of f.
Since the factorization Π and the corresponding Π' contribute to (47) with the opposite signs, the contribution of all factorizations of class 1b) cancels with the contribution of factorizations of class 2). $\square$

Proof of 5.22. In the proof we will consider all factorizations $(0, \pi_1) \cdot (0, \pi_2) \cdots (0, \pi_p) = (\mathcal{U}, \gamma)$ with the requirement that $\pi_i \neq e$ for all i, *i.e.* $(\pi_1, \ldots, \pi_p) \in \mathcal{S}_p$, as in the proof of Lemma 5.23. Sometimes we will require in addition that $\pi_1 = (1, k)$ with $k \neq 1$. To simplify the notation we will not explicitly state every time that $\pi_i \neq e$. Since $\gamma(1) \neq 1$ we have $\delta(\mathcal{U}, \gamma) = 0$. When γ is a transposition the right hand side of equation (46) is -1; so we can assume that γ is not a transposition. So by Proposition 5.21 we have

$$\begin{aligned}
\mu(\mathcal{U}, \gamma) &= \sum_{p=1}^{\infty} \sum_{\substack{(0,\pi_1)\cdots(0,\pi_p) \\ =(\mathcal{U},\gamma)}} (-1)^p \overset{(5.23)}{=} \sum_{p=1}^{\infty} \sum_{\substack{(0,(1,k))\cdots(0,\pi_p) \\ =(\mathcal{U},\gamma)}} (-1)^p \\
&= \sum_{p=2}^{\infty} \sum_{\substack{(0,(1,k)),(\mathcal{V},\pi) \\ (0,(1,k))\cdot(\mathcal{V},\pi)=(\mathcal{U},\gamma)}} \sum_{\substack{(0,\pi_2)\cdots(0,\pi_p) \\ =(\mathcal{V},\pi)}} (-1)^p \\
&= - \sum_{\substack{(0,(1,k)),(\mathcal{V},\pi) \\ (0,(1,k))\cdot(\mathcal{V},\pi)=(\mathcal{U},\gamma)}} \sum_{p=2}^{\infty} \sum_{\substack{(0,\pi_2)\cdots(0,\pi_p) \\ =(\mathcal{V},\pi)}} (-1)^{p-1} \\
&= - \sum_{\substack{(0,(1,k)),(\mathcal{V},\pi) \\ (0,(1,k))\cdot(\mathcal{V},\pi)=(\mathcal{U},\gamma)}} \sum_{p=2}^{\infty} \sum_{\substack{(0,\pi_2)\cdots(0,\pi_p) \\ =(\mathcal{V},\pi)}} (-1)^{p-1} \\
&= - \sum_{\substack{(0,(1,k)),(\mathcal{V},\pi) \\ (0,(1,k))\cdot(\mathcal{V},\pi)=(\mathcal{U},\gamma)}} \mu(\mathcal{V}, \pi)
\end{aligned}$$

$\square$

One observes that the recursion formulas for the Möbius function and for ζ^{*2} look very similar. However, there are some significant differences. The recursion for ζ^{*2} effectively expresses ζ^{*2} for n points in terms of ζ^{*2} for $n-1$ points. The recursion for the Möbius function does not reduce the number of points. Nevertheless, at least for first and second order one can match the two recursions and connect the values of the Möbius function with the values of the function ζ^{*2} (i.e., with the number of non-crossing partitions and non-crossing annular permutations). In order to see this let us first specify the meaning of Theorem 5.22 for first and second order. In first order we get

$$\mu(1_n, \gamma_n) = - \sum_{1 \leq k \leq n-1} \mu(1_k, \gamma_k) \mu(1_{n-k}, \gamma_{n-k}),$$

which shows that $(-1)^n \mu(1_{n+1}, \gamma_{n+1})$ and $\zeta^{*2}(1_n, \gamma_n)$ satisfy the same recursion (namely the one for the Catalan numbers). This is, of course, just the well-known fact [Kre72, Spe94] that the Möbius function on non-crossing partitions is given by the signed and shifted Catalan numbers. In second order our recursion reads

$$\begin{aligned}(-1)\mu(1_{m+n}, \gamma_{m,n}) &= m \cdot \mu(1_{m+n}, \gamma_{m+n}) \\ &+ \sum_{1 \leq k \leq n-1} \big(\mu(1_{m+k}, \gamma_{m,k})\mu(1_{n-k}, \gamma_{n-k}) + \mu(1_{m+n-k}, \gamma_{m,n-k})\mu(1_k, \gamma_k)\big),\end{aligned}$$

which we recognize — by taking into account the shifted relation between μ and ζ^{*2} on the first level — as the recursion for $(-1)^{m+n}\zeta^{*2}(1_{m+n}, \gamma_{m,n})$. Let us collect these explicit results about the Möbius function in the following theorem.

Theorem 5.24. *We have for $m, n \in \mathbb{N}$ that*

$$\mu(1_n, \gamma_n) = (-1)^{n-1} \cdot \#NC(n-1) = (-1)^{n-1} \cdot c_{n-1}$$

and

$$\mu(1_{m+n}, \gamma_{m,n}) = (-1)^{m+n} \cdot \#S_{NC}(m, n) = (-1)^{m+n} \cdot c_{m,n}.$$

For higher orders we were not able to match the values of μ with those of ζ^{*2}.

6. R-transform formulas

Let us consider the situation that two multiplicative functions f and h on $\mathcal{PS}$ are related by $h = f * \zeta$. We want to understand what this means for the relations between the numbers $\kappa_n := f(1_n, \gamma_n)$ and $\kappa_{m,n} := f(1_{m+n}, \gamma_{m,n})$ on one side and the numbers $\alpha_n := h(1_n, \gamma_n)$ and $\alpha_{m,n} := h(1_{m+n}, \gamma_{m,n})$ on the other side. In particular, we want to express this in terms of the generating power series of these numbers,

$$C(x) := 1 + \sum_{n \geq 1} \kappa_n x^n, \qquad C(x, y) := \sum_{m,n \geq 1} \kappa_{m,n} x^m y^n$$

and

$$M(x) := 1 + \sum_{n \geq 1} \alpha_n x^n, \qquad M(x, y) := \sum_{m,n \geq 1} \alpha_{m,n} x^m y^n.$$

(Note that the above summation corresponds to putting formally

$$f(1_0, \gamma_0) := 1 \qquad \text{and} \qquad f(1_0, \gamma_{0,0}) := 0$$

for a multiplicative f. Our notation is motivated by the fact that the most important realization of the relation $h = f * \zeta$ will be the situation where the α's are the correlation moments and the κ's the corresponding cumulants, thus M is a moment series and C is a cumulant series.) On the first order level we have

$$\alpha_n = \sum_{\pi \in NC(n)} f(0_\pi, \pi),$$

which is the usual moment-cumulant formula of free probability theory, and it is well-known [Spe94] that this is equivalent to

$$C\big(xM(x)\big) = M(x).$$

Our main goal now is to derive the analogue of this for the second order level. There we have

$$\alpha_{m,n} = \sum_{\pi \in S_{NC}(m,n)} f(0_\pi, \pi) + \sum_{\substack{\pi_1\times\pi_2 \in NC(m)\times NC(n)\\ |\mathcal{V}|=|\pi_1\times\pi_2|+1}} f(\mathcal{V}, \pi_1\times\pi_2).$$

It turns out that the second term, the sum over disconnected partitions, is quite easy to deal with. The first term, the sum over connected annular permutations, looks much more involved, however, one can handle this also if one realizes that one can reduce this first term to the second one. Namely, one can sum over all connected annular permutations by first bundling all through-cycles into one through-cycle and secondly decomposing this through-cycle into sub-cycles all of which are through-cycles. In this way one can reduce the problem of dealing with all annular non-crossing permutations to the problem of considering permutations with exactly one through-cycle and the problem of considering permutations where all cycles are through-cycles. The first problem corresponds exactly to the above sum over disconnected partitions. So we can write

$$\sum_{\pi \in S_{NC}(m,n)} f(0_\pi, \pi) = \sum_{\substack{\pi_1\times\pi_2 \in NC(m)\times NC(n)\\ |\mathcal{V}|=|\pi_1\times\pi_2|+1}} \tilde{f}(\mathcal{V}, \pi_1\times\pi_2),$$

where $\tilde{f}$ is now the multiplicative function corresponding to

$$\tilde{f}(1_n, \gamma_n) = \tilde{\kappa}_n, \qquad \tilde{f}(1_{m+n}, \gamma_{m,n}) = \tilde{\kappa}_{m,n}$$

with

$$\tilde{\kappa}_n := \kappa_n$$

and

$$\tilde{\kappa}_{m,n} := \sum_{\pi \in S_{NC}^{all}(m,n)} f(0_\pi, \pi).$$

Thus we can combine this to get finally

$$\begin{aligned}
\alpha_{m,n} &= \sum_{\substack{\pi_1\times\pi_2 \in NC(m)\times NC(n)\\ |\mathcal{V}|=|\pi_1\times\pi_2|+1}} \big(f(\mathcal{V}, \pi_1\times\pi_2) + \tilde{f}(\mathcal{V}, \pi_1\times\pi_2)\big)\\
&= \sum_{\substack{\pi_1\times\pi_2 \in NC(m)\times NC(n)\\ |\mathcal{V}|=|\pi_1\times\pi_2|+1}} g(\mathcal{V}, \pi_1\times\pi_2),
\end{aligned}$$

where g is the multiplicative function corresponding to

$$g(1_n, \gamma_n) = \tilde{\alpha}_n, \qquad g(1_{m+n}, \gamma_{m,n}) = \tilde{\alpha}_{m,n}$$

with

$$\tilde{\alpha}_n = \tilde{\kappa}_n = \kappa_n$$

and

$$\tilde{\alpha}_{m,n} = \kappa_{m,n} + \tilde{\kappa}_{m,n}.$$

So we have to translate the relation between $\tilde{\kappa}_{m,n}$ and f and the relation between $\alpha_{m,n}$ and g into relations between the corresponding formal power series.

PROPOSITION 6.1. *Let f be a multiplicative function on $\mathcal{PS}$ with*

$$f(1_n, \gamma_n) =: \kappa_n \qquad \text{and} \qquad C(x) := 1 + \sum_{n\geq 1} \kappa_n x^n.$$

Put

$$\tilde{\kappa}_{m,n} := \sum_{\pi \in S_{NC}^{all}(m,n)} f(0_\pi, \pi),$$

where $S_{NC}^{all}(m,n)$ denotes the permutations in $S_{NC}(m,n)$ for which all cycles are through-cycles. Consider the corresponding generating power series

$$\tilde{C}(x,y) := \sum_{m,n\geq 1} \tilde{\kappa}_{m,n} x^m y^n.$$

Then we have

$$\tilde{C}(x,y) = -xy \frac{\partial^2}{\partial x \partial y} \log\Big(\frac{xC(y) - yC(x)}{x-y}\Big),$$

or equivalently

$$\tilde{C}(x,y) = -xy \Big(\frac{\big(C(x) - xC'(x)\big)\big(C(y) - yC'(y)\big)}{\big(xC(y) - yC(x)\big)^2} - \frac{1}{(x-y)^2} \Big).$$

Proof. Note that we can parameterize an element $\pi \in S_{NC}^{all}(m,n)$ in a bijective way by specifying the number of cycles, the number of elements on each circle for all cycles, the position of a fixed element (let's say 1) in its cycle and the first element on the other circle of this cycle. Let us denote the number of cycles by r, the number of elements of the cycles on the first circle by $i_1, \dots, i_r$ and the number of elements of those cycles on the other circle by $j_1, \dots, j_r$. Thus the l-th cycle contains $i_l + j_l$ elements and makes the contribution $\kappa_{i_l + j_l}$ in the calculation of $\tilde{\kappa}_{m,n}$. We normalize things so that the first cycle contains the element 1. Fixing $i_1, \dots, i_r$ and $j_1, \dots, j_r$ we thus have i_1 possibilities for where 1 sits in the first cycle and n possibilities for the first element of this cycle on the other circle. This means we have

$$\tilde{\kappa}_{m,n} = \sum_{r\geq 1} \sum_{\substack{i_1,\dots,i_r \geq 1 \\ i_1+\dots+i_r = m}} \sum_{\substack{j_1,\dots,j_r\geq 1 \\ j_1+\dots+j_r = n}} i_1 n \kappa_{i_1+j_1} \cdots \kappa_{i_r+j_r}$$

and thus

$$\tilde{C}(x,y) = \sum_{r\geq 1} \sum_{i_1,\dots,i_r\geq 1} \sum_{j_1,\dots,j_r\geq 1} i_1(j_1+\cdots j_r)\kappa_{i_1+j_1}\cdots\kappa_{i_r+j_r}x^{i_1}\cdots x^{i_r}y^{j_1}\cdots y^{j_r}$$

$$= \sum_{r\geq 1} \sum_{i_1,\dots,i_r\geq 1} \sum_{j_1,\dots,j_r\geq 1} i_1 y\frac{\partial}{\partial y}\big(\kappa_{i_1 gb+j_1}\cdots\kappa_{i_r+j_r}x^{i_1}\cdots x^{i_r}y^{j_1}\cdots y^{j_r}\big)$$

$$= \sum_{r\geq 1} y\frac{\partial}{\partial y}\Big(\big(\sum_{i_1,j_1\geq 1} i_1\kappa_{i_1+j_1}x^{i_1}y^{j_1}\big)\cdot\big(\sum_{i_2,j_2\geq 1}\kappa_{i_2+j_2}x^{i_2}y^{j_2}\big)\cdots\big(\sum_{i_r,j_r\geq 1}\kappa_{i_r+j_r}x^{i_r}y^{j_r}\big)\Big)$$

Let us now use the notation

$$\hat{C}(x,y) := \sum_{i,j\geq 1}\kappa_{i+j}x^i y^j.$$

Then we can continue with

$$\begin{aligned}\tilde{C}(x,y) &= \sum_{r\geq 1} y\frac{\partial}{\partial y}\Big(\big(x\frac{\partial}{\partial x}\hat{C}(x,y)\big)\cdot\hat{C}(x,y)^{r-1}\Big)\\ &= \sum_{r\geq 1} xy\frac{\partial}{\partial y}\Big(\frac{1}{r}\frac{\partial}{\partial x}\big(\hat{C}(x,y)^r\big)\Big)\\ &= xy\frac{\partial}{\partial y}\frac{\partial}{\partial x}\Big(\sum_{r\geq 1}\frac{1}{r}\hat{C}(x,y)^r\Big)\\ &= -xy\frac{\partial}{\partial y}\frac{\partial}{\partial x}\log\big(1-\hat{C}(x,y)\big)\end{aligned}$$

The assertions follow now by noting that

$$\hat{C}(x,y) = 1 - \frac{xC(y)-yC(x)}{x-y}$$

and by working out the partial derivatives. □

Proposition 6.2. *Let g be a multiplicative function on $\mathcal{PS}$. Put*

$$\tilde{\alpha}_{m,n} := g(1_{m+n},\gamma_{m,n})$$

and denote its generating power series of second order by

$$H(x,y) := \sum_{m,n\geq 1}\tilde{\alpha}_{m,n}x^m y^n.$$

Put

$$\alpha_n := (g*\zeta)(1_n,\gamma_n)$$

and

$$\alpha_{m,n} := \sum_{\substack{(\mathcal{V},\pi_1\times\pi_2)\\ |\mathcal{V}|=|\pi_1\times\pi_2|+1}} g(\mathcal{V},\pi)$$

and denote the corresponding generating functions by

$$M(x) := 1+\sum_{n\geq 1}\alpha_n x^n \qquad \textit{and} \qquad M(x,y) := \sum_{m,n\geq 1}\alpha_{m,n}x^m y^n.$$

Then we have the relation

$$M(x,y) = H(xM(x), yM(y)) \cdot \Big(1 + x\frac{M'(x)}{M(x)}\Big) \cdot \Big(1 + y\frac{M'(y)}{M(y)}\Big).$$

Proof. Let us do the summation in the definition of $\alpha_{m,n}$ in the way that we first fix the two cycles $V_1 \in \pi_1$ and $V_2 \in \pi_2$ which are connected by $\mathcal{V}$ and sum over all possibilities for fixed V_1, V_2. If V_1 has k elements and V_2 has l elements then this contributes the factor $\tilde{\alpha}_{k,l}$. Furthermore, $\pi_1 \backslash V_1$ decomposes into k independent non-crossing partitions and the summations over them (for fixed V_1) gives the α_i for the intervals between consecutive elements from V_1. (Of course, we are counting here modulo m.) For the final summation over V_1 we have to notice that there are two different possibilities: either a fixed number (let's say 1) is an element of V_1 - in which case we can specify the situation by prescribing the number k of elements of V_1 and the differences $i_1, \dots, i_k$ between consecutive elements in V_1 - or 1 is not an element of V_1, — in which case we need an extra factor i_1, because we have now i_1 different possibilities how 1 can lie between two consecutive elements of V_1. Since we have the same situation for V_2 we can thus write $\alpha_{m,n}$ in the form

$$\alpha_{m,n} = \sum_{k,l\geq 1} \sum_{\substack{i_1,\dots,i_k\geq 0\\ k+i_1+\dots+i_k=m}} \sum_{\substack{j_1,\dots,j_l\geq 0\\ l+j_1+\dots+j_l=n}} \tilde{\alpha}_{k,l}\alpha_{i_1}\cdots\alpha_{i_k}\alpha_{j_1}\cdots\alpha_{j_l}\Big(1 + i_1 + j_1 + i_1 j_1\Big).$$

Translating this into generating power series gives the assertion. □

The combination of the previous two propositions, with

$$H(x,y) = C(x,y) + \tilde{C}(x,y),$$

gives now our main result.

Theorem 6.3. *Let f and h be multiplicative functions on $\mathcal{PS}$ which are related by*

$$h = f * \zeta.$$

Denote

$$\kappa_n := f(1_n, \gamma_n), \qquad \kappa_{m,n} := f(1_{m+n}, \gamma_{m,n})$$

and

$$\alpha_n := h(1_n, \gamma_n), \qquad \alpha_{m,n} := h(1_{m+n}, \gamma_{m,n})$$

and define the corresponding generating power series

$$C(x) := 1 + \sum_{n\geq 1} \kappa_n x^n, \qquad C(x,y) := \sum_{m,n\geq 1} \kappa_{m,n} x^m y^n$$

and

$$M(x) := 1 + \sum_{n\geq 1} \alpha_n x^n, \qquad M(x,y) := \sum_{m,n\geq 1} \alpha_{m,n} x^m y^n.$$

Then we have as formal power series the first order relation

$$C(xM(x)) = M(x) \tag{49}$$

and for the second order

$$M(x,y) = H\big(xM(x), yM(y)\big) \cdot \frac{\frac{d}{dx}(xM(x))}{M(x)} \cdot \frac{\frac{d}{dy}(yM(y))}{M(y)}, \tag{50}$$

where

$$H(x,y) := C(x,y) - xy \frac{\partial^2}{\partial x \partial y} \log\Big(\frac{xC(y) - yC(x)}{x-y}\Big), \tag{51}$$

or equivalently,

$$\begin{aligned} M(x,y) = C\big(xM(x), yM(y)\big) \cdot \frac{\frac{d}{dx}(xM(x))}{M(x)} \cdot \frac{\frac{d}{dy}(yM(y))}{M(y)} \\ + xy\Big(\frac{\frac{d}{dx}(xM(x)) \cdot \frac{d}{dy}(yM(y))}{(xM(x) - yM(y))^2} - \frac{1}{(x-y)^2}\Big). \end{aligned} \tag{52}$$

Proof. The formulation (50) and (51) follows directly from a combination of Propositions 6.1 and 6.2. In order to reformulate this to (52) one uses the equivalence of the two formulas in Proposition 6.1 and the fact that $C(xM(x)) = M(x)$ yields

$$1 - xC'(xM(x)) = \frac{M(x)}{\frac{d}{dx}(xM(x))}.$$

□

If we go over from the moment generating series M to a kind of Cauchy transform like quantity G, then these formulas take on a particularly nice form.

COROLLARY 6.4. *Consider the same situation and notations as in Theorem 6.3. In terms of*

$$G(x) := \frac{1}{x}M(1/x), \quad G(x,y) := \frac{1}{xy}M(1/x, 1/y), \quad \mathcal{R}(x,y) := \frac{1}{xy}C(x,y)$$

the Equation (52) can be written as

$$G(x,y) = G'(x)G'(y)\Big\{\mathcal{R}(G(x), G(y)) + \frac{1}{(G(x) - G(y))^2}\Big\} - \frac{1}{(x-y)^2}. \tag{53}$$

$\mathcal{R}(x,y)$ is the second order R-transform. Note that Voiculescu's first order R-transform $\mathcal{R}$ is defined by the relation $C(x) = 1 + x\mathcal{R}(x)$, and equation (49) says for this

$$\frac{1}{G(x)} + \mathcal{R}(G(x)) = x,$$

i.e., that $G(x)$ and $K(x) := \frac{1}{x} + \mathcal{R}(x)$ are inverses of each other under composition.

EXAMPLE 6.5. Let us apply our formulas to some examples.
1) If we put f to be the multiplicative function with $\kappa_2 = 1$ and all other κ_n and all $\kappa_{m,n}$ vanishing, then $h = f * \zeta$ counts the non-crossing pairings, i.e., in this case $M(x)$ is the generating function of the number of non-crossing

pairings (on one circle) and $M(x,y)$ is the generating function of the number of non-crossing annular pairings (on two circles). Let us calculate it by using the above theorem.
We have

$$C(x) = 1 + x^2, \qquad C(x,y) = 0$$

and we know that M is the generating function of number of non-crossing pairings on a circle. In this case

$$\hat{C}(x,y) = xy,$$

and thus

$$H(x,y) = -xy \frac{\partial^2}{\partial x \partial y} \log(1-xy) = \frac{xy}{(1-xy)^2},$$

which yields the result

$$M(x,y) = xy \cdot \frac{\frac{d}{dx}(xM(x)) \cdot \frac{d}{dy}(yM(y))}{\big(1 - xyM(x)M(y)\big)^2}.$$

Related formulas are known in the physical literature, see, e.g, [FMP78], [AJM90], [ACKM93] and also [BZ93], [KKP95].
2) If we put $f = \zeta$ then $h = \zeta * \zeta$ counts the non-crossing permutations, i.e., in this case $M(x)$ is the generating function of the number of non-crossing permutations (which is the same as non-crossing partition) on one circle and $M(x,y)$ is the generating function of the number of annular non-crossing permutations (on two circles).
We have

$$C(x) = \frac{1}{1-x}, \qquad C(x,y) = 0.$$

In this case

$$\hat{C}(x,y) = \frac{1-x-y}{(1-x)(1-y)},$$

and thus

$$H(x,y) = -xy \frac{\partial^2}{\partial x \partial y} \log(1-xy) = \frac{xy}{(1-x-y)^2},$$

which yields

$$M(x,y) = xy \cdot \frac{\frac{d}{dx}(xM(x)) \cdot \frac{d}{dy}(yM(y))}{\big(1 - xM(x) - yM(y)\big)^2}.$$

3) Let us finally see whether we can extract the value of the Möbius function from our formula. Since we have $\delta = \mu * \zeta$, our formula with

$$M(x) = 1 + x, \qquad M(x,y) = 0$$

should allow to solve for $C(x,y)$ which is then the generating function for the annular Möbius function. Note that we already know $M(x)$ in this case to be the generating function of the disc Möbius function.

If $M(x,y)$ vanishes identically this implies that $H(x,y)$ vanishes identically, leading to the identity

$$C(x,y) = xy\frac{\partial^2}{\partial x \partial y}\log(\frac{xC(y)-yC(x)}{x-y})$$
$$= xy\Big(\frac{(C(x)-xC'(x))\cdot(C(y)-yC'(y))}{(xC(y)-yC(x))^2} - \frac{1}{(x-y)^2}\Big)$$

7. Higher order freeness and corresponding cumulants

7.1. Abstract framework.

Definition 7.1. A *higher-order (non-commutative) probability space*, or briefly HOPS, $(\mathcal{A},\varphi)$ consists of a unital algebra $\mathcal{A}$ and a collection $\varphi = (\varphi_n)_{n\in\mathbb{N}}$ of maps ($n \in \mathbb{N}$)

$$\varphi_n : \underbrace{\mathcal{A}\times\cdots\times\mathcal{A}}_{n \text{ times}} \to \mathbb{C},$$

which are linear and tracial in each of its n arguments and which are symmetric under exchange of its n arguments and which satisfy

$$\varphi_1(1) = 1$$

and

$$\varphi_n(1,a_2,\dots,a_n) = 0$$

for all $n \geq 2$ and all $a_2,\dots,a_n \in \mathcal{A}$.

Of course, we can include the usual (first order) non-commutative probability space $(\mathcal{A},\varphi_1)$ into this framework by putting all higher φ_n equal to zero. In the same way we recover a second order non-commutative probability space $(\mathcal{A},\varphi_1,\varphi_2)$ by putting $\varphi_n = 0$ for all $n \geq 3$.

Definition 7.2. 1) We denote by $\mathcal{PS}(\mathcal{A})$ the set of partitioned permutations decorated with elements from $\mathcal{A}$, i.e.,

$$\mathcal{PS} = \bigcup_{n\in\mathbb{N}}\big(\mathcal{PS}(n)\times\mathcal{A}^n\big).$$

2) For a function

$$f : \mathcal{PS}(\mathcal{A}) \to \mathbb{C}$$
$$(\mathcal{V},\pi)\times(a_1,\dots,a_n) \mapsto f(\mathcal{V},\pi)[a_1,\dots,a_n]$$

and a function

$$g : \mathcal{PS} \to \mathbb{C}$$

we define their convolution

$$f * g : \mathcal{PS}(\mathcal{A}) \to \mathbb{C}$$

by

$$(f*g)(\mathcal{U},\gamma)[a_1,\dots,a_n] := \sum_{\substack{(\mathcal{V},\pi),(\mathcal{W},\sigma)\in\mathcal{PS}(n)\\(\mathcal{V},\pi)\cdot(\mathcal{W},\sigma)=(\mathcal{U},\gamma)}} f(\mathcal{V},\pi)[a_1,\dots,a_n]\cdot g(\mathcal{W},\sigma)$$

for all $(\mathcal{U},\gamma)\in\mathcal{PS}(n)$ and all $a_1,\dots,a_n\in\mathcal{A}$.

Definition 7.3. A function $f:\mathcal{PS}(\mathcal{A})\to\mathbb{C}$ is called *multiplicative* if we have

$$f(\mathcal{V},\pi)[a_1,\dots,a_n]=\prod_{B\in\mathcal{V}} f(1_B,\pi|_B)[(a_1,\dots,a_n)_B]$$

and

$$f(1_n,\sigma^{-1}\pi\sigma)[a_{\sigma(1)},\dots,a_{\sigma(n)}]=f(1_n,\pi)[a_1,\dots,a_n]$$

for all $a_1,\dots,a_n\in\mathcal{A}$ and all $\pi,\sigma\in S(n)$.

Note that this extension of our formalism on multiplicative functions on $\mathcal{PS}$ and their convolution from the last section is not changing the results from the last section. The structure of all formulas remains the same; one just has to insert the $a_1,\dots,a_n$ as dummy variables at the right positions. Thus, in particular, δ is still the unit for this extended convolution and $f=g*\zeta$ is equivalent to $g=f*\mu$ for multiplicative f,g on $\mathcal{PS}(\mathcal{A})$. And again, the convolution of a multiplicative function on $\mathcal{PS}(\mathcal{A})$ with a multiplicative function on $\mathcal{PS}$ gives a multiplicative function on $\mathcal{PS}(\mathcal{A})$.
It is clear that a multiplicative function f on $\mathcal{PS}(\mathcal{A})$ is uniquely determined by the values of $f(1_n,\gamma_{n(1),\dots,n(r)})[a_1,\dots,a_n]$ (where we put $n:=n(1)+\cdots+n(r)$) for all $r\in\mathbb{N}$, all $n(1),\dots,n(r)\in\mathbb{N}$ and all $a_1,\dots,a_n\in\mathcal{A}$.

7.2. Moment and cumulant functions. Let us now apply this formalism to get moment and cumulant functions for higher order probability spaces. So let a HOPS $(\mathcal{A},\varphi)$ be given. We will use the φ_n to produce a multiplicative "moment" function on $\mathcal{PS}(\mathcal{A})$, which we will also denote by φ. Namely, we put

$$\varphi(1_n,\gamma_{n(1),\dots,n(r)})[a_1,\dots,a_n] \\ :=\varphi_r(a_1\cdots a_{n(1)};\dots;a_{n(1)+\cdots n(r-1)+1}\cdots a_n)$$

and extend this by multiplicativity. (Note that we need the φ_n to be tracial in their arguments for this extension.)
Here is an example for our function φ.

$$\varphi\big(\{1,3,4\}\{2\},(1,3)(2)(4)\big)[a_1,a_2,a_3,a_4]=\varphi_2(a_1a_3,a_4)\cdot\varphi_1(a_2)$$

Definition 7.4. For a given HOPS $(\mathcal{A},\varphi)$ we define the corresponding *(higher order) free cumulants* as a function on $\mathcal{PS}(\mathcal{A})$ by

$$\kappa=\varphi*\mu,$$

or more explicitly

$$\kappa(\mathcal{U},\gamma)[a_1,\dots,a_n]:=\sum_{\substack{(\mathcal{V},\pi),(\mathcal{W},\sigma)\in\mathcal{PS}(n)\\(\mathcal{V},\pi)\cdot(\mathcal{W},\sigma)=(\mathcal{U},\gamma)}}\varphi(\mathcal{V},\pi)[a_1,\dots,a_n]\cdot\mu(\mathcal{W},\sigma),$$

for all $n\in\mathbb{N}$, $(\mathcal{U},\gamma)\in\mathcal{PS}(n)$, $a_1,\dots,a_n\in\mathcal{A}$.

As we noted before the definition above is equivalent to the statement $\varphi = \kappa * \zeta$, i.e.,

$$\varphi(\mathcal{U},\gamma)[a_1,\dots,a_n] = \sum_{(\mathcal{V},\pi)\in\mathcal{PS}_{NC}(\mathcal{U},\gamma)} \kappa(\mathcal{V},\pi)[a_1,\dots,a_n]$$

for all $(\mathcal{U},\gamma)[a_1,\dots,a_n] \in \mathcal{PS}(\mathcal{A})$.
Furthermore, as with φ, κ is also a multiplicative function on $\mathcal{PS}(\mathcal{A})$. Thus in the same way as all $\varphi(\mathcal{U},\gamma)$ are determined by the knowledge of all

$$\varphi(1_n,\gamma_{n(1),\dots,n(r)})[a_1,\dots,a_n] \\ = \varphi_r(a_1\cdots a_{n(1)};\dots;a_{n(1)+\cdots+n(r-1)+1}\cdots a_{n(1)+\cdots+n(r)})$$

the free cumulants $\kappa(\mathcal{U},\gamma)$ are determined by the values of

$$\kappa(1_n,\gamma_{n(1),\dots,n(r)})[a_1,\dots,a_n] \\ =: \kappa_{n(1),\dots,n(r)}(a_1,\dots,a_{n(1)};\dots;a_{n(1)+\cdots+n(r-1)+1},\dots,a_{n(1)+\cdots+n(r)}).$$

Remark 7.5. Note that whereas on the level of φ we also know (by definition) that we can multiply elements along the cycles of π (and thus we do not need a comma as separator for those elements along a cycle), this is not true for κ. Thus we have, e.g.,

$$\varphi(1_3,(1,2)(3))[a_1,a_2,a_3] = \varphi_2(a_1a_2;a_3) = \varphi(1_2,(1),(2))[a_1a_2;a_3],$$

but no clear relation exists among

$$\kappa(1_3,(1,2)(3))[a_1,a_2,a_3] = \kappa_{2,1}(a_1,a_2;a_3)$$

and

$$\kappa(1_2,(1),(2))[a_1a_2;a_3] = \kappa_{1,1}[a_1a_2;a_3].$$

Note also that since our convolution on $\mathcal{PS}$ coincides on the first level with the usual convolution of multiplicative functions on non-crossing partitions, the above definition of cumulants reduces on the first level to the usual free cumulants.

7.3. Higher order freeness. Equipped with the notion of cumulants we can now define "freeness" by the requirement of vanishing of mixed cumulants.

Definition 7.6. We say that a family $(\mathcal{X}_i)_{i\in I}$ of subsets of $\mathcal{A}$ is *free (of all orders)* if we have the following vanishing of mixed cumulants: For all $n \geq 2$ and all $a_k \in \mathcal{X}_{i(k)}$ $(1 \leq k \leq n)$ such that $i(p) \neq i(q)$ for some $1 \leq p,q \leq n$ we have

$$\kappa(1_n,\pi)[a_1,\dots,a_n] = 0$$

for all $\pi \in S(n)$.

Example 7.7. Let us see that this definition includes the definition of Voiculescu [VDN92] for (first order) freeness and the definition of Mingo and Speicher [MS06] for second order freeness.
1) On the first level this follows from the fact that our cumulants reduce then to the usual free cumulants and it is well-known that freeness is equivalent to

the vanishing of mixed cumulants. One can see it directly as follows: Let us consider $a_k \in \mathcal{X}_{i(k)}$ with $i(k) \neq i(k+1)$ and $\varphi_1(a_k) = 0$ for all $k = 1, \dots, n$. Then we have

$$\varphi_1(a_1 \cdots a_n) = \varphi(1_n, \gamma_n)[a_1, \dots, a_n] = \sum_{\pi \in NC(n)} \kappa(0_\pi, \pi)[a_1, \dots, a_n].$$

However the vanishing of mixed moments means now that the only π which contribute are those which do not connect elements from different sets. Furthermore, the fact that all our variables are centered excludes singletons. But then it is easy to see that there are no such π at all, so the sum is zero.
2) Now we have to consider two cyclically alternating and centered tuples $a_1, \dots, a_m$ and $b_1, \dots, b_n$. Then we have

$$\begin{aligned}\varphi_2(a_1 \cdots a_m; b_1 \cdots b_n) &= \varphi(1_{m+n}, \gamma_{m,n})[a_1, \dots, a_m, b_1, \dots, b_n] \\ &= \sum_{(\mathcal{V},\pi) \in \mathcal{PS}_{NC}(m,n)} \kappa(\mathcal{V}, \pi)[a_1, \dots, a_m, b_1, \dots, b_n].\end{aligned}$$

Again, the vanishing of mixed moments requires that $(\mathcal{V}, \pi)$ connects only elements from the same set and the centeredness of the elements excludes singletons. It is then easy to see that, for $n \geq 2$, the only possibilities for such $(\mathcal{V}, \pi)$ arise for $m = n$ and they have to be disc permutations $(0_\pi, \pi)$ which are pairings $(a_1, b_{1+s})(a_2, b_{2+s}) \cdots (a_n, b_{n+s})$ for some s. The factors $k(1_2, (...))[a_k, b_{k+s}]$ are just $\varphi_1(a_2 b_{2+s})$, so that one finally gets, for $n \geq 2$, the formula

$$\varphi_2(a_1 \cdots a_m; b_1 \cdots b_n) = \delta_{mn} \sum_{k=1}^{n} \varphi_1(a_1 b_{1+s}) \cdots \varphi_1(a_n b_{n+s}).$$

For $n = m = 1$ one gets with

$$\varphi_2(a_1; b_1) = k_2(a_1, a_2) + k_{1,1}(a_1; b_1)$$

the conclusion that $\varphi_2(a_1; b_1)$ has to vanish if a_1 and b_1 are from different sets. Nothing is required if both are from the same set. We see that we get exactly the defining properties for second order freeness from [MS06].
3) It would be nice to be able to reformulate in a similar way the definition of higher order freeness in terms of the φ instead of the cumulants. However, the situation with more than two circles is getting much more involved and we are not aware of such a reformulation for third and higher order freeness.

As in the case of the first order freeness, one sees immediately that constants are free from everything.

Proposition 7.8. *Let $(\mathcal{A}, \varphi)$ be a HOPS. Then $\{1\}$ is free of all orders from every subset $\mathcal{X} \subset \mathcal{A}$.*

Proof. We have to prove that

$$\kappa(1_n, \gamma_{n(1),\dots,n(r)})[1, a_2, \dots, a_n] = 0,$$

unless $n = 1$. We will do this by induction on n. The case $n = 2$ is clear because

$$\kappa(1_2, (12))[1, a_2] = \varphi_1(1 \cdot a_2) - \varphi_1(1) \cdot \varphi_1(a_2) = 0$$

and

$$\kappa(1_2, (1)(2))[1, a_2] = \varphi_2(1; a_2) = 0.$$

In general, one has

$$\begin{aligned}
&\varphi(1_n, \gamma_{n(1),\dots,n(r)})[1, a_2, \dots, a_n] \\
&\qquad = \sum_{(\mathcal{V},\pi)\in\mathcal{PS}_{NC}(n(1),\dots,n(r))} \kappa(\mathcal{V}, \pi)[1, a_2, \dots, a_n] \\
&\qquad = \kappa(1_n, \gamma_{n(1),\dots,n(r)})[1, a_2, \dots, a_n] \\
&\qquad\qquad + \sum_{\substack{(\mathcal{V},\pi)\in\mathcal{PS}_{NC}(n(1),\dots,n(r)) \\ |(\mathcal{V},\pi)|<|(1_n,\gamma_{n(1),\dots,n(r)})|}} \kappa(\mathcal{V}, \pi)[1, a_2, \dots, a_n]
\end{aligned}$$

By induction hypothesis, in the later sum exactly terms of the form $(\{1\} \cup \tilde{\mathcal{V}}, (1) \cup \tilde{\pi})$ with

$$(\tilde{\mathcal{V}}, \tilde{\pi}) \in \mathcal{PS}_{NC}(n(1) - 1, n(2), \dots, n(r))$$

contribute. In the case $n(1) > 1$ the sum over those yields

$$\varphi(1_{n-1}, \gamma_{n(1)-1,n(2),\dots,n(r)}[a_2, \dots, a_n].$$

In this case, also

$$\varphi(1_n, \gamma_{n(1),\dots,n(r)})[1, a_2, \dots, a_n] = \varphi(1_{n-1}, \gamma_{n(1)-1,n(2),\dots,n(r)}[a_2, \dots, a_n],$$

and thus $\kappa(1_n, \gamma_{n(1),\dots,n(r)})[1, a_2, \dots, a_n] = 0$. If, on the other side, $n(1) = 1$ (i.e., 1 is the only element on its circle), then we have to set

$$\mathcal{PS}_{NC}(0, n(2), \dots, n(r)) = \emptyset,$$

because then the first circle cannot be connected to the others if we ask 1 to be a cycle of its own. But this means that in this case

$$\kappa(1_n, \gamma_{n(1),\dots,n(r)})[1, a_2, \dots, a_n] = \varphi(1_n, \gamma_{n(1),\dots,n(r)})[1, a_2, \dots, a_n]$$

However, for $n(1) = 1$ and $n > 1$ we have

$$\varphi(1_n, \gamma_{1,\dots,n(r)})[1, a_2, \dots, a_n] = 0.$$

□

Note that our definition of freeness behaves clearly very nicely with respect to decompositions of our sets. For example, we have that $\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3$ are free if and only if $\mathcal{X}_1$ and $\mathcal{X}_2 \cup \mathcal{X}_3$ are free and $\mathcal{X}_2$ and $\mathcal{X}_3$ are free. Thus we can reduce the investigation of freeness to the understanding of freeness for the case of two sets. A characterization for this is given in the next theorem.

THEOREM 7.9. *Let $(\mathcal{A}, \varphi)$ be a higher order probability space and consider two subsets of $\mathcal{X}_1, \mathcal{X}_2 \subset \mathcal{A}$. Then the following are equivalent.*

(1) *The sets $\mathcal{X}_1, \mathcal{X}_2$ are free of all orders.*

(2) *The sets* $\mathcal{X}_1 \cup \{1\}$, $\mathcal{X}_2 \cup \{1\}$ *are free of all orders.*

(3) *We have*

$$\varphi(\mathcal{U},\gamma)[a_1b_1,\dots,a_nb_n] = \sum_{(\mathcal{V},\pi)\cdot(\mathcal{W},\sigma)=(\mathcal{U},\gamma)} \kappa(\mathcal{V},\pi)[a_1,\dots,a_n]\cdot\varphi(\mathcal{W},\sigma)[b_1,\dots,b_n]$$

for all $n \in \mathbb{N}$, *all* $(\mathcal{U},\gamma) \in \mathcal{PS}(n)$ *and all* $a_1,\dots,a_n \in \mathcal{X}_1 \cup \{1\}$, $b_1,\dots,b_n \in \mathcal{X}_2 \cup \{1\}$.

(4) *We have*

$$\varphi(\mathcal{U},\gamma)[a_1b_1,\dots,a_nb_n] = \sum_{(\mathcal{V},\pi)\cdot(\mathcal{W},\sigma)=(\mathcal{U},\gamma)} \varphi(\mathcal{V},\pi)[a_1,\dots,a_n]\cdot\kappa(\mathcal{W},\sigma)[b_1,\dots,b_n]$$

for all $n \in \mathbb{N}$, *all* $(\mathcal{U},\gamma) \in \mathcal{PS}(n)$ *and all* $a_1,\dots,a_n \in \mathcal{X}_1 \cup \{1\}$, $b_1,\dots,b_n \in \mathcal{X}_2 \cup \{1\}$.

(5) *We have*

$$\kappa(\mathcal{U},\gamma)[a_1b_1,\dots,a_nb_n] = \sum_{(\mathcal{V},\pi)\cdot(\mathcal{W},\sigma)=(\mathcal{U},\gamma)} \kappa(\mathcal{V},\pi)[a_1,\dots,a_n]\cdot\kappa(\mathcal{W},\sigma)[b_1,\dots,b_n]$$

for all $n \in \mathbb{N}$, *all* $(\mathcal{U},\gamma) \in \mathcal{PS}(n)$ *and all* $a_1,\dots,a_n \in \mathcal{X}_1 \cup \{1\}$, $b_1,\dots,b_n \in \mathcal{X}_2 \cup \{1\}$.

In order to prove this we would like to write $\varphi(\mathcal{U},\gamma)[a_1b_1,\dots,a_nb_n]$ in the form $\varphi(\hat{\mathcal{U}},\hat{\gamma})[a_1,b_1,\dots,a_n,b_n]$. Let us introduce the following formalism for this. Let $(\mathcal{U},\gamma) \in \mathcal{PS}(n)$ be a partitioned permutation of the numbers $1,2,3,\dots,n$. Double now this set of numbers by introducing a copy $\bar{1},\bar{2},\bar{3},\dots,\bar{n}$ and interleave the new and old numbers as follows:

$$1,\bar{1},2,\bar{2},3,\bar{3},\dots,n,\bar{n}.$$

If we induce now $(\mathcal{U},\gamma)$ on $1,2,\dots,n$ to $(\hat{\mathcal{U}},\hat{\gamma})$ on $1,\bar{1},\dots,n,\bar{n}$ by putting

$$\hat{\gamma}(k) = \bar{k} \qquad \text{and} \qquad \hat{\gamma}(\bar{k}) = \gamma(k),$$

then this has exactly the desired property. The vanishing of mixed cumulants means that in the factorizations of $(\hat{\mathcal{U}},\hat{\gamma})$ in $(\mathcal{V},\pi)$ times a disc permutation we are only interested in $(\mathcal{V},\pi)$ which have the property that each block of $\mathcal{V}$ contains either only unbarred numbers or only barred numbers, i.e., $(\mathcal{V},\pi)$ must be of the form $(\mathcal{V}_a \cup \mathcal{V}_b, \pi_a \cup \pi_b)$ with

$$(\mathcal{V}_a,\pi_a) \in \mathcal{PS}(1,\dots,n) \qquad \text{and} \qquad (\mathcal{V}_b,\pi_b) \in \mathcal{PS}(\bar{1},\dots,\bar{n}).$$

Let us first observe some simple relations between the quantities on $1,\dots,n$ and their relatives on $1,\bar{1},\dots,n,\bar{n}$.

LEMMA 7.10. *1) We have*

$$|\hat{\gamma}| = n + |\gamma|, \qquad |\hat{\mathcal{U}}| = n + |\mathcal{U}|,$$

and thus

$$|(\hat{\mathcal{U}}, \hat{\gamma})| = n + |(\mathcal{U}, \gamma)|.$$

2) We have

$$|\pi_a \cup \pi_b| = |\pi_a| + |\pi_b|, \qquad |\mathcal{V}_a \cup \mathcal{V}_b| = |\mathcal{V}_a| + |\mathcal{V}_b|,$$

and thus

$$|(\mathcal{V}_a \cup \mathcal{V}_b, \pi_a \cup \pi_b)| = |(\mathcal{V}_a, \pi_a)| + |(\mathcal{V}_b, \pi_b)|.$$

3) We have that $(\pi_a \cup \pi_b)\hat{\gamma}$ maps unbarred to barred and barred to unbarred elements and, for all $k = 1, \dots, n$,

$$[(\pi_a \cup \pi_b)\hat{\gamma}]^2(\bar{k}) = \pi_b \pi_a \gamma(k),$$

thus

$$|(\pi_a \cup \pi_b)\hat{\gamma}| = n + |\pi_b \pi_a \gamma|$$

Proof. Only the third part is non-trivial. To see this observe

$$(\pi_a \cup \pi_b)\hat{\gamma}(\bar{k}) = \pi_a(\gamma(k))$$

and thus

$$[(\pi_a \cup \pi_b)\hat{\gamma}]^2(\bar{k}) = \pi_b \overline{\pi_a(\gamma(k))},$$

which is our first equation, with the identification of $\pi_b \in S(\bar{1}, \dots, \bar{1})$ with the corresponding permutation in $S(1, \dots, n)$. Since the mapping between barred and unbarred elements is clear, this yields that $(\pi_a \cup \pi_b)\hat{\gamma}$ and $\pi_b \pi_a \gamma$ have the same number of orbits which gives the last equation. □

This lemma allows us to characterize the contributing factorizations in $(\hat{\mathcal{U}}, \hat{\gamma})$ in terms of special factorizations of $(\mathcal{U}, \gamma)$.

PROPOSITION 7.11. *The statement*

$$(\mathcal{V}_a \cup \mathcal{V}_b, \pi_a \cup \pi_b) \in \mathcal{PS}_{NC}(\hat{\mathcal{U}}, \hat{\gamma})$$

is equivalent to the statement

$$(\mathcal{V}_a, \pi_a) \cdot (\mathcal{V}_b, \pi_b) \in \mathcal{PS}_{NC}(\mathcal{U}, \gamma),$$

where in the last product we identify $(\mathcal{V}_b, \pi) \in \mathcal{PS}(\bar{1}, \dots, \bar{n})$ with the corresponding element in $\mathcal{PS}(1, \dots, n)$.

Proof. Note that $(\mathcal{V}_a \cup \mathcal{V}_b, \pi_a \cup \pi_b) \in \mathcal{PS}_{NC}(\hat{\mathcal{U}}, \hat{\gamma})$ is equivalent to

$$|(\mathcal{V}_a \cup \mathcal{V}_b, \pi_a \cup \pi_b)| + |(\pi_a \cup \pi_b)^{-1}\hat{\gamma}| = |(\hat{\mathcal{U}}, \hat{\gamma})| \tag{54}$$

and

$$\hat{\mathcal{U}} = (\mathcal{V}_a \cup \mathcal{V}_b) \vee \hat{\gamma}. \tag{55}$$

On the other hand, $(\mathcal{V}_a, \pi_a) \cdot (\mathcal{V}_b, \pi_b) \in \mathcal{PS}_{NC}(\mathcal{U}, \gamma)$, means

$$(\mathcal{V}_a, \pi_a) \cdot (\mathcal{V}_b, \pi_b) \cdot (0_{\pi_b^{-1}\pi_a^{-1}\gamma}, \pi_b^{-1}\pi_a^{-1}\gamma) = (\mathcal{U}, \gamma),$$

which is equivalent to

$$|(\mathcal{V}_a,\pi_a)|+|(\mathcal{V}_b,\pi_b)|+|\pi_b^{-1}\pi_a^{-1}\gamma|=|(\mathcal{U},\gamma)| \tag{56}$$

and

$$\mathcal{U}=\mathcal{V}_a\vee\mathcal{V}_b\vee\gamma. \tag{57}$$

Equations (54) and (56) are, by Lemma 7.10, equivalent.
The equivalence between (55) and (57) is also easily checked. □

Equipped with these tools we can now prove our main Theorem 7.9.

Proof. The equivalences between (3), (4), and (5) follow by convolving with the ζ or the μ function. That (2) is actually the same as (1) follows from Prop. 7.8.
(1) $\Longrightarrow$ (3): We have

$$\begin{aligned}\varphi(\mathcal{U},\gamma)[a_1b_1,\dots,a_nb_n]&=\varphi(\hat{\mathcal{U}},\hat{\gamma})[a_1,b_1,\dots,a_n,b_n]\\&=\sum_{(\mathcal{V},\pi)\cdot(\mathcal{W},\sigma)=(\hat{\mathcal{U}},\hat{\gamma})}\kappa(\mathcal{V},\pi)[a_1,b_1,\dots,a_n,b_n]\cdot\zeta(\mathcal{W},\sigma)\\&=\sum_{(\mathcal{V},\pi)\in\mathcal{PS}_{NC}(\hat{\mathcal{U}},\hat{\gamma})}\kappa(\mathcal{V},\pi)[a_1,b_1,\dots,a_n,b_n]\end{aligned}$$

By our assumption on the vanishing of mixed cumulants, only $(\mathcal{V},\pi)$ of the form $(\mathcal{V}_1\cup\mathcal{V}_2,\pi_a\cup\pi_b)$ with

$$(\mathcal{V}_a,\pi_a)\in\mathcal{PS}(1,\dots,n)\qquad\text{and}\qquad(\mathcal{V}_b,\pi_b)\in\mathcal{PS}(\bar{1},\dots,\bar{n})$$

contribute and, by the above Proposition 7.11,

$$(\mathcal{V}_a\cup\mathcal{V}_b,\pi_a\cup\pi_b)\in\mathcal{PS}_{NC}(\hat{\mathcal{U}},\hat{\gamma})$$

is equivalent to

$$(\mathcal{V}_a,\pi_a)\cdot(\mathcal{V}_b,\pi_b)\in\mathcal{PS}_{NC}(\mathcal{U},\gamma).$$

Thus we can continue with

$$\begin{aligned}&\varphi(\mathcal{U},\gamma)[a_1b_1,\dots,a_nb_n]\\&=\sum_{(\mathcal{V}_a\cup\mathcal{V}_b,\pi_a\cup\pi_b)\in\mathcal{PS}_{NC}(\hat{\mathcal{U}},\hat{\gamma})}\kappa(\mathcal{V}_a,\pi_a)[a_1,a_2,\dots,a_n]\cdot\kappa(\mathcal{V}_b,\pi_b)[b_1,b_2,\dots,b_n]\\&=\sum_{(\mathcal{V}_a,\pi_a)\cdot(\mathcal{V}_b,\pi_b)\in\mathcal{PS}_{NC}(\mathcal{U},\gamma)}\kappa(\mathcal{V}_a,\pi_a)[a_1,a_2,\dots,a_n]\cdot\kappa(\mathcal{V}_b,\pi_b)[b_1,b_2,\dots,b_n]\\&=\sum_{(\mathcal{V}_a,\pi_a)\cdot(\mathcal{V}_b,\pi_b)\cdot(\mathcal{W},\sigma)=(\mathcal{U},\gamma)}\kappa(\mathcal{V}_a,\pi_a)[a_1,a_2,\dots,a_n]\cdot\\&\qquad\qquad\cdot\kappa(\mathcal{V}_b,\pi_b)[b_1,b_2,\dots,b_n]\cdot\zeta(\mathcal{W},\sigma)\\&=\sum_{(\mathcal{V}_a,\pi_a)\cdot(\mathcal{V},\pi)=(\mathcal{U},\gamma)}\kappa(\mathcal{V}_a,\pi_a)[a_1,\dots,a_n]\cdot\varphi(\mathcal{V},\pi)[b_1,\dots,b_n].\end{aligned}$$

(3) $\Longrightarrow$ (1): Note that (3) allows us to calculate all moments of elements from $\mathcal{X}_1\cup\mathcal{X}_2$ out of the moments of elements from $\mathcal{X}_1$ and the moments of elements

from $\mathcal{X}_2$. (In order to do so, we also have to allow some of the a's or b's to be equal to the unit 1.) Since this calculation rule is the same as for free sets, this shows that the sets $\mathcal{X}_1$ and $\mathcal{X}_2$ must be free. $\square$

This theorem is now the key ingredient to transfer freeness from sets to their generated algebras.

Theorem 7.12. *Let $(\mathcal{A}, \varphi)$ be a HOPS and consider subsets $(\mathcal{X}_i)_{i\in I}$. For each $i \in I$, let $\mathcal{A}_i$ be the unital algebra generated by elements from $\mathcal{X}_i$. Then the following are equivalent.*

(1) *The subsets $(\mathcal{X}_i)_{i\in I}$ are free of all orders.*
(2) *The subalgebras $(\mathcal{A}_i)_{i\in I}$ are free of all orders.*

Proof. Since the cumulant $\kappa(\mathcal{V},\pi)[a_1,\dots,a_n]$ is a multi-linear functional in the n variables $a_1,\dots,a_n$, it is clear that taking sums of elements within the sets $\mathcal{X}_i$ preserves freeness. What we have to see is that also taking products preserves freeness. Since we can iterate our arguments, it suffices to see the following: if $\mathcal{X}_1$ and $\mathcal{X}_2$ are free, then also $\mathcal{X}_1 \cup \{a_0 a_1 \mid a_0, a_1 \in \mathcal{X}_1\}$ and $\mathcal{X}_2$ are free. Adding one product after the other to $\mathcal{X}_1$ and by Theorem 7.9 it is enough to show that

$$\varphi(\mathcal{U},\gamma)[a_0a_1b_1, a_2b_2\dots,a_nb_n] = \sum_{(\mathcal{V},\pi)\cdot(\mathcal{W},\sigma)=(\mathcal{U},\gamma)} \varphi(\mathcal{V},\pi)[a_0a_1,a_2\dots,a_n]\cdot\kappa(\mathcal{W},\sigma)[b_1,\dots,b_n]$$

for all $n\in\mathbb{N}$, all $(\mathcal{U},\gamma)\in\mathcal{PS}(n)$ and all $a_0,a_1,\dots,a_n\in\mathcal{X}_1\cup\{1\}$, $b_1,\dots,b_n\in\mathcal{X}_2\cup\{1\}$. Let us induce $(\mathcal{U},\pi)\in\mathcal{PS}(1,\dots,n)$ to $(\hat{\mathcal{U}},\hat\pi)\in\mathcal{PS}(0,1,\dots,n)$ by requiring that $\hat{\mathcal{W}}$ and $\hat\pi$ restricted to $1,\dots,n$ agree with $\mathcal{W}$ and π, respectively, and that 0 and 1 are in the same block of $\hat{\mathcal{W}}$ and $\hat\pi(0)=1$. Then we can calculate

$$\begin{aligned}\varphi(\mathcal{U},\gamma)[a_0a_1b_1, a_2b_2\dots,a_nb_n] &= \varphi(\hat{\mathcal{U}},\hat\pi)[a_01, a_1b_1, a_2b_2,\dots,a_nb_n]\\ &= \sum_{(\mathcal{V},\pi)\cdot(\mathcal{W},\sigma)=(\hat{\mathcal{U}},\hat\gamma)} \varphi(\mathcal{V},\pi)[a_0,a_1,a_2\dots,a_n]\cdot\kappa(\mathcal{W},\sigma)[1,b_1,\dots,b_n].\end{aligned}$$

By Proposition 7.8 we know that $\kappa(\mathcal{W},\sigma)[1,b_1,\dots,b_n]$ is only different from zero if $\mathcal{W}$ has 0 as a singleton, i.e., $(\mathcal{W},\sigma)$ has to be of the form

$$\mathcal{W}=\{0\}\cup\tilde{\mathcal{W}}, \qquad \sigma=(0)\tilde\sigma,$$

with

$$(\tilde{\mathcal{W}},\tilde\sigma)\in\mathcal{PS}(1,\dots,n).$$

But then we must have that $\pi(0)=1$ and 0 and 1 must be in the same block of $\mathcal{V}$. Thus there is a unique $(\mathcal{V}',\pi')$ so that $(\mathcal{V},\pi)=(\hat{\mathcal{V}}',\hat\pi')$ and

$$(\mathcal{V},\pi)\cdot(\mathcal{W},\sigma)=(\hat{\mathcal{U}},\hat\gamma)$$

is equivalent to

$$(\mathcal{V}',\pi')\cdot(\tilde{\mathcal{W}},\tilde\sigma)=(\mathcal{U},\gamma).$$

Note also that in this situation

$$\kappa(\mathcal{W},\sigma)[a_0,a_1,,\dots,a_n]=\kappa(\tilde{\mathcal{W}},\tilde{\sigma})[b_1,b_2,\dots,b_n].$$

and

$$\varphi(\hat{\mathcal{V}}',\hat{\pi}')[a_0,a_1,,\dots,a_n]=\varphi(\mathcal{V}',\pi')[a_0a_1,a_2,\dots,a_n].$$

So we can continue the above calculation as follows

$$\begin{aligned}&\varphi(\mathcal{U},\gamma)[a_0a_1b_1,a_2b_2\dots,a_nb_n]\\&=\sum_{(\mathcal{V}',\pi')\cdot(\tilde{\mathcal{W}},\tilde{\sigma})=(\mathcal{U},\gamma)}\varphi(\mathcal{V}',\pi')[a_0a_1,a_2\dots,a_n]\cdot\kappa(\tilde{\mathcal{W}},\tilde{\sigma})[b_1,\dots,b_n],\end{aligned}$$

which is exactly what we had to show. □

7.4. DISTRIBUTION OF ONE RANDOM VARIABLE. For the case where we restrict our attention to just one random variable $a\in\mathcal{A}$ we introduce the following notation.

NOTATION 7.13. Let $(\mathcal{A},\varphi)$ be a HOPS and let $a\in\mathcal{A}$.
1) For, $(\mathcal{V},\pi)\in\mathcal{PS}(n)$, we will write

$$\varphi^a(\mathcal{V},\pi):=\varphi(\mathcal{V},\pi)\underbrace{[a,\dots,a]}_{n\text{-times}}$$

and

$$\kappa^a(\mathcal{V},\pi):=\kappa(\mathcal{V},\pi)\underbrace{[a,\dots,a]}_{n\text{-times}}.$$

2) A *Young diagram* is a $\lambda=(\lambda_1,\dots,\lambda_l)$ for some $l\in\mathbb{N}$ and $\lambda_1,\dots,\lambda_l\in\mathbb{N}$ with $\lambda_1\geq\lambda_2\geq\cdots\geq\lambda_l$. We put $|\lambda|:=\lambda_1+\cdots+\lambda_l$ (the total number of boxes of the Young diagram λ). The set of all Young diagrams will be denoted by $\mathbf{Y}$.
3) The information about the higher order moments of a can also be parameterized by Young diagrams as follows: for $\lambda=(\lambda_1,\dots,\lambda_l)$ we put

$$\varphi^a(\lambda):=\varphi(1_{|\lambda|},\pi)\underbrace{[a,\dots,a]}_{n\text{-times}}=\varphi_l(a^{\lambda_1},\dots,a^{\lambda_l})$$

where π is any permutation whose conjugacy class corresponds to λ (i.e., $\pi\in S_{|\lambda|}$ has cycles of length $\lambda_1,\dots,\lambda_l$. The collection of all higher order moments $(\varphi^a(\lambda))_{\lambda\in\mathbf{Y}}$ is called the *(higher order) distribution of* a.
4) Similarly as for moments, we put

$$\kappa^a(\lambda):=\kappa(1_{|\lambda|},\pi)\underbrace{[a,\dots,a]}_{n\text{-times}},$$

where π is any permutation whose conjugacy class corresponds to λ.

Remark 7.14. For first and second order moments and cumulants, we used in Section 2 also the following notations:

$$\alpha_n:=\varphi^a(1_n,\gamma_n)\qquad\alpha^a_{m,n}:=\varphi^a(1_{m+n},\gamma_{m,n}),$$

and

$$\kappa_n^a := \kappa^a(1_n, \gamma_n) \qquad \kappa_{m,n} := \kappa^a(1_{m+n}, \gamma_{m,n}),$$

where γ_n and $\gamma_{m,n}$ are permutations with one cycle and two cycles, respectively.

The vanishing of mixed cumulants translates in this framework into the additivity of the cumulants for sums of free variables.

Theorem 7.15. *Let $(\mathcal{A}, \varphi)$ be a HOPS and $a, b \in \mathcal{A}$ free of all orders. Then we have*

$$\kappa^{a+b}(\lambda) = \kappa^a(\lambda) + \kappa^b(\lambda)$$

for all $\lambda \in \mathbf{Y}$.

Proof. By the multilinearity of the cumulants and the vanishing of mixed cumulants for free variable, we have for any $n \in \mathbb{N}$ and $\pi \in S_n$:

$$\begin{aligned}\kappa^{a+b}(1_n, \pi) &= \kappa(1_n, \pi)[a+b, \dots, a+b] \\ &= \kappa(1_n, \pi)[a, \dots, a] + \kappa(1_n, \pi)[b, \dots, b] \\ &= \kappa^a(1_n, \pi) + \kappa^b(1_n, \pi).\end{aligned}$$

□

8. Random matrices, Itzykson-Zuber integrals and higher order freeness

8.1. Asymptotic higher order freeness of random matrices. Let us now come back to our original motivation for our theory – the asymptotic behavior of random matrices. In order to reformulate our calculations from Section 4 in our language of higher order freeness we still need to define the notion of "asymptotic freeness".

Definition 8.1. 1) Let $(\mathcal{A}, \varphi)$ and, for each $N \in \mathbb{N}$, $(\mathcal{A}_N, \varphi^{(N)})$ be HOPSs. Let I be an index set and for each $i \in I$, $a_i \in \mathcal{A}$ and $a_i^{(N)} \in \mathcal{A}_N$ $(N \in \mathbb{N})$. We say that the family $(a_i^{(N)} \mid i \in I)$ converges, for $N \to \infty$, to $(a_i \mid i \in I)$, denoted by

$$(a_i^{(N)})_{i \in I} \to (a_i)_{i \in I},$$

if we have for all $n \in \mathbb{N}$ and all polynomials $p_1, \dots, p_n$ in $|I|$-many non-commuting indeterminates that

$$\lim_{N\to\infty} \varphi_n^{(N)} \Big(p_1\big((a_i^{(N)})_{i\in I}\big), \dots, p_n\big((a_i^{(N)})_{i\in I}\big)\Big) \\ = \varphi_n \left(p_1\big((a_i)_{i\in I}\big), \dots, p_n\big((a_i)_{i\in I}\big)\right). \tag{58}$$

2) Let, for each $N \in \mathbb{N}$, $(\mathcal{A}_N, \varphi^{(N)})$ be HOPSs. Let I be an index set and, for each $i \in I$ and $N \in \mathbb{N}$, $a_i^{(N)} \in \mathcal{A}_N$. We say that the sequence of families $(a_i^{(N)})_{i\in I}$ has a *limit distribution of all orders* if there exists a HOPS $(\mathcal{A}, \varphi)$ such that

$$(a_i^{(N)})_{i \in I} \to (a_i)_{i \in I},$$

for some $a_i \in \mathcal{A}$ $(i \in I)$

3) Let, for each $N \in \mathbb{N}$, $(\mathcal{A}_N, \varphi^{(N)})$ be HOPSs. Let I be an index set and, for each $i \in I$ and $N \in \mathbb{N}$, $a_i^{(N)} \in \mathcal{A}_N$. Let $I = I_1 \cup \cdots \cup I_k$ be a decomposition of I into k disjoint subsets. We say that the sets $\{a_i^{(N)} \mid i \in I_1\}, \dots, \{a_i^{(N)} \mid i \in I_k\}$ are *asymptotically free of all orders* if there exists a HOPS $(\mathcal{A}, \varphi)$ such that

$$(a_i^{(N)})_{i \in I} \to (a_i)_{i \in I},$$

for some $a_i \in \mathcal{A}$ $(i \in I)$ and such that the sets $\{a_i \mid i \in I_1\}, \dots, \{a_i \mid i \in I_k\}$ are free of all orders in $(\mathcal{A}, \varphi)$.

With this notation and by invoking Theorem 7.9 we can reformulate our main result on random matrices, Theorem 4.4, in the following form.

Theorem 8.2. *Let $\mathcal{M}_N := M_N \otimes L^{\infty-}(\Omega)$ be an ensemble of $N \times N$-random matrices. Define rescaled correlation functions $\tilde{\varphi}^{(N)} = (\tilde{\varphi}_n^{(N)})_{n \in \mathbb{N}}$ on $\mathcal{M}_N$ by ($n \in \mathbb{N}$, $D_1, \dots, D_n \in \mathcal{M}_N$)*

$$\tilde{\varphi}_n^{(N)}(D_1, \dots, D_n) := k_n(\mathrm{Tr}(D_1), \dots, \mathrm{Tr}(D_n)) \cdot N^{2-n}. \tag{59}$$

Assume that we have, for each $N \in \mathbb{N}$, subalgebras $\mathcal{A}_N, \mathcal{B}_N \in \mathcal{M}_N$ such that

(1) *$\mathcal{A}_N$ is a unitarily invariant ensemble,*
(2) *$\mathcal{A}_N$ and $\mathcal{B}_N$ are independent.*

Let $(A_i^{(N)})_{i \in I}$ be a family of elements in $(\mathcal{A}_N, \tilde{\varphi}^{(N)})$ which has a higher order limit distribution and let $(B_j^{(N)})_{j \in J}$ ($N \in \mathbb{N}$) be a family of elements in $(\mathcal{B}_N, \tilde{\varphi}^{(N)})$ which has a higher order limit distribution. Then the families $\{A_i^{(N)} \mid i \in I\}$ and $\{B_j^{(N)} \mid j \in J\}$ are asymptotically free of all orders.

8.2. Itzykson-Zuber integrals.

Definition 8.3. For $N \times N$ matrices A_N, B_N their *Itzykson-Zuber integral* is defined as the following function in $z \in \mathbb{C}$:

$$\mathrm{IZ}(z, A_N, B_N) := N^{-2} \log E(e^{zN\mathrm{Tr}(A_N U B_N U^*)}),$$

where U denotes a Haar unitary $N \times N$-random matrix.

Consider now a sequence of such matrices A_N and B_N. Note that A_N and B_N are non-random, thus all distributions of order higher than 1 vanish identically. If we assume that A_N and B_N have a first order (eigenvalue) limit distribution for $N \to \infty$, then it is known (see [Col03]) that each Taylor coefficient about zero of $z \to IZ(z, A_N, B_N)$ admits a limit as $N \to \infty$. Note that the effect of the Haar unitary random matrix in the above Itzykson-Zuber integral was to make A_N and UB_NU^* asymptotically free of all orders. We show now that this kind of result extends also to the case of random matrices A_N and B_N, and that our theory allows to identify the limit of the Taylor coefficients very precisely.

Theorem 8.4. *Let $A = (A_N)_{N \in \mathbb{N}}$ and $B = (B_N)_{N \in \mathbb{N}}$ be two ensembles of $N \times N$-random matrices which are asymptotically free of all orders with respect* *to the rescaled correlation functions $\tilde{\varphi}^{(N)}$. Denote the corresponding limiting distribution of $(A_N)_{N\in\mathbb{N}}$ by φ^a and the corresponding limit distribution of $(B_N)_{N\in\mathbb{N}}$ by φ^b. Then, as formal power series in z, we have*

(60)

$$\lim_{N\to\infty} N^{-2}\log \mathrm{E}[e^{zN\mathrm{Tr}(A_N B_N)}] = \sum_{n=1}^{\infty}\frac{z^n}{n!}\sum_{\substack{(\mathcal{V},\pi),(\mathcal{W},\sigma)\in\mathcal{PS}(n)\\ (\mathcal{V},\pi)\cdot(\mathcal{W},\sigma)=(1_n,e)}}\kappa^a(\mathcal{V},\pi)\cdot\varphi^b(\mathcal{W},\sigma).$$

Proof. Recall that the logarithm of the exponential generating series of the moments of a random variable is the exponential generating series of the classical cumulants of that variable. Thus we have

$$\begin{aligned}N^{-2}\cdot \log\mathrm{E}[e^{zNA_NB_N}]&\\ &= N^{-2}\sum_{n=1}^{\infty}k_n(N\mathrm{Tr}(A_NB_N),\dots,N\mathrm{Tr}(A_NB_N))\cdot\frac{z^n}{n!}\\ &=\sum_{n=1}^{\infty}N^{n-2}\cdot\varphi^{(N)}(1_n,e)[A_NB_N,\dots,A_NB_N]\cdot\frac{z^n}{n!}.\end{aligned}$$

By our assumption that A_N and B_N are asymptotically free with respect to $\tilde{\varphi}^{(N)}_n = N^{n-2}\varphi^N_n$, this converges to

$$\sum_{n=1}^{\infty}\varphi(1_n,e)[ab,\dots,ab]\cdot\frac{z^n}{n!},$$

where a and b are free of all orders with respect to φ. Theorem 7.9 yields then the assertion. □

In a forthcoming work we will investigate the relevance of higher order freeness for Itzykson-Zuber integrals more detailed, in particular, in comparison with and extension of results of Zinn-Justin [ZJ99], Collins [Col03], and Guionnet and Maida [GM05].

9. Appendix: Surfaced permutations

In this appendix we will present a more geometrical view on partitioned permutations. As we shall see in the following, partitioned permutations are just special cases of "surfaced permutations"; in particular the results of this article can be equivalently formulated in the language of surfaced permutations. On the other hand, for the purpose of this article we do not need anything more than just partitioned permutations and the Reader not interested in surfaced permutations may skip this Section without much harm.

9.1. Motivations. Our goal is to study factorizations of permutations, i.e. solutions $(\pi_1,\dots,\pi_k)$ of the equation

$$\gamma = \pi_1\cdots\pi_k,$$

where $\gamma\in S_n$ is some fixed permutation and $\pi_1,\dots,\pi_k\in S_n$ are subject to some additional constraints, depending on a particular context. Typically, one

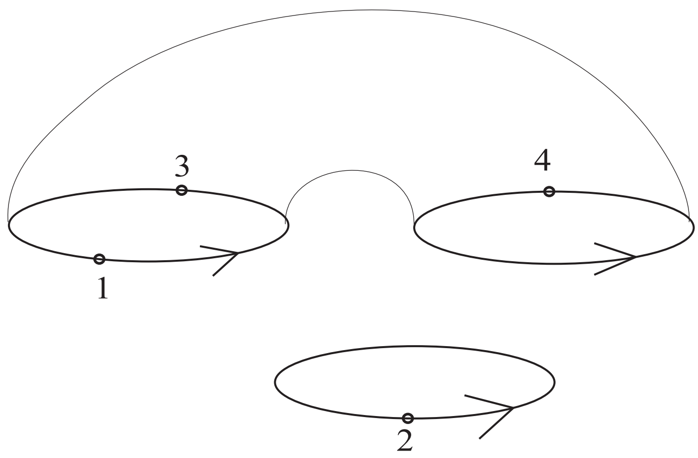


Figure 1. Example of a surfaced permutation. Its support is equal to $(1,3)(2)(4) \in S_4$. This surfaced permutation corresponds to a partitioned permutation $\big(\{1,3,4\}\{2\},(1,3)(2)(4)\big)$

.

of these constrains concerns $|\pi_1| + \cdots + |\pi_k|$, the other one concerns the orbits of the action of $\pi_1, \dots, \pi_k$.
It would be very useful to equip permutations $\pi_1, \dots, \pi_k$ with some additional structure in such a way that the product $\tilde{\pi}_1 \cdots \tilde{\pi}_k$ of the resulting enriched permutations $\tilde{\pi}_1, \dots, \tilde{\pi}_k$ would carry both the information about the product $\pi_1 \cdots \pi_k$ of permutations and the information about $|\pi_1| + \cdots + |\pi_k|$. As we shall see in the following, surfaced permutations provide an appropriate tool.

9.2. Definition. We say that $\sigma = (S, j)$ is a *surfaced permutation* of some finite set A if S is a two–dimensional surface with a fixed orientation and with a boundary ∂S and if $j : A \to \partial S$ is a injection. We can think about the information carried by j as follows: some of the points on the boundary ∂S are distinguished and carry different labels from the set A. We also require that every connected component of ∂S carries at least one distinguished point. An example of a surfaced permutation is presented on Figure 1.
We identify surfaced permutations (S_1, j_1), (S_2, j_2) of the same set A if there exists a orientation preserving homeomorphism $f : S_1 \to S_2$ such that $f \circ j_1 = j_2$. The set of surfaced permutations of set $\{1, \dots, n\}$ will be denoted by $\mathcal{SS}_n$.

9.3. Surfaced permutations and the usual permutations. Let $(S, j) \in \mathcal{SS}_n$; the boundary ∂S with the inherited orientation from S is just a collection of oriented circles with some distinguished points labeled $1, \dots, n$ marked on them. In this way we can define a permutation $\sigma \in S_n$, called the support of (S, j), the cycles of which correspond to connected components of ∂S, as it can

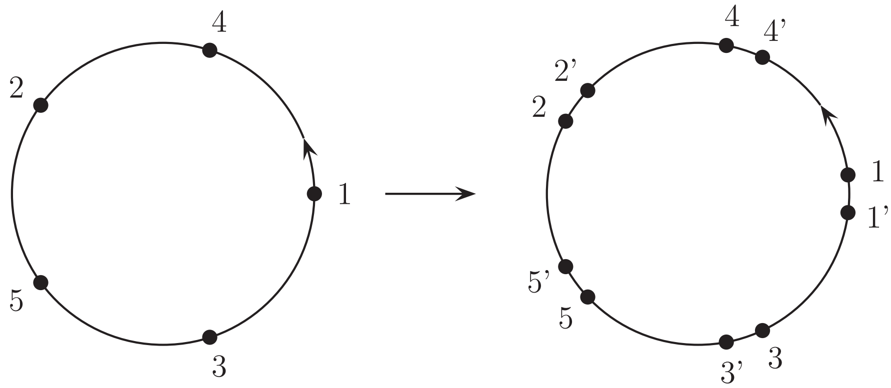


Figure 2. Convention for splitting labels.

be seen on Figure 1. It is therefore a good idea to think that a surfaced permutation is just a (usual) permutation $\sigma \in S_n$ equipped with some additional information carried by the surface S.

A surfaced permutation $(S, j) \in \mathcal{SS}_n$ can be uniquely specified (up to the equivalence relation) by its support $\sigma \in S_n$ and by specifying the shape of the connected components of S. The latter information is given by an equivalence relation on cycles of σ (each class corresponds to a connected component of S) and furthermore for each class of this relation we should specify the genus of the corresponding connected component of S. Above it should be understood that the genus of a surface S with a boundary is by definition equal to the genus of a surface S' without boundary obtained from S by gluing a disc to every connected component of ∂S; for example both a disc and the lateral surface of a cylinder have genus zero.

9.4. Surfaced permutations and partitioned permutations.
Among surfaced permutations a special class will be very important for our purposes, namely surfaced permutations (S, j) such that each connected component of S has genus zero. It is easy to see that there is a bijection between such surfaced permutations (S, j) and partitioned permutations $(\mathcal{V}, \sigma)$ given as follows: σ is the support of (S, j) and $\mathcal{V}$ is the partition given by connected components of S.

9.5. Products of surfaced permutations. Let surfaced permutations $(S_1, j_1), (S_2, j_2) \in \mathcal{SS}_n$ be given. On the boundary of S_2 there are marked points labeled by numbers $1, \dots, n$; let us split every marked point k into a consecutive pair of points k and k', as it is presented on the example from Figure 2. In the second step, for each $k \in \{1, \dots, n\}$ we glue a small neighborhood of the vertex $k \in \partial S_1$ to a small neighborhood of the vertex $k' \in \partial S_2$ in such a way that the orientations of S_1 and S_2 coincide. In this way we obtain a new surface S which has marked points on its boundary ∂S and these are exactly the vertices from ∂S_2 labeled $1, \dots, n$; we denote the resulting surfaced permutation by (S, j) and we call it a product $(S_1, j_1)(S_2, j_2)$ of the original surfaced permutations. This choice of gluing surfaces S_1 and S_2 implies that the support of $(S_1, j_1)(S_2, j_2)$ is equal to the product of the support of (S_1, j_1) and the support of (S_2, j_2).

It is not difficult to explain now the definition of the product of partitioned permutations (Definition 4.9): we treat partitioned permutations as surfaced permutations and compute their product; if the genus of the resulting surface is zero we can identify it with another partitioned permutation, otherwise we set the product to be zero.

It is not difficult to show that for surfaced permutations the product is associative and the associativity of the product of partitioned permutations is a simple corollary.

References

[AJM90] J. Ambjørn, J. Jurkiewicz, and Yu. M. Makeenko: Multiloop correlators for two-dimensional quantum gravity. *Physics Letters B.* 251: 517–524, 1990.

[ACKM93] J. Ambjørn, L. Chekhov, C. F. Kristjansen, and Yu. M. Makeenko: Matrix model calculations beyond the spherical limit. *Nuclear Physics B* 404: 127–172, 1993.

[AZ06] G. W. Anderson, O. Zeitouni: A CLT for a band matrix model. *Probab. Theory Relat. Fields*, 134: 283–338, 2006.

[BS04] Z. D. Bai and J. Silverstein: CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Prob.*, 32: 533–605, 2004.

[Bia97] P. Biane: Some properties of crossings and partitions. *Discrete Math.*, 175(1-3):41–53, 1997.

[BMS00] M. Bousquet-Mélou and G. Schaeffer: Enumeration of planar constellations. *Adv. in Appl. Math.*, 24(4):337–368, 2000.

[BZ93] É. Brézin, A. Zee: Universality of the correlations between eigenvalues of large random matrices. *Nuclear Phys.* B 402 (1993), no. 3, 613–627.

[CC06] M. Capitaine and M. Casalis: Cumulants for random matrices as convolutions over the symmetric group. *Probab. Theory Relat. Fields*, 136 (2006), 19-36.

[Col03] B. Collins: Moments and cumulants of polynomial random variables on unitary groups, the Itzykson-Zuber integral, and free probability. *Int. Math. Res. Not.*, (17):953–982, 2003.

[CŚ06] B. Collins and P. Śniady: Integration with respect to the Haar measure on unitary, orthogonal and symplectic groups. *Comm. Math. Phy.*, 264, (2006), 773 - 795.

[Dia03] P. Diaconis: Patterns in Eigenvalues: The 70th Josiah Willard Gibbs Lecture. *Bulletin of the AMS*, 40: 155-178, 2003.

[FMP78] J. B. French, P. A. Mello, A. Pandey: Statistical properties of many-particle spectra. II. Two-point correlations and fluctuations. *Ann. of Physics*, 113 (1978), no. 2, 277–293.

[GM05] A. Guionnet and M. Maïda: A Fourier view on the R-transform and related asymptotics of spherical integrals. *J. Funct. Anal.*, 222(2):435–490, 2005.

[Joh98] K. Johansson: On fluctuations of eigenvalues of random Hermitian matrices. *Duke Math. J.*, 91(1):151–204, 1998.

[Jon99] V. F. R. Jones: Planar algebras. I. `math.QA/9909027`, 1999, 122 pp.

[KKP95] A. Khorunzhy, B. Khoruzhenko, L. Pastur: On the $1/N$ corrections to the Green functions of random matrices with independent entries. *J. Phys.* A 28 (1995), L31–L35.

[Kre72] G. Kreweras: Sur les partitions non croisées d'un cycle. *Discrete Math.*, 1(4):333–350, 1972.

[LS59] V. P. Leonov and A. N. Shiryaev: On a method of semi-invariants. *Theory of Probability and its Applications*, 4, 319–329, 1959.

[MN04] J. Mingo and A. Nica: Annular noncrossing permutations and partitions, and second-order asymptotics for random matrices. *Int. Math. Res. Not.*, (28):1413–1460, 2004.

[MŚS07] J. Mingo, P. Śniady, and R. Speicher. Second order freeness and fluctuations of random matrices: II. Unitary random matrices. *Adv. in Math.* 209: 212 – 240, (2007).

[MS06] J. Mingo and R. Speicher: Second Order Freeness and Fluctuations of Random Matrices: I. Gaussian and Wishart matrices and Cyclic Fock spaces. *J. Funct. Anal.*, 235, 2006, pp. 226-270.

[NSp97] A. Nica and R. Speicher: A "Fourier transform" for multiplicative functions on non-crossing partitions. *J. Algebraic Combin.*, 6(2):141–160, 1997.

[NSp06] A. Nica and R. Speicher: Lectures on the Combinatorics of Free Probability, London Mathematical Society Lecture Note Series, New York: Cambridge University Press, 2006.

[OZ84] K. H. O'Brien and J.-B. Zuber: A note on U(N) integrals in the large N limit. *Phys. Lett. B*, 144(5-6):407–408, 1984.

[Oko00] A. Okounkov: Random matrices and random permutations. *Int. Math. Res. Not.*, 2000, no. 20: 1043–1095, 2000.

[Rad06] F. Radulescu: Combinatorial aspects of Connes's embedding conjecture and asymptotic distribution of traces of products of unitaries, *Operator theory* 20, 197–205, Theta Ser. Adv. Math., 6, Theta, Bucharest, 2006

[Rot64] G.-C. Rota: On the foundations of combinatorial theory. I. Theory of Möbius functions. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 2: 340–368, 1964.

[Śni06] P. Śniady: Gaussian fluctuations of characters of symmetric groups and of Young diagrams. *Probab. Theory Related Fields* 136: 263–297, 2006.

[Spe94] R. Speicher: Multiplicative functions on the lattice of noncrossing partitions and free convolution. *Math. Ann.*, 298(4):611–628, 1994.

[VDN92] D. V. Voiculescu, K. J. Dykema, and A. Nica: *Free random variables.* American Mathematical Society, Providence, RI, 1992.

[Voi85] D. Voiculescu: Symmetries of some reduced free product C^*-algebras. In *Operator Algebras and their Connections with Topology and Ergodic Theory* (Lecture Notes in Mathematics, vol. 1132, Springer-Verlag): 556-588, 1985.

[Voi86] D. Voiculescu: Addition of certain non-commuting random variables. *J. Funct. Anal.*, 66: 323–346, 1986.

[Voi91] D. Voiculescu: Limit laws for random matrices and free products. *Invent. Math.*, 104(1):201–220, 1991.

[Wei78] D. Weingarten: Asymptotic behavior of group integrals in the limit of infinite rank. *J. Mathematical Phys.*, 19(5):999–1001, 1978.

[ZJ99] P. Zinn-Justin. Adding and multiplying random matrices: a generalization of Voiculescu's formulas. *Phys. Rev. E (3)*, 59(5, part A):4884–4888, 1999.

Benoît Collins
University of Ottawa
585 King Edward Ave.
Ottawa, ON, K1N 6N5, Canada
and
CNRS, Institut Camille Jordan
Université Claude Bernard Lyon 1
43 boulevard du 11 novembre 1918
69622 Villeurbanne Cedex,
France
bcollins@uottawa.ca

James A. Mingo
Queen's University,
Department of Mathematics
and Statistics
Jeffery Hall, Kingston, ON
K7L 3N6, Canada
mingo@mast.queensu.ca

Piotr Śniady
Instytut Matematyczny
Uniwersytet Wroclawski
pl Grunwaldzki 2/4
50-384 Wroclaw
Poland
Piotr.Sniady@math.uni.wroc.pl

Roland Speicher
Queen's University,
Department of Mathematics
and Statistics
Jeffery Hall, Kingston, ON,
K7L 3N6, Canada
speicher@mast.queensu.ca