Fields Institute Monographs 35

The Fields Institute for Research in Mathematical Sciences

James A. Mingo
Roland Speicher

Free Probability and Random Matrices

FIELDS

Springer

Fields Institute Monographs

VOLUME 35

The Fields Institute for Research in Mathematical Sciences

The Fields Institute is a centre for research in the mathematical sciences, located in Toronto, Canada. The Institutes mission is to advance global mathematical activity in the areas of research, education and innovation. The Fields Institute is supported by the Ontario Ministry of Training, Colleges and Universities, the Natural Sciences and Engineering Research Council of Canada, and seven Principal Sponsoring Universities in Ontario (Carleton, McMaster, Ottawa, Queen's, Toronto, Waterloo, Western and York), as well as by a growing list of Affiliate Universities in Canada, the U.S. and Europe, and several commercial and industrial partners.

More information about this series at http://www.springer.com/series/10502

James A. Mingo • Roland Speicher

Free Probability and Random Matrices

FIELDS The Fields Institute for Research in the Mathematical Sciences

Springer

James A. Mingo
Department of Mathematics and Statistics
Queen's University
Kingston, ON, Canada

Roland Speicher
FB Mathematik
Universität des Saarlandes
Saarbrücken, Saarland, Germany

ISSN 1069-5273 ISSN 2194-3079 (electronic)
Fields Institute Monographs
ISBN 978-1-4939-6941-8 ISBN 978-1-4939-6942-5 (eBook)
DOI 10.1007/978-1-4939-6942-5

Library of Congress Control Number: 2017930940

Mathematics Subject Classification (2010): 05A15, 15B52, 46L54, 60B20

Cover illustration: Drawing of J.C. Fields by Keith Yeomans

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer Science+Business Media LLC
The registered company address is: 233 Spring Street, New York, NY 10013, U.S.A.

Dedicated to Dan-Virgil Voiculescu,
who gave us freeness, and Betina and Jill,
who gave us the liberty to work on it.

Contents

Introduction

This book is an invitation to the world of free probability theory.

Free probability is a quite young mathematical theory with many avatars. It owes its existence to the visions of one man, Dan-Virgil Voiculescu, who created it out of nothing at the beginning of the 1980s and pushed it forward ever since. The subject had a relatively slow start in its first decade but took on a lot of momentum later on.

It started in the theory of operator algebras, showed its beautiful combinatorial structure via non-crossing partitions, made contact with the world of random matrices, and reached out to many other subjects like representation theory of large groups, quantum groups, invariant subspace problem, large deviations, quantum information theory, subfactors, or statistical inference. Even in physics and engineering, many people have heard of it and find it useful and exciting.

One of us (RS) has already written, jointly with Alexandru Nica, a monograph [137] on the combinatorial side of free probability theory. Whereas combinatorics will also show up in the present book, our intention here is different: we want to give a flavour of the breadth of the subject; hence this book will cover a variety of different facets, occurrences, and applications of free probability; instead of going in depth in one of them, our goal is to give the basic ideas and results in many different directions and show how all this is related.

This means that we have to cover subjects as different as random matrices and von Neumann algebras. This should, however, not to be considered a peril but as a promise for the good things to come.

We have tried to make the book accessible to both random matrix and operator algebra (and many more) communities without requiring too many prerequisites. Whereas our presentation of random matrices should be mostly self-contained, on the operator algebraic side, we try to avoid the technical parts of the theory as much as possible. We hope that the main ideas about von Neumann algebras are comprehensible even without knowing what a von Neumann algebra is. In particular, in Chapters 1–5, no von Neumann algebras will make their appearance.

The book is a mixture between textbook and research monograph. We actually cover many of the important developments of the subject in recent years, for which no coherent introduction in monograph style has existed up to now.

Chapters 1, 2, 3, 4, and 6 describe in a self-contained way the by now well-established basic body of the theory of free probability. Chapters 1 and 4 deal with the relation of free probability with random matrices; Chapter 1 is more of a motivating nature, whereas Chapter 4 provides the rigorous results. Chapter 6 provides the relation to operator algebras and the free group factor isomorphism problem, which initiated the whole theory. Chapter 2 presents the combinatorial side of the theory; as this is dealt with in much more detail in the monograph [137], we sometimes refer to the latter for details. Chapter 3 gives a quite extensive and self-contained account of the analytic theory of free convolution. We put there quite some emphasis on the subordination formulation, which is the modern state of the art for dealing with such questions and which cannot be found in this form anywhere else.

The other chapters deal with parts of the theory where the final word is not yet spoken, but where important progress has been achieved and which surely will survive in one or the other form in future versions of free probability. In those chapters, we often make references to the original literature for details of the proofs. Nevertheless we try also there to provide intuition and motivation for what and why. We hope that those chapters invite also some of the readers to do original work in the subject.

Chapter 5 is on second order freeness; this theory intends to deal with fluctuations of random matrices in the same way as freeness does this with the average. Whereas the combinatorial aspect of this theory is far evolved, the analytic status awaits a better understanding.

Free entropy has at the moment two incarnations with very different flavour. The microstates approach is treated in Chapter 7, whereas the non-microstates approach is in Chapter 8. Both approaches have many interesting and deep results and applications—however, the final form of free entropy theory (hoping that there is only one) still has to be found.

Operator-valued free probability has evolved in recent years into a very powerful generalization of free probability theory; this is made clear by its applicability to much bigger classes of random matrices and by its use for calculating the distribution of polynomials in free variables. The operator-valued theory is developed and its use demonstrated in Chapters 9 and 10.

In Chapter 11, we present the Brown measure, a generalization of the spectral distribution from the normal to the non-normal case. In particular, we show how free probability (in its operator-valued version) allows one to calculate such Brown measures. Again there is a relation with random matrices; the Brown measure is the canonical candidate for the eigenvalue distribution of non-normal random matrix models (where the eigenvalues are not real, but complex).

After having claimed to cover many of the important directions of free probability, we have now to admit that there are at least as many which unfortunately did not make it into the book. One reason for this is that free probability is still very fast evolving with new connections popping up quite unexpectedly.

So we are, for example, not addressing such exciting topics as free stochastic and Malliavin calculus [39, 108, 114], or the rectangular version of free probability [28], or the strong version of asymptotic freeness [48, 58, 88], or free monotone transport [83], or the relation with representation theory [35, 72] or with quantum groups [16, 17, 44, 73, 110, 118, 148], or the quite recent new developments around bifreeness [52, 81, 100, 196], traffic freeness [50, 122], or connections to Ramanujan graphs via finite free convolution [124]. Instead of trying to add more chapters to a never-ending (and never-published) book, we prefer just to stop where we are and leave the remaining parts for others.

We want to emphasize that some of the results in this book owe their existence to the book writing itself and our endeavour to fill apparent gaps in the existing theory. Examples of this are our proof of the asymptotic freeness of Wigner matrices from deterministic matrices in Section 4.4 (for which there exists now also another proof in the book [7]), the fact that finite free Fisher information implies the existence of a density in Proposition 8.18, or the results about the absence of algebraic relations and zero divisors in the case of finite free Fisher information in Theorems 8.13 and 8.32.

Our presentation benefited a lot from input by others. In particular, we like to mention Serban Belinschi and Hari Bercovici for providing us with a proof of Proposition 8.18 and Uffe Haagerup for allowing us to use his manuscript of his talk at the Fields Institute as the basis for Chapter 11. With the exception of Sections 11.9 and 11.10, we are mainly following his notes in Chapter 11. Chapter 3 relied substantially on input and feedback from the experts on the subject. Many of the results and proofs around subordination were explained to us by Serban Belinschi, and we also got a lot of feedback from JC Wang and John Williams. We are also grateful to N. Raj Rao for help with his RMTool package which was used in our numerical simulations.

The whole idea of writing this book started from a lectures series on free probability and random matrices which we gave at the Fields Institute, Toronto, in the fall of 2007 within the Thematic Program on Operator Algebras. Notes of our lectures were taken by Emily Redelmeier and by Jonathan Novak, and the first draft of the book was based on these notes.

We had the good fortune to have Uffe Haagerup around during this programme, and he agreed to give one of the lectures, on his work on the Brown measure. As mentioned above, the notes of his lecture became the basis of Chapter 11.

What are now Chapters 5, 8, 9, and 10 were not part of the lectures at the Fields Institute, but were added later. Those additional chapters cover in big parts also results which did not yet exist in 2007. So this gives us at least some kind of excuse that the finishing of the book took so long.

Much of Chapter 8 is based on classes on “Random matrices and free entropy” and “Non-commutative distributions” which one of us (RS) taught at Saarland University during the winter terms 2013/2014 and 2014/2015, respectively. The final outcome of this chapter owes a lot to the support of Tobias Mai for those classes.

Chapter 9 is based on work of RS with Wlodek Bryc, Reza Rashidi Far, and Tamer Oraby on block random matrices in a wireless communications (MIMO) context and on various lectures of RS for engineering audiences, where he tried to convince them of the relevance and usefulness of operator-valued methods in wireless problems. Chapter 10 benefited a lot from the work of Carlos Vargas on free deterministic equivalents in his PhD thesis and from the joint work of RS with Serban Belinschi and Tobias Mai around linearization and the analytic theory of operator-valued free probability. The algorithms, numerical simulations, and histograms for eigenvalue distributions in Chapter 10 and Brown measures in Chapter 11 are done with great expertise and dedication by Tobias Mai.

There are exercises scattered throughout the text. The intention is to give readers an opportunity to test their understanding. In some cases, where the result is used in a crucial way or where the calculation illustrates basic ideas, a solution is provided at the end of the book.

In addition to the already mentioned individuals, we owe a lot of thanks to people who read preliminary versions of the book and gave useful feedback, which helped to improve the presentation and correct some mistakes. We want to mention in particular Marwa Banna, Arup Bose, Mario Diaz, Yinzheng Gu, Todd Kemp, Felix Leid, Josué Vázquez, Hao-Wei Wang, and Guangqu Zheng.

Further thanks are due to the Fields Institute for the great environment they offered us during the already mentioned thematic programme on operator algebras and for the opportunity to publish our work in their Monographs series. The writing of this book, as well as many of the reported results, would not have been possible without financial support from various sources; in particular, we want to mention a Killam Fellowship for RS in 2007 and 2008, which allowed him to participate in the thematic programme at the Fields Institute and thus get the whole project started, and the ERC Advanced Grant "Non-commutative distributions in free probability" of RS, which provided time and resources for the finishing of the project. Many of the results we report here were supported by grants from the Canadian and German Science Foundations NSERC and DFG, respectively, by Humboldt Fellowships for Serban Belinschi and John Williams for stays at Saarland University, and by DAAD German-French Procope exchange programmes between Saarland University and the Universities of Besançon and of Toulouse.

As we are covering a wide range of topics, there might come a point where one gets a bit exhausted from our book. There are, however, some alternatives, like the standard references [97, 137, 197, 198] or survey articles [37, 84, 141, 142, 156, 162, 164, 165, 183, 191, 192] on (some aspects of) free probability. Our advice: take a break, enjoy those, and then come back motivated to learn more from our book.

Chapter 1
Asymptotic Freeness of Gaussian Random Matrices

In this chapter we shall introduce a principal object of study: Gaussian random matrices. This is one of the few ensembles of random matrices for which one can do explicit calculations of the eigenvalue distribution. For this reason the Gaussian ensemble is one of the best understood. Information about the distribution of the eigenvalues is carried by it moments: $\{\mathrm{E}(\mathrm{tr}(X^k))\}_k$ where E is the expectation, tr denotes the normalized trace (i.e. $\mathrm{tr}(I_N) = 1$), and X is an $N \times N$ random matrix.

One of the achievements of the free probability approach to random matrices is to isolate the property called asymptotic freeness. If X and Y are asymptotically free, then we can approximate the moments of $X + Y$ and XY from the moments of X and Y; moreover this approximation becomes exact in the large N limit. In its exact form, this relation is called freeness, and we shall give its definition at the end of this chapter, §1.12. In Chapter 2 we shall explore the basic properties of freeness and relate these to some new objects called free cumulants. To motivate the definition of freeness, we shall show in this chapter that independent Gaussian random matrices are asymptotically free, thus showing that freeness arises naturally.

To begin this chapter, we shall review enough of the elementary properties of Gaussian random variables to demonstrate asymptotic freeness.

We want to add right away the disclaimer that we do not attempt to give a comprehensive introduction into the vast subject of random matrices. We concentrate on aspects which have some relevance for free probability theory; still this should give the uninitiated reader a good idea what random matrices are and why they are so fascinating and have become a centrepiece of modern mathematics. For more on its diversity, beauty, and depth, one should have a look at [7, 69, 171] or on the collection of survey articles on various aspects of random matrices in [2].

J.A. Mingo, R. Speicher, *Free Probability and Random Matrices*,
Fields Institute Monographs 35, DOI 10.1007/978-1-4939-6942-5_1

1.1 Moments and cumulants of random variables

Let ν be a probability measure on $\mathbb{R}$. If $\int_{\mathbb{R}} |t|^n \, d\nu(t) < \infty$, we say that ν has a moment of order n, and the n^{th} moment is denoted $\alpha_n = \int_{\mathbb{R}} t^n \, d\nu(t)$.

Exercise 1. If ν has a moment of order n, then ν has all moments of order m for $m < n$.

The integral $\varphi(t) = \int e^{ist} \, d\nu(s)$ (with $i = \sqrt{-1}$) is always convergent and is called the *characteristic function* of ν. It is always uniformly continuous on $\mathbb{R}$ and $\varphi(0) = 1$, so for $|t|$ small enough $\varphi(t) \notin (-\infty, 0]$ and we can define the continuous function $\log(\varphi(t))$. If ν has a moment of order n, then φ has a derivative of order n, and conversely if φ has a derivative of order n, then ν has a moment of order n when n is even and a moment of order $n-1$ when n is odd (see Lukacs [119, Corollary 1 to Theorem 2.3.1]). Moreover $\alpha_n = i^{-n}\varphi^{(n)}(0)$, so if φ has a power series expansion, it has to be

$$\varphi(t) = \sum_{n \geq 0} \alpha_n \frac{(it)^n}{n!}.$$

Thus if ν has a moment of order $m+1$, we can write

$$\log(\varphi(t)) = \sum_{n=1}^{m} k_n \frac{(it)^n}{n!} + o(t^m) \qquad \text{with} \qquad k_n = i^{-n} \frac{d^n}{dt^n} \log(\varphi(t)) \Big|_{t=0}.$$

The numbers $\{k_n\}_n$ are the *cumulants* of ν. To distinguish them from the free cumulants, which will be defined in the next chapter, we will call $\{k_n\}_n$ the *classical* cumulants of ν. The moments $\{\alpha_n\}_n$ of ν and the cumulants $\{k_n\}_n$ of ν each determine the other through the *moment-cumulant formulas*:

$$\alpha_n = \sum_{\substack{1\cdot r_1 + \cdots + n \cdot r_n = n \\ r_1, \ldots, r_n \geq 0}} \frac{n!}{(1!)^{r_1} \cdots (n!)^{r_n} r_1! \cdots r_n!} k_1^{r_1} \cdots k_n^{r_n} \tag{1.1}$$

$$k_n = \sum_{\substack{1\cdot r_1 + \cdots + n \cdot r_n = n \\ r_1, \ldots, r_n \geq 0}} \frac{(-1)^{r_1 + \cdots + r_n - 1}(r_1 + \cdots + r_n - 1)! \, n!}{(1!)^{r_1} \cdots (n!)^{r_n} r_1! \cdots r_n!} \alpha_1^{r_1} \cdots \alpha_n^{r_n}. \tag{1.2}$$

Both sums are over all non-negative integers $r_1, \ldots, r_n$ such that $1 \cdot r_1 + \cdots + n \cdot r_n = n$. We shall see below in Exercises 4 and 12 how to use partitions to simplify these formidable equations.

A very important random variable is the *Gaussian* or *normal* random variable. It has the distribution

$$P(t_1 \leq X \leq t_2) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{t_1}^{t_2} e^{-(t-a)^2/(2\sigma^2)} \, dt$$

where a is the mean and σ^2 is the variance. The characteristic function of a Gaussian random variable is

$$\varphi(t) = \exp\left(iat - \frac{\sigma^2 t^2}{2}\right), \qquad \text{thus} \qquad \log\varphi(t) = a\frac{(it)^1}{1!} + \sigma^2\frac{(it)^2}{2!}.$$

Hence for a Gaussian random variable, all cumulants beyond the second are 0.

Exercise 2. Suppose ν has a fifth moment and we write

$$\varphi(t) = 1 + \alpha_1\frac{(it)}{1!} + \alpha_2\frac{(it)^2}{2!} + \alpha_3\frac{(it)^3}{3!} + \alpha_4\frac{(it)^4}{4!} + o(t^4)$$

where $\alpha_1, \alpha_2, \alpha_3$, and α_4 are the first four moments of ν. Let

$$\log(\varphi(t)) = k_1\frac{(it)}{1!} + k_2\frac{(it)^2}{2!} + k_3\frac{(it)^3}{3!} + k_4\frac{(it)^4}{4!} + o(t^4).$$

Using the Taylor series for $\log(1+x)$, find a formula for $\alpha_1, \alpha_2, \alpha_3$, and α_4 in terms of k_1, k_2, k_3, and k_4.

1.2 Moments of a Gaussian random variable

Let X be a Gaussian random variable with mean 0 and variance 1. Then by definition

$$P(t_1 \le X \le t_2) = \frac{1}{\sqrt{2\pi}}\int_{t_1}^{t_2} e^{-t^2/2}dt.$$

Let us find the moments of X. Clearly, $\alpha_0 = 1$, $\alpha_1 = 0$, and by integration by parts

$$\alpha_n = \mathrm{E}(X^n) = \int_{\mathbb{R}} t^n e^{-t^2/2}\frac{dt}{\sqrt{2\pi}} = (n-1)\alpha_{n-2} \quad \text{for } n \ge 2.$$

Thus

$$\alpha_{2n} = (2n-1)(2n-3)\cdots 5\cdot 3\cdot 1 =: (2n-1)!!$$

and $\alpha_{2n-1} = 0$ for all n.

Let us find a combinatorial interpretation of these numbers. For a positive integer n, let $[n] = \{1, 2, 3, \ldots, n\}$, and $\mathcal{P}(n)$ denote all *partitions* of the set $[n]$, i.e. $\pi = \{V_1, \ldots, V_k\} \in \mathcal{P}(n)$ means $V_1, \ldots, V_k \subseteq [n]$, $V_i \neq \emptyset$ for all i, $V_1 \cup \cdots \cup V_k = [n]$, $V_i \cap V_j = \emptyset$ for $i \neq j$; $V_1, \ldots, V_k$ are called the *blocks* of π. We let $\#(\pi)$ denote the number of blocks of π and $\#(V_i)$ the number of elements in the block V_i. A partition is a *pairing* if each block has size 2. The pairings of $[n]$ will be denoted $\mathcal{P}_2(n)$.

Let us count $|\mathcal{P}_2(2n)|$, the number of pairings of $[2n]$. 1 must be paired with something and there are $2n-1$ ways of choosing it. Thus

$$|\mathcal{P}_2(n)| = (2n-1)|\mathcal{P}_2(n-2)| = (2n-1)!!.$$

So $\mathrm{E}(X^{2n}) = |\mathcal{P}_2(2n)|$. There is a deeper connection between moments and partitions known as Wick's formula (see Section 1.5).

Exercise 3. We say that a partition of $[n]$ has *type* $(r_1, \ldots, r_n)$ if it has r_i blocks of size i. Show that the number of partitions of $[n]$ of type $(r_1, r_2, \ldots, r_n)$ is

$$\frac{n!}{(1!)^{r_1}(2!)^{r_2}\cdots(n!)^{r_n} r_1! r_2! \cdots r_n!}.$$

Using the type of a partition, there is a very simple expression for the moment-cumulant relations above. Moreover this expression is quite amenable for calculation. If π is a partition of $[n]$ and $\{k_i\}_i$ is any sequence, let $k_\pi = k_1^{r_1} k_2^{r_2} \cdots k_n^{r_n}$ where r_i is the number of blocks of π of size i. Using this notation the first of the moment-cumulant relations can be written

$$\alpha_n = \sum_{\pi \in \mathcal{P}(n)} k_\pi. \tag{1.3}$$

The second moment-cumulant relation can be written (see Exercise 13)

$$k_n = \sum_{\pi \in \mathcal{P}(n)} (-1)^{\#(\pi)-1} (\#(\pi) - 1)!\, \alpha_\pi. \tag{1.4}$$

The simplest way to do calculations with relations like those above is to use formal power series (see Stanley [167, §1.1]).

Exercise 4. Let $\{\alpha_n\}$ and $\{k_n\}$ be two sequences satisfying (1.3). In this exercise we shall show that as formal power series

$$\log\left(1 + \sum_{n=1}^{\infty} \alpha_n \frac{z^n}{n!}\right) = \sum_{n=1}^{\infty} k_n \frac{z^n}{n!}. \tag{1.5}$$

(i) Show that by differentiating both sides of (1.5) it suffices to prove

$$\sum_{n=0}^{\infty} \alpha_{n+1} \frac{z^n}{n!} = \left(1 + \sum_{n=1}^{\infty} \alpha_n \frac{z^n}{n!}\right) \sum_{n=0}^{\infty} k_{n+1} \frac{z^n}{n!}. \tag{1.6}$$

(ii) By grouping the terms in $\sum_\pi k_\pi$ according to the size of the block containing 1, show that

$$\alpha_n = \sum_{\pi \in \mathcal{P}(n)} k_\pi = \sum_{m=0}^{n-1} \binom{n-1}{m} k_{m+1} \alpha_{n-m-1}.$$

(iii) Use the result of (ii) to prove (1.6).

1.3 Gaussian vectors

Let $X : \Omega \to \mathbb{R}^n$, $X = (X_1, \ldots, X_n)$ be a random vector. We say that X is Gaussian if there is a positive definite $n \times n$ real symmetric matrix B such that

$$\mathrm{E}(X_{i_1} \cdots X_{i_k}) = \int_{\mathbb{R}^n} t_{i_1} \cdots t_{i_k} \frac{\exp(-\langle Bt, t\rangle/2)dt}{(2\pi)^{n/2}\det(B)^{-1/2}}$$

where $\langle \cdot, \cdot \rangle$ denotes the standard inner product on $\mathbb{R}^n$. Let $C = (c_{ij})$ be the covariance matrix, that is, $c_{ij} = \mathrm{E}\big([X_i - \mathrm{E}(X_i)] \cdot [X_j - \mathrm{E}(X_j)]\big)$.

In fact $C = B^{-1}$, and if $X_1, \ldots, X_n$ are independent, then B is a diagonal matrix; see Exercise 5. If $Y_1, \ldots, Y_n$ are independent Gaussian random variables, A is an invertible real matrix, and $X = AY$, then X is a Gaussian random vector, and every Gaussian random vector is obtained in this way. If $X = (X_1, \ldots, X_n)$ is a complex random vector, we say that X is a complex Gaussian random vector if $(\mathrm{Re}(X_1), \mathrm{Im}(X_1), \ldots, \mathrm{Re}(X_n), \mathrm{Im}(X_n))$ is a real Gaussian random vector.

Exercise 5. Let $X = (X_1, \ldots, X_n)$ be a Gaussian random vector with density $\sqrt{\det(B)(2\pi)^{-n}}\ \exp(-\langle Bt, t\rangle/2)$. Let $C = (c_{ij}) = B^{-1}$.

(i) Show that B is diagonal if and only if $\{X_1, \ldots, X_n\}$ are independent.
(ii) By first diagonalizing B, show that $c_{ij} = \mathrm{E}\big([X_i - \mathrm{E}(X_i)] \cdot [X_j - \mathrm{E}(X_j)]\big)$.

1.4 The moments of a standard complex Gaussian random variable

Suppose X and Y are independent real Gaussian random variables with mean 0 and variance 1. Then $Z = (X + iY)/\sqrt{2}$ is a complex Gaussian random variable with mean 0 and variance $\mathrm{E}(Z\overline{Z}) = \frac{1}{2}\mathrm{E}(X^2 + Y^2) = 1$. We call Z a *standard complex Gaussian random variable*. Moreover, for such a complex Gaussian variable, we have

$$\mathrm{E}(Z^m\overline{Z}^n) = \begin{cases} 0, & m \neq n \\ m!, & m = n \end{cases}.$$

Exercise 6. Let $Z = (X + iY)/\sqrt{2}$ be a standard complex Gaussian random variable with mean 0 and variance 1.

(i) Show that

$$\mathrm{E}(Z^m\overline{Z}^n) = \frac{1}{\pi}\int_{\mathbb{R}^2} (t_1 + it_2)^m (t_1 - it_2)^n e^{-(t_1^2+t_2^2)} dt_1 dt_2.$$

By switching to polar coordinates, show that

$$\mathrm{E}(Z^m \overline{Z}^n) = \frac{1}{\pi} \int_0^{2\pi} \int_0^{\infty} r^{m+n+1} e^{i\theta(m-n)} e^{-r^2}\, dr d\theta.$$

(*ii*) Show that $\mathrm{E}(Z^m \overline{Z}^n) = 0$ for $m \neq n$ and that $\mathrm{E}(|Z|^{2n}) = n!$.

1.5 Wick's formula

Let $(X_1, \ldots X_n)$ be a real Gaussian random vector and $i_1, \ldots, i_k \in [n]$. Wick's formula gives a simple expression for $\mathrm{E}(X_{i_1} \cdots X_{i_k})$. If k is even and $\pi \in \mathcal{P}_2(k)$, let $\mathrm{E}_\pi(X_1, \ldots, X_k) = \prod_{(r,s)\in\pi} \mathrm{E}(X_r X_s)$. For example, if $\pi = \{(1,3)(2,6)(4,5)\}$, then $\mathrm{E}_\pi(X_1, X_2, X_3, X_4, X_5, X_6) = \mathrm{E}(X_1X_3)\mathrm{E}(X_2X_6)\mathrm{E}(X_4X_5)$. E_π is a k-linear functional. The fact that only pairings arise in Wick's formula is a consequence of the observation on page 3 that for a Gaussian random variable, all cumulants above the second vanish.

Theorem 1. *Let $(X_1, \ldots, X_n)$ be a real Gaussian random vector. Then*

$$\mathrm{E}(X_{i_1} \cdots X_{i_k}) = \sum_{\pi\in\mathcal{P}_2(k)} \mathrm{E}_\pi(X_{i_1}, \ldots, X_{i_k}) \quad \textit{for any } i_1, \ldots, i_k \in [n]. \tag{1.7}$$

Proof: Suppose that the covariance matrix C of $(X_1, \ldots, X_n)$ is diagonal, i.e. the X_i's are independent. Consider $(i_1, \ldots, i_k)$ as a function $[k] \to [n]$. Let $\{a_1, \ldots, a_r\}$ be the range of i and $A_j = i^{-1}(a_j)$. Then $\{A_1, \ldots, A_r\}$ is a partition of $[k]$ which we denote $\ker(i)$. Let $|A_t|$ be the number of elements in A_t. Then $\mathrm{E}(X_{i_1} \cdots X_{i_k}) = \prod_{t=1}^r \mathrm{E}(X_{a_t}^{|A_t|})$. Let us recall that if X is a real Gaussian random variable of mean 0 and variance c, then for k even $\mathrm{E}(X^k) = c^{k/2} \times |\mathcal{P}_2(k)| = \sum_{\pi\in\mathcal{P}_2(k)} \mathrm{E}_\pi(X, \ldots, X)$ and for k odd $\mathrm{E}(X^k) = 0$. Thus we can write the product $\prod_t \mathrm{E}(X_{a_t}^{|A_t|})$ as a sum $\sum_{\pi\in\mathcal{P}_2(k)} \mathrm{E}_\pi(X_{i_1}, \ldots, X_{i_k})$ where the sum runs over all π's which only connect elements in the same block of $\ker(i)$. Since $\mathrm{E}(X_{i_r} X_{i_s}) = 0$ for $i_r \neq i_s$, we can relax the condition that π only connect elements in the same block of $\ker(i)$. Hence $\mathrm{E}(X_{i_1} \cdots X_{i_k}) = \sum_{\pi\in\mathcal{P}_2(k)} \mathrm{E}_\pi(X_{i_1}, \ldots, X_{i_k})$.

Finally let us suppose that C is arbitrary. Let the density of $(X_1, \ldots, X_n)$ be $\exp(-\langle Bt, t\rangle/2)[(2\pi)^{n/2} \det(B)^{-1/2}]^{-1}$ and choose an orthogonal matrix O such that $D = O^{-1}BO$ is diagonal. Let

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = O^{-1} \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}.$$

Then $(Y_1, \ldots, Y_n)$ is a real Gaussian random vector with the diagonal covariance matrix D^{-1}. Then

$$\begin{aligned}
\mathrm{E}(X_{i_1} \cdots X_{i_k}) &= \sum_{j_1,\ldots,j_k=1}^{n} o_{i_1 j_1} o_{i_2 j_2} \cdots o_{i_k j_k} \mathrm{E}(Y_{j_1} Y_{j_2} \cdots Y_{j_k}) \\
&= \sum_{j_1,\ldots,j_k=1}^{n} o_{i_1 j_1} \cdots o_{i_k j_k} \sum_{\pi \in \mathcal{P}_2(k)} \mathrm{E}_\pi(Y_{j_1}, \ldots, Y_{j_k}) \\
&= \sum_{\pi \in \mathcal{P}_2(k)} \mathrm{E}_\pi(X_{i_1}, \ldots, X_{i_k}).
\end{aligned}$$

□

Since both sides of equation (1.7) are k-linear, we can extend by linearity to the complex case.

Corollary 2. *Suppose $(X_1, \ldots, X_n)$ is a complex Gaussian random vector; then*

$$\mathrm{E}(X_{i_1}^{(\varepsilon_1)} \cdots X_{i_k}^{(\varepsilon_k)}) = \sum_{\pi \in \mathcal{P}_2(k)} \mathrm{E}_\pi(X_{i_1}^{(\varepsilon_1)}, \ldots, X_{i_k}^{(\varepsilon_k)}) \tag{1.8}$$

for all $i_1, \ldots, i_k \in [n]$ and all $\varepsilon_1, \ldots, \varepsilon_k \in \{0, 1\}$, where we have used the notation $X_i^{(0)} := X_i$ and $X_i^{(1)} := \overline{X_i}$.

Formulas (1.7) and (1.8) are usually referred to as *Wick's formula* after the physicist Gian-Carlo Wick [200], who introduced them in 1950 as a fundamental tool in quantum field theory; one should notice, though, that they had already appeared much earlier, in 1918, in the work of the statistician Leon Isserlis [101].

Exercise 7. Let $Z_1, \ldots, Z_s$ be independent standard complex Gaussian random variables with mean 0 and $\mathrm{E}(|Z_i|^2) = 1$. Show that

$$\mathrm{E}(Z_{i_1} \cdots Z_{i_n} \overline{Z_{j_1}} \cdots \overline{Z_{j_n}}) = |\{\sigma \in S_n \mid i = j \circ \sigma\}|.$$

S_n denotes the symmetric group on $[n]$. Note that this is consistent with part (*iii*) of Exercise 6.

1.6 Gaussian random matrices

Let X be an $N \times N$ matrix with entries f_{ij} where $f_{ij} = x_{ij} + \sqrt{-1}\, y_{ij}$ is a complex Gaussian random variable normalized such that $\sqrt{N} f_{ij}$ is a standard complex Gaussian random variable, i.e. $\mathrm{E}(f_{ij}) = 0$, $\mathrm{E}(|f_{ij}|^2) = 1/N$ and

(*i*) $f_{ij} = \overline{f_{ji}}$,
(*ii*) $\{x_{ij}\}_{i \geq j} \cup \{y_{ij}\}_{i > j}$ are independent.

Then X is a *self-adjoint Gaussian random matrix*. Such a random matrix is often called a GUE random matrix (GUE = Gaussian unitary ensemble).

Exercise 8. Let X be an $N \times N$ GUE random matrix, with entries $f_{ij} = x_{ij} + \sqrt{-1}\, y_{ij}$ normalized so that $\mathrm{E}(|f_{ij}|^2) = 1/N$.

(i) Consider the random N^2-vector

$$(x_{11}, \dots, x_{NN}, x_{12}, \dots, x_{1N}, \dots, x_{N-1,N}, y_{12}, \dots, y_{N-1,N}).$$

Show that the density of this vector is $c\, e^{-N\mathrm{Tr}(X^2)/2} dX$ where c is a constant and $dX = \prod_{i=1}^{N} dx_{ii} \prod_{i<j} dx_{ij}\, dy_{ij}$ is Lebesgue measure on $\mathbb{R}^{N^2}$.

(ii) Evaluate the constant c.

1.7 A genus expansion for the GUE

Let us calculate $\mathrm{E}(\mathrm{Tr}(Y^k))$, for $Y = (g_{ij})$ a $N \times N$ GUE random matrix. We first suppose for convenience that the entries of Y have been normalized so that $\mathrm{E}(|g_{ij}|^2) = 1$. Now

$$\mathrm{E}(\mathrm{Tr}(Y^k)) = \sum_{i_1,\dots,i_k=1}^{N} \mathrm{E}(g_{i_1 i_2} g_{i_2 i_3} \cdots g_{i_k i_1}).$$

By Wick's formula (1.8), $\mathrm{E}(g_{i_1 i_2} g_{i_2 i_3} \cdots g_{i_k i_1}) = 0$ whenever k is odd, and otherwise

$$\mathrm{E}(g_{i_1 i_2} g_{i_2 i_3} \cdots g_{i_{2k} i_1}) = \sum_{\pi \in \mathcal{P}_2(2k)} \mathrm{E}_\pi(g_{i_1 i_2}, g_{i_2 i_3}, \dots, g_{i_{2k} i_1}).$$

Now $\mathrm{E}(g_{i_r i_{r+1}} g_{i_s i_{s+1}})$ will be 0 unless $i_r = i_{s+1}$ and $i_s = i_{r+1}$ (using the convention that $i_{2k+1} = i_1$). If $i_r = i_{s+1}$ and $i_s = i_{r+1}$, then $\mathrm{E}(g_{i_r i_{r+1}} g_{i_s i_{s+1}}) = \mathrm{E}(|g_{i_r i_{r+1}}|^2) = 1$. Thus given $(i_1, \dots, i_{2k})$, $\mathrm{E}(g_{i_1 i_2} g_{i_2 i_3} \cdots g_{i_{2k} i_1})$ will be the number of pairings π of $[2k]$ such that for each pair (r, s) of π, $i_r = i_{s+1}$ and $i_s = i_{r+1}$.

In order to easily count these, we introduce the following notation. We regard the $2k$-tuple $(i_1, \dots, i_{2k})$ as a function $i : [2k] \to [N]$. A pairing $\pi = \{(r_1, s_1)(r_2, s_2), \dots, (r_k, s_k)\}$ of $[2k]$ will be regarded as a permutation of $[2k]$ by letting (r_i, s_i) be the transposition that switches r_i with s_i and $\pi = (r_1, s_1) \cdots (r_k, s_k)$ as the product of these transpositions. We also let γ_{2k} be the permutation of $[2k]$ which has the one cycle $(1, 2, 3, \dots, 2k)$. With this notation our condition on the pairings has a simple expression. Let π be a pairing of $[2k]$ and (r, s) be a pair of π. The condition $i_r = i_{s+1}$ can be written as $i(r) = i(\gamma_{2k}(\pi(r)))$ since $\pi(r) = s$ and $\gamma_{2k}(\pi(r)) = s + 1$. Thus $\mathrm{E}_\pi(g_{i_1 i_2}, g_{i_2 i_3}, \dots, g_{i_{2k} i_1})$ will be 1 if i is constant on the orbits of $\gamma_{2k}\pi$ and 0 otherwise. For a permutation σ, let $\#(\sigma)$ denote the number of cycles of σ. Thus

$$
\begin{aligned}
\mathrm{E}(\mathrm{Tr}(Y^{2k})) &= \sum_{i_1,\ldots,i_{2k}=1}^{N} \left| \left\{ \pi \in \mathcal{P}_2(2k) \;\middle|\; \begin{array}{l} i \text{ is constant on the} \\ \text{orbits of } \gamma_{2k}\pi \end{array} \right\} \right| \\
&= \sum_{\pi \in \mathcal{P}_2(2k)} \left| \left\{ i : [2k] \to [N] \;\middle|\; \begin{array}{l} i \text{ is constant on the} \\ \text{orbits of } \gamma_{2k}\pi \end{array} \right\} \right| \\
&= \sum_{\pi \in \mathcal{P}_2(2k)} N^{\#(\gamma_{2k}\pi)}.
\end{aligned}
$$

We summarize this result in the statement of the following theorem.

Theorem 3. *Let* $Y_N = (g_{ij})$ *be a* $N \times N$ GUE *random matrix with entries normalized so that* $\mathrm{E}(|g_{ij}|^2) = 1$ *for all* i *and* j. *Then*

$$\mathrm{E}(\mathrm{Tr}(Y_N^{2k})) = \sum_{\pi \in \mathcal{P}_2(2k)} N^{\#(\gamma_{2k}\pi)}.$$

Moreover, for $X_N = N^{-1/2} Y_N = (f_{ij})$, *with the normalization* $\mathrm{E}(|f_{ij}|^2) = 1/N$, *we have*

$$\mathrm{E}(\mathrm{tr}(X_N^{2k})) = \sum_{\pi \in \mathcal{P}_2(2k)} N^{\#(\gamma_{2k}\pi)-k-1}.$$

Here, Tr *denotes the usual, unnormalized trace, whereas* $\mathrm{tr} = \frac{1}{N}\mathrm{Tr}$ *is the normalized trace.*

The expansion in this theorem is usually addressed as *genus expansion*. In the next section, we will elaborate more on this.

In the mathematical literature, this genus expansion appeared for the first time in the work of Harer and Zagier [91] in 1986 but was mostly overlooked for a while, until random matrices became mainstream also in mathematics in the new millennium; in physics, on the other side, such expansions were kind of folklore and at the basis of Feynman diagram calculations in quantum field theory; see, for example, [45, 172, 207].

1.8 Non-crossing partitions and permutations

In order to find the limit of $\mathrm{E}(\mathrm{tr}(X_N^{2k}))$, we have to understand the sign of the quantity $\#(\gamma_{2k}\pi) - k - 1$. We shall show that for all pairings $\#(\gamma_{2k}\pi) - k - 1 \leq 0$ and identify the pairings for which we have equality. As we shall see that the π's for which we have equality are the non-crossing pairings, let us begin by reviewing some material on non-crossing partitions from [137, Lecture 9].

Let π be a partition of $[n]$. If we can find $i < j < k < l$ such that i and k are in one block V of π and j and l are in another block W of π, we say that V and W *cross*. If no pair of blocks of π cross, then we say π is *non-crossing*. We denote the set of non-crossing partitions of $[n]$ by $NC(n)$. The set of non-crossing pairings of $[n]$ is denoted $NC_2(n)$. We discuss this more fully in §2.2

Given a permutation $\pi \in S_n$, we consider all possible factorizations into products of transpositions. For example, we can write $(1,2,3,4) = (1,4)(1,3)(1,2) = (1,2)(1,4)(2,3)(1,4)(3,4)$. We let $|\pi|$ be the least number of transpositions needed to factor π. In the example above, $|(1,2,3,4)| = 3$. From this definition we see that $|\pi\sigma| \leq |\pi| + |\sigma|$, $|\pi^{-1}| = |\pi|$, and $|e| = 0$, that is, $|\cdot|$ is a *length function* on S_n.

There is a very simple relation between $|\pi|$ and $\#(\pi)$, namely, for $\pi \in S_n$ we have $|\pi| + \#(\pi) = n$. There is a simple observation that will be used to establish this and many other useful inequalities. Let $(i_1, \ldots, i_k)$ be a cycle of a permutation π and $1 \leq m < n \leq k$. Then $(i_1, \ldots, i_k)(i_m, i_n) = (i_1, \ldots, i_m, i_{n+1}, \ldots, i_k)(i_{m+1}, \ldots, i_n)$. From this we immediately see that if π is a permutation and $\tau = (r, s)$ is a transposition, then $\#(\pi\tau) = \#(\pi) + 1$ if r and s are in the same cycle of π and $\#(\pi\tau) = \#(\pi) - 1$ if r and s are in different cycles of π. Thus we easily deduce that for any transpositions $\tau_1, \ldots, \tau_k$ in S_n we have $\#(\tau_1 \cdots \tau_k) \geq n - k$ as, starting with the identity permutation (with n cycles), each transposition τ_i can reduce the number of cycles by at most 1. This shows that $\#(\pi) \geq n - |\pi|$. On the other hand, we have for any cycle $(i_1, \ldots, i_k) = (i_1, i_k)(i_1, i_{k-1}) \cdots (i_1, i_2)$ is the product of $k - 1$ transpositions. Thus $|\pi| \leq n - \#(\pi)$. See [137, Lecture 23] for a further discussion.

Let us return to our original problem and let π be a pairing of $[2k]$. We regard π as a permutation in S_{2k} as above. Then $\#(\pi) = k$, so $|\pi| = k$. Also $|\gamma_{2k}| = 2k - 1$. The triangle inequality gives us $|\gamma_{2k}| \leq |\pi| + |\gamma_{2k}\pi|$ (since $\pi = \pi^{-1}$) or $\#(\gamma_{2k}\pi) \leq k + 1$. This shows that $\#(\gamma_{2k}\pi) - k - 1 \leq 0$ for all pairings π. Next we have to identify for which π's we have equality. For this we use a theorem of Biane which embeds $NC(n)$ into S_n.

We let $\gamma_n = (1, 2, 3, \ldots, n)$. Let π be a partition of $[n]$. We can arrange the elements of the blocks of π in increasing order and consider these blocks to be the cycles of a permutation, also denoted π. When we regard π as a permutation, $\#(\pi)$ also denotes the number of cycles of π. Biane's result is that π is non-crossing, as a partition, if and only if, the triangle inequality $|\gamma_n| \leq |\pi| + |\pi^{-1}\gamma_n|$ becomes an equality. In terms of cycles, this means $\#(\pi) + \#(\pi^{-1}\gamma_n) \leq n + 1$ with equality only if π is non-crossing. This is a special case of a theorem which states that for π and σ, any two permutations of $[n]$ such that the subgroup generated by π and σ acts transitively on $[n]$, there is an integer $g \geq 0$ such that $\#(\pi) + \#(\pi^{-1}\sigma) + \#(\sigma) = n + 2(1 - g)$, and g is the minimal genus of a surface upon which the "graph" of π relative to σ can be embedded. See [61, Propriété II.2] and Fig. 1.1. Thus we can say that π is non-crossing with respect to σ if $|\sigma| = |\pi| + |\pi^{-1}\sigma|$. We shall need this relation in Chapter 5. An easy corollary of the equation $\#(\pi) + \#(\pi^{-1}\sigma) + \#(\sigma) = n + 2(1-g)$ is that if π is a pairing of $[2k]$ and $\#(\gamma_{2k}\pi) < k + 1$, then $\#(\gamma_{2k}\pi) < k$.

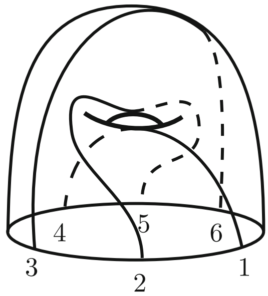


Fig. 1.1 A surface of genus 1 with the pairing $(1,4)(2,5)$ $(3,6)$ drawn on it

Example 4. $\gamma = (1,2,3,4,5,6)$, $\pi = (1,4)(2,5)(3,6)$, $\#(\pi) = 3$, $\#(\gamma_6) = 1$, $\#(\pi^{-1}\gamma_6) = 2$, $\#(\pi) + \#(\pi^{-1}\gamma_6) + \#(\gamma_6) = 6$; therefore $g = 1$.

If $g = 0$, the surface is a sphere and the graph is planar, and we say π is *planar* relative to γ. When γ has one cycle, "planar relative to γ" is what we have been calling a non-crossing partition; for a proof of Biane's theorem, see [137, Proposition 23.22].

Proposition 5. *Let $\pi \in S_n$; then $\pi \in NC(n)$ if and only if $|\pi| + |\pi^{-1}\gamma_n| = |\gamma_n|$.*

Corollary 6. *If π is a pairing of $[2k]$ then $\#(\gamma_{2k}\pi) \leq k-1$ unless π is non-crossing in which case $\#(\gamma_{2k}\pi) = k+1$.*

1.9 Wigner's semi-circle law

Consider again our GUE matrices $X_N = (f_{ij})$ with normalization $\mathrm{E}(\left|f_{ij}\right|^2) = \frac{1}{N}$. Then, by Theorem 3, we have

$$\begin{aligned}\mathrm{E}(\mathrm{tr}(X_N^{2k})) &= N^{-(k+1)} \sum_{\pi \in \mathcal{P}_2(2k)} N^{\#(\gamma_{2k}\pi)} \\ &= \sum_{\pi \in \mathcal{P}_2(2k)} N^{-2g_\pi}, \end{aligned} \tag{1.9}$$

because $\#(\pi^{-1}\gamma) = \#(\gamma\pi^{-1})$ for any permutations π and γ, and if π is a pairing, then $\pi = \pi^{-1}$. Thus $C_k := \lim_{N\to\infty} \mathrm{E}(\mathrm{tr}(X_N^{2k}))$ is the number of non-crossing pairings of $[2k]$, i.e. the cardinality of $NC_2(2k)$. It is well-known that this is the k-th Catalan number $\frac{1}{k+1}\binom{2k}{k}$ (see [137, Lemma 8.9], or (2.5) in the next chapter).

Since the Catalan numbers are the moments of the semi-circle distribution, we have arrived at Wigner's famous semi-circle law [201], which says that the spectral measures of $\{X_N\}_N$, relative to the state $\mathrm{E}(\mathrm{tr}(\cdot))$, converge to $(2\pi)^{-1}\sqrt{4-t^2}\,dt$, i.e. the expected proportion of eigenvalues of X between a and b is asymptotically $(2\pi)^{-1}\int_a^b \sqrt{4-t^2}\,dt$. See Fig. 1.2.

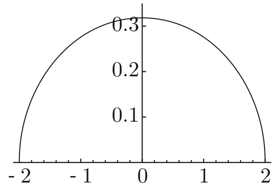


Fig. 1.2 The graph of $(2\pi)^{-1}\sqrt{4-t^2}$. The $2k^{th}$ moment of the semi-circle law is the Catalan number $C_k = (2\pi)^{-1}\int_{-2}^{2} t^{2k}\sqrt{4-t^2}dt$

Theorem 7. *Let $\{X_N\}_N$ be a sequence of* GUE *random matrices, normalized so that $E(|f_{ij}|^2) = 1/N$ for the entries of X_N. Then*

$$\lim_N \mathrm{E}(\mathrm{tr}(X_N^k)) = \frac{1}{2\pi}\int_{-2}^{2} t^k\sqrt{4-t^2}\,dt.$$

If we arrange that all the X_N's are defined on the same probability space $X_N : \Omega \to M_N(\mathbb{C})$, we can say something stronger: $\{\mathrm{tr}(X_N^k)\}_N$ converges to the k^{th} moment $(2\pi)^{-1}\int_{-2}^{2} t^k\sqrt{4-t^2}\,dt$ almost surely. We shall prove this in Chapters 4 and 5. See Theorem 4.4 and Remark 5.14.

1.10 Asymptotic freeness of independent GUE's

Suppose that for each N, $X_1, \ldots, X_s$ are independent $N \times N$ GUE's. For notational simplicity we suppress the dependence on N. Suppose $m_1, \ldots, m_r$ are positive integers and $i_1, i_2, \ldots, i_r \in [s]$ such that $i_1 \neq i_2, i_2 \neq i_3, \ldots, i_{r-1} \neq i_r$. Consider the random $N \times N$ matrix $Y_N := (X_{i_1}^{m_1} - c_{m_1}I)(X_{i_2}^{m_2} - c_{m_2}I)\cdots(X_{i_r}^{m_r} - c_{m_r}I)$, where c_m is the asymptotic value of the m-th moment of X_i (note that this is the same for all i); i.e. c_m is zero for m odd and the Catalan number $C_{m/2}$ for m even.

Each factor is centred asymptotically and adjacent factors have independent entries. We shall show that $\mathrm{E}(\mathrm{tr}(Y_N)) \to 0$ and we shall call this property *asymptotic freeness*. This will then motivate Voiculescu's definition of freeness.

First let us recall the principle of inclusion-exclusion (see Stanley [167, Vol. 1, Chap. 2]). Let S be a set and $E_1, \ldots, E_r \subseteq S$. Then

$$\begin{aligned}|S \setminus (E_1 \cup \cdots \cup E_r)| &= |S| - \sum_{i=1}^{r}|E_i| + \sum_{i_1 \neq i_2}|E_{i_1} \cap E_{i_2}| + \cdots \\ &\quad + (-1)^k \sum_{\substack{i_1,\ldots,i_k \\ \text{distinct}}} |E_{i_1} \cap \cdots \cap E_{i_k}| + \cdots + (-1)^r\,|E_1 \cap \cdots \cap E_r|; \qquad (1.10)\end{aligned}$$

for example, $|S \setminus (E_1 \cup E_2)| = |S| - (|E_1| + |E_2|) + |E_1 \cap E_2|$.

We can rewrite the right-hand side of (1.10) as

$$|S \setminus (E_1 \cup \cdots \cup E_r)| = \sum_{\substack{M \subseteq [r] \\ M=\{i_1,\dots,i_m\}}} (-1)^m \, |E_{i_1} \cap \cdots \cap E_{i_m}| = \sum_{M \subseteq [r]} (-1)^{|M|} \left| \bigcap_{i \in M} E_i \right|$$

provided we make the convention that $\bigcap_{i \in \emptyset} E_i = S$ and $(-1)^{|\emptyset|} = 1$.

Notation 8. *Let $i_1, \dots, i_m \in [s]$. We regard these labels as the colours of the matrices $X_{i_1}, X_{i_2}, \dots, X_{i_m}$. Given a pairing $\pi \in \mathcal{P}_2(m)$, we say that π respects the colours $i := (i_1, \dots, i_m)$, or to be brief, π respects i, if $i_r = i_p$ whenever (r, p) is a pair of π. Thus π respects i if and only if π only connects matrices of the same colour.*

Lemma 9. *Suppose $i_1, \dots, i_m \in [s]$ are positive integers. Then*

$$\mathrm{E}(\mathrm{tr}(X_{i_1} \cdots X_{i_m})) = |\{\pi \in NC_2(m) \mid \pi \text{ respects } i\}| + O(N^{-2}).$$

Proof: The proof proceeds essentially in the same way as for the genus expansion of moments of one GUE matrix.

$$\begin{aligned}
\mathrm{E}(\mathrm{tr}(X_{i_1} \cdots X_{i_m})) &= \sum_{j_1,\dots,j_m} \mathrm{E}(f^{(i_1)}_{j_1 j_2} \cdots f^{(i_m)}_{j_m, j_1}) \\
&= \sum_{j_1,\dots,j_m} \sum_{\pi \in \mathcal{P}_2(m)} \mathrm{E}_\pi(f^{(i_1)}_{j_1, j_2}, \dots, f^{(i_m)}_{j_m, j_1}) \\
&= \sum_{\substack{\pi \in \mathcal{P}_2(m) \\ \pi \text{ respects } i}} \sum_{j_1,\dots,j_m} \mathrm{E}_\pi(f^{(i_1)}_{j_1, j_2}, \dots, f^{(i_m)}_{j_m, j_1}) \\
&\overset{\text{by (1.9)}}{=} \sum_{\substack{\pi \in \mathcal{P}_2(m) \\ \pi \text{ respects } i}} N^{-2g_\pi} \\
&= |\{\pi \in NC_2(m) \mid \pi \text{ respects } i\}| + O(N^{-2}).
\end{aligned}$$

The penultimate equality follows in the same way as in the calculations leading to Theorem 3; for this note that we have for π which respects i that

$$\mathrm{E}_\pi(f^{(i_1)}_{j_1, j_2}, \dots, f^{(i_m)}_{j_m, j_1}) = \mathrm{E}_\pi(f^{(1)}_{j_1, j_2}, \dots, f^{(1)}_{j_m, j_1}),$$

so for the contribution of such a π which respects i, it does not play a role any more that we have several matrices instead of one. □

Theorem 10. *If* $i_1 \neq i_2, i_2 \neq i_3, \dots, i_{r-1} \neq i_r$, *then* $\lim_N \mathrm{E}(\mathrm{tr}(Y_N)) = 0$.

Proof: Let $I_1 = \{1, \dots, m_1\}$, $I_2 = \{m_1 + 1, \dots, m_1 + m_2\}, \dots, I_r = \{m_1 + \cdots + m_{r-1} + 1, \dots, m_1 + \cdots + m_r\}$ and $m = m_1 + \cdots + m_r$. Then

$$\begin{aligned}
&\mathrm{E}\left(\mathrm{tr}((X_{i_1}^{m_1} - c_{m_1} I) \cdots (X_{i_r}^{m_r} - c_{m_r} I))\right) \\
&\quad = \sum_{M \subseteq [r]} (-1)^{|M|} \Big[\prod_{i \in M} c_{m_i} \Big] \mathrm{E}\Big(\mathrm{tr}\Big(\prod_{j \notin M} X_{i_j}^{m_j} \Big)\Big) \\
&\quad = \sum_{M \subseteq [r]} (-1)^{|M|} \Big[\prod_{i \in M} c_{m_i} \Big] |\{\pi \in NC_2(\cup_{j \notin M} I_j) \mid \pi \text{ respects } i\}| + O(N^{-2}).
\end{aligned}$$

Let $S = \{\pi \in NC_2(m) \mid \pi \text{ respects } i\}$ and $E_j = \{\pi \in S \mid \text{elements of } I_j \text{ are only paired among themselves }\}$. Then

$$\Big| \bigcap_{j \in M} E_j \Big| = \Big[\prod_{j \in M} c_{m_j} \Big] \Big| \{\pi \in NC_2(\cup_{j \notin M} I_j) \mid \pi \text{ respects } i\} \Big|.$$

Thus

$$\mathrm{E}\left(\mathrm{tr}((X_{i_1}^{m_1} - c_{m_1} I) \cdots (X_{i_r}^{m_r} - c_{m_r} I))\right) = \sum_{M \subseteq [r]} (-1)^{|M|} \Big| \bigcap_{j \in M} E_j \Big| + O(N^{-2}).$$

So we must show that

$$\sum_{M \subseteq [r]} (-1)^{|M|} \Big| \bigcap_{j \in M} E_j \Big| = 0.$$

However, by inclusion-exclusion, this sum equals $|S \setminus (E_1 \cup \cdots \cup E_r)|$. Now $S \setminus (E_1 \cup \cdots \cup E_r)$ is the set of pairings of $[m]$ respecting i such that at least one element of each interval is connected to another interval. However this set is empty because elements of $S \setminus (E_1 \cup \cdots \cup E_r)$ must connect each interval to at least one other interval in a non-crossing way and thus form a non-crossing partition of the intervals $\{I_1, \dots, I_r\}$ without singletons, in which no pair of adjacent intervals are in the same block, and this is impossible. □

1.11 Freeness and asymptotic freeness

Let $X_{N,1}, \dots, X_{N,s}$ be independent $N \times N$ GUE random matrices. For each N let $\mathcal{A}_{N,i}$ be the polynomials in $X_{N,i}$ with complex coefficients. Let $\mathcal{A}_N$ be the algebra generated by $\mathcal{A}_{N,1}, \dots, \mathcal{A}_{N,s}$. For $A \in \mathcal{A}_N$ let $\varphi_N(A) = \mathrm{E}(\mathrm{tr}(A))$. Thus $\mathcal{A}_{N,1}, \dots, \mathcal{A}_{N,s}$ are unital subalgebras of the unital algebra $\mathcal{A}_N$ with state φ_N.

We have just shown in Theorem 7 that given a polynomial p we have that $\lim_N \varphi_N(A_{N,i})$ exists where $A_{N,i} = p(X_{N,i})$. Moreover we have from Theorem 10 that given polynomials $p_1, \ldots, p_r$ and positive integers $j_1, \ldots, j_r$ such that

- $\lim_N \varphi_N(A_{N,i}) = 0$ for $i = 1, 2, \ldots, r$
- $j_1 \neq j_2, j_2 \neq j_3, \ldots, j_{r-1} \neq j_r$

that $\lim_N \varphi_N(A_{N,1}A_{N,2} \cdots A_{N,r}) = 0$, where $A_{N,i} = p_i(X_{N,j_i})$. We thus say that the subalgebras $\mathcal{A}_{N,1}, \ldots, \mathcal{A}_{N,s}$ are asymptotically free because, in the limit as N tends to infinity, they satisfy the freeness property of Voiculescu. We state this below; in the next chapter, we shall explore freeness in detail. Note that asymptotic freeness implies that for any polynomials $p_1, \ldots, p_r$ and $i_1, \ldots, i_r \in [s]$ we have that $\lim_N \varphi_N(p_1(X_{N,i_1}) \cdots p_r(X_{N,i_r}))$ exists. So the random variables $\{X_{N,1}, \ldots, X_{N,s}\}$ have a joint limit distribution and it is the distribution of free random variables.

Definition 11. Let $(\mathcal{A}, \varphi)$ be a unital algebra with a unital linear functional. Suppose $\mathcal{A}_1, \ldots, \mathcal{A}_s$ are unital subalgebras. We say that $\mathcal{A}_1, \ldots, \mathcal{A}_s$ are *freely independent* (or just *free*) with respect to φ if whenever we have $r \geq 2$ and $a_1, \ldots, a_r \in \mathcal{A}$ such that

- $\varphi(a_i) = 0$ for $i = 1, \ldots, r$
- $a_i \in \mathcal{A}_{j_i}$ with $1 \leq j_i \leq s$ for $i = 1, \ldots, r$
- $j_1 \neq j_2, j_2 \neq j_3, \ldots, j_{r-1} \neq j_r$

we must have $\varphi(a_1 \cdots a_r) = 0$. We can say this succinctly as *the alternating product of centred elements is centred.*

1.12 Basic properties of freeness

We adopt the general philosophy of regarding freeness as a non-commutative analogue of the classical notion of independence in probability theory. Thus we refer to it often as *free independence*.

Definition 12. In general we refer to a pair $(\mathcal{A}, \varphi)$, consisting of a unital algebra $\mathcal{A}$ and a unital linear functional $\varphi : \mathcal{A} \to \mathbb{C}$ with $\varphi(1) = 1$, as a *non-commutative probability space*. If $\mathcal{A}$ is a $*$-algebra and φ is a *state*, i.e. in addition to $\varphi(1) = 1$ also positive (which means $\varphi(a^*a) \geq 0$ for all $a \in \mathcal{A}$), then we call $(\mathcal{A}, \varphi)$ a $*$-probability space. If $\mathcal{A}$ is a C^*-algebra and φ a state, $(\mathcal{A}, \varphi)$ is a *C^*-probability space*. Elements of $\mathcal{A}$ are called *non-commutative random variables* or just random variables.

If $(\mathcal{A}, \varphi)$ is a $*$-probability space and $\varphi(x^*x) = 0$ only when $x = 0$, we say that φ is *faithful*. If $(\mathcal{A}, \varphi)$ is a non-commutative probability space, we say that φ is *non-degenerate* if we have $\varphi(yx) = 0$ for all $y \in \mathcal{A}$ implies that $x = 0$ and $\varphi(xy) = 0$ for all $y \in \mathcal{A}$ implies that $x = 0$. By the Cauchy-Schwarz inequality, for a state on a $*$-probability space, "non-degenerate" and "faithful" are equivalent. If $\mathcal{A}$ is a von Neumann algebra and φ is a faithful normal state, i.e. continuous with respect to the weak-* topology, $(\mathcal{A}, \varphi)$ is called a *W^*-probability space*. If φ is also a trace, i.e.

$\varphi(ab) = \varphi(ba)$ for all $a, b \in \mathcal{A}$, then it is a *tracial W^*-probability space*. For a tracial W^*-probability space, we will usually write (M, τ) instead of $(\mathcal{A}, \varphi)$.

Proposition 13. *Let $(\mathcal{B}, \varphi)$ be a non-commutative probability space. Consider unital subalgebras $\mathcal{A}_1, \dots, \mathcal{A}_s \subset \mathcal{B}$ which are free. Let $\mathcal{A}$ be the algebra generated by $\mathcal{A}_1, \dots, \mathcal{A}_s$. Then $\varphi|_{\mathcal{A}}$ is determined by $\varphi|_{\mathcal{A}_1}, \dots, \varphi|_{\mathcal{A}_s}$ and the freeness condition.*

Proof: Elements in the generated algebra $\mathcal{A}$ are linear combinations of words of the form $a_1 \cdots a_k$ with $a_j \in \mathcal{A}_{i_j}$ for some $i_j \in \{1, \dots, s\}$ which meet the condition that neighbouring elements come from different subalgebras. We need to calculate $\varphi(a_1 \cdots a_k)$ for such words. Let us proceed in an inductive fashion.

We know how to calculate $\varphi(a)$ for $a \in A_i$ for some $i \in \{1, \dots, s\}$.

Now suppose we have a word of the form $a_1 a_2$ with $a_1 \in \mathcal{A}_{i_1}$ and $a_2 \in \mathcal{A}_{i_2}$ with $i_1 \neq i_2$. By the definition of freeness, this implies

$$\varphi[(a_1 - \varphi(a_1)1)(a_2 - \varphi(a_2)1)] = 0.$$

But

$$(a_1 - \varphi(a_1)1)(a_2 - \varphi(a_2)1) = a_1 a_2 - \varphi(a_2)a_1 - \varphi(a_1)a_2 + \varphi(a_1)\varphi(a_2)1.$$

Hence we have

$$\varphi(a_1 a_2) = \varphi\big[\varphi(a_2)a_1 + \varphi(a_1)a_2 - \varphi(a_1)\varphi(a_2)1\big] = \varphi(a_1)\varphi(a_2).$$

Continuing in this fashion, we know that $\varphi(\mathring{a}_1 \cdots \mathring{a}_k) = 0$ by the definition of freeness, where $\mathring{a}_i = a_i - \varphi(a_i)1$ is a centred random variable. But then

$$\varphi(\mathring{a}_1 \cdots \mathring{a}_k) = \varphi(a_1 \cdots a_k) + \text{lower order terms in } \varphi,$$

where the lower order terms are already dealt with by induction hypothesis. $\square$

Remark 14. Let $(\mathcal{A}, \varphi)$ be a non-commutative probability space. For any subalgebra $\mathcal{B} \subset \mathcal{A}$, we let $\mathring{\mathcal{B}} = \mathcal{B} \cap \ker \varphi$. Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be unital subalgebras of $\mathcal{A}$; we let $\mathcal{A}_1 \vee \mathcal{A}_2$ be the subalgebra of $\mathcal{A}$ generated algebraically by $\mathcal{A}_1$ and $\mathcal{A}_2$. With this notation we can restate Proposition 13 as follows. If $\mathcal{A}_1$ and $\mathcal{A}_2$ are free, then

$$\ker \varphi|_{\mathcal{A}_1 \vee \mathcal{A}_2} = \sum_{n \geq 1}^{\oplus} \sum_{\alpha_1 \neq \cdots \neq \alpha_n}^{\oplus} \mathring{\mathcal{A}}_{\alpha_1} \mathring{\mathcal{A}}_{\alpha_2} \cdots \mathring{\mathcal{A}}_{\alpha_n} \tag{1.11}$$

where $\alpha_1, \dots, \alpha_n \in \{1, 2\}$.

For subalgebras $\mathcal{C} \subset \mathcal{B} \subset \mathcal{A}$, we shall let $\mathcal{B} \ominus \mathcal{C} = \{b \in \mathcal{B} \mid \varphi(cb) = 0$ for all $c \in \mathcal{C}\}$. When $\varphi|_{\mathcal{C}}$ is non-degenerate, we have $\mathcal{C} \ominus \mathcal{C} = \{0\}$.

Exercise 9. Let $(\mathcal{A}, \varphi)$ be a non-commutative probability space. Suppose $\mathcal{A}_1, \mathcal{A}_2 \subset \mathcal{A}$ are unital subalgebras and are free with respect to φ. If $\varphi|_{\mathcal{A}_1}$ is non-degenerate, then

$$(\mathcal{A}_1 \vee \mathcal{A}_2) \ominus \mathcal{A}_1 = \mathring{\mathcal{A}}_2 \oplus \sum_{n \geq 2}^{\oplus} \sum_{\alpha_1 \neq \cdots \neq \alpha_n}^{\oplus} \mathring{\mathcal{A}}_{\alpha_1} \mathring{\mathcal{A}}_{\alpha_2} \cdots \mathring{\mathcal{A}}_{\alpha_n}. \tag{1.12}$$

Definition 15. Let $(\mathcal{A}, \varphi)$ be a non-commutative probability space. Elements $a_1, \ldots, a_s \in \mathcal{A}$ are said to be *free* or *freely independent* if the generated unital subalgebras $\mathcal{A}_i = \mathrm{alg}(1, a_i)$ $(i = 1, \ldots, s)$ are free in $\mathcal{A}$ with respect to φ. If $(\mathcal{A}, \varphi)$ is a $*$-probability space, then we say that $a_1, \ldots, a_s \in \mathcal{A}$ are $*$-*free* if the generated unital $*$-subalgebras $\mathcal{B}_i = \mathrm{alg}(1, a_i, a_i^*)$ $(i = 1, \ldots, s)$ are free in $\mathcal{A}$ with respect to φ. In the same way, $(*$-)freeness between sets of variables is defined by the freeness of the generated unital $(*$-)subalgebras.

In terms of random variables, Proposition 13 says that mixed moments of free variables are calculated in a specific way out of the moments of the separate variables. This is in clear analogy to the classical notion of independence.

Let us look at some examples for such calculations of mixed moments. For example, if a, b are freely independent, then $\varphi[(a - \varphi(a)1)(b - \varphi(b)1)] = 0$, implying $\varphi(ab) = \varphi(a)\varphi(b)$.

In a slightly more complicated example, let $\{a_1, a_2\}$ be free from b. Then applying the state to the corresponding centred word:

$$\varphi[(a_1 - \varphi(a_1)1)(b - \varphi(b)1)(a_2 - \varphi(a_2)1)] = 0,$$

hence the linearity of φ gives

$$\varphi(a_1 b a_2) = \varphi(a_1 a_2)\varphi(b). \tag{1.13}$$

A similar calculation shows that if $\{a_1, a_2\}$ is free from $\{b_1, b_2\}$, then

$$\begin{aligned} \varphi(a_1 b_1 a_2 b_2) = \; & \varphi(a_1 a_2)\varphi(b_1)\varphi(b_2) \\ & + \varphi(a_1)\varphi(a_2)\varphi(b_1 b_2) - \varphi(a_1)\varphi(a_2)\varphi(b_1)\varphi(b_2). \end{aligned} \tag{1.14}$$

It is important to note that while free independence is analogous to classical independence, it is not a generalization of the classical case. Classical commuting random variables a, b are free only in trivial cases, $\varphi(aabb) = \varphi(abab)$, but the left-hand side is $\varphi(aa)\varphi(bb)$, while the right-hand side is $\varphi(a^2)\varphi(b)^2 + \varphi(a)^2\varphi(b^2) - \varphi(a)^2\varphi(b)^2$, which implies $\varphi[(a - \varphi(a))^2] \cdot \varphi[(b - \varphi(b))^2] = 0$. But then (note that states in classical probability spaces are always positive and faithful) one of the factors inside φ must be 0, so that one of a, b must be a scalar.

Observe that while freeness gives a concrete rule for calculating mixed moments, this rule is a priori quite complicated. We will come back to this question for a better understanding of this rule in the next chapter. For the moment let us just note the following.

Proposition 16. *Let $(\mathcal{A}, \varphi)$ be a non-commutative probability space. The subalgebra of scalars $\mathbb{C}1$ is free from any other unital subalgebra $\mathcal{B} \subset \mathcal{A}$.*

Proof: Let $a_1 \cdots a_k$ be an alternating word in centred elements of $\mathbb{C}1, \mathcal{B}$. The case $k = 1$ is trivial, otherwise we have at least one $a_j \in \mathbb{C}1$. But then $\varphi(a_j) = 0$ implies $a_j = 0$, so $a_1 \cdots a_k = 0$. Thus obviously $\varphi(a_1 \cdots a_k) = 0$. □

1.13 Classical moment-cumulant formulas

At the beginning of this chapter, we introduced the cumulants of a probability measure ν via the logarithm of its characteristic function: if $\{\alpha_n\}_n$ are the moments of ν and

$$\sum_{n\geq 1} k_n \frac{z^n}{n!} = \log\left(1 + \sum_{n\geq 1} \alpha_n \frac{z^n}{n!}\right) \tag{1.15}$$

is the logarithm of the moment-generating function, then $\{k_n\}_n$ are the cumulants of ν. We gave without proof two formulas (1.1) and (1.2) showing how to compute the n^{th} moment from the first n cumulants and conversely.

In the exercises below, we shall prove equations (1.1) and (1.2) as well as show the very simple restatements in terms of set partitions

$$\alpha_n = \sum_{\pi \in \mathcal{P}(n)} k_\pi \qquad \text{and} \qquad k_n = \sum_{\pi \in \mathcal{P}(n)} (-1)^{\#(\pi)-1} (\#(\pi) - 1)!\, \alpha_\pi .$$

The simplicity of these formulas, in particular the first, makes them very useful for computation. Moreover they naturally lead to the moment-cumulant formulas for the free cumulants in which the set $\mathcal{P}(n)$ of all partitions of $[n]$ is replaced by $NC(n)$, the set of non-crossing partitions of $[n]$. This will be taken up in Chapter 2.

It was shown in Exercise 4 that if we have two sequences $\{\alpha_n\}_n$ and $\{k_n\}_n$ such that $\alpha_n = \sum_{\pi \in \mathcal{P}(n)} k_\pi$, then we have (1.15) as relation between their exponential power series. In Exercises 11 and 12, this is proved again starting from the formal power series relation and ending with the first moment-cumulant relation. This can be regarded as a warm-up for Exercises 13 and 14 when we prove the second half of the moment-cumulant relation:

$$k_n = \sum_{\pi \in \mathcal{P}(n)} (-1)^{\#(\pi)-1} (\#(\pi) - 1)! \alpha_\pi .$$

This formula can also be proved by the general theory of Möbius inversion in $\mathcal{P}(n)$ after identifying the Möbius function on $\mathcal{P}(n)$ (see [137, Ex. 10.33]).

So far we have only considered cumulants of a single random variable; we need an extension to several random variables so that k_n becomes a n-linear functional. We begin with mixed moments and extend the notation used in Section 1.5. Let $\{X_i\}_i$ be a sequence of random variables and $\pi \in \mathcal{P}(n)$; we let

$$\mathrm{E}_\pi(X_1, \ldots, X_n) = \prod_{\substack{V \in \pi \\ V=(i_1,\ldots,i_l)}} \mathrm{E}(X_{i_1} X_{i_2} \cdots X_{i_l}).$$

Then we set

$$k_n(X_1, \ldots, X_n) = \sum_{\pi \in \mathcal{P}(n)} (-1)^{\#(\pi)-1} (\#(\pi) - 1)!\, \mathrm{E}_\pi(X_1, \ldots, X_n).$$

We then define k_π as above; namely, for $\pi \in \mathcal{P}(n)$, we set

$$k_\pi(X_1, \ldots, X_n) = \prod_{\substack{V \in \pi \\ V=(i_1,\ldots,i_l)}} k_l(X_{i_1}, \ldots, X_{i_l}).$$

Our moment-cumulant formula can be recast as a multilinear moment-cumulant formula

$$\mathrm{E}(X_1 \cdots X_n) = \sum_{\pi \in \mathcal{P}(n)} k_\pi(X_1, \ldots, X_n).$$

Another formula we shall need is the product formula of Leonov and Shiryaev for cumulants (see [137, Theorem 11.30]). Let $n_1, \ldots, n_r$ be positive integers and $n = n_1 + \cdots + n_r$. Given random variables $X_1, \ldots, X_n$, let $Y_1 = X_1 \cdots X_{n_1}$, $Y_2 = X_{n_1+1} \cdots X_{n_1+n_2}, \ldots, Y_r = X_{n_1+\cdots+n_{r-1}+1} \cdots X_{n_1+\cdots+n_r}$. Then

$$k_r(Y_1, \ldots, Y_r) = \sum_{\substack{\pi \in \mathcal{P}(n) \\ \pi \vee \tau = 1_n}} k_\pi(X_1, \ldots, X_n) \tag{1.16}$$

where the sum runs over all $\pi \in \mathcal{P}(n)$ such that $\pi \vee \tau = 1_n$ and $\tau \in \mathcal{P}(n)$ is the partition with r blocks

$$\{(1, \ldots, n_1), (n_1 + 1, \ldots, n_1 + n_2), \cdots, (n_1 + \cdots + n_{r-1} + 1, \ldots, n_1 + \cdots + n_r)\}$$

and $1_n \in \mathcal{P}(n)$ is the partition with one block. Here $\vee$ denotes the join in the lattice of all partitions (see [137, Remark 9.19]).

In the next chapter, we will have in (2.19) an analogue of (1.16) for free cumulants.

1.14 Additional exercises

Exercise 10. (i) Let $\sum_{n=1}^{\infty} \beta_n z^n$ be a formal power series. Using the power series expansion for e^x, show that as a formal power series

$$\exp\Big(\sum_{n=1}^{\infty} \beta_n z^n\Big) = 1 + \sum_{n=1}^{\infty} \sum_{m=1}^{n} \sum_{\substack{l_1,\dots,l_m \geq 1 \\ l_1+\cdots+l_m=n}} \frac{\beta_{l_1}\cdots\beta_{l_m}}{m!} z^n.$$

(ii) Show

$$\exp\Big(\sum_{n=1}^{\infty} \beta_n z^n\Big) = 1 + \sum_{n=1}^{\infty} \sum_{\substack{r_1,\dots,r_n \geq 0 \\ 1\cdot r_1+\cdots+n\cdot r_n=n}} \frac{\beta_1^{r_1}\cdots\beta_n^{r_n}}{r_1!r_2!\cdots r_n!} z^n.$$

Use this to prove equation (1.1).

Exercise 11. Let $\sum_{n=1}^{\infty} \frac{\beta_n}{n!} z^n$ be a formal power series. For a partition π of type $(r_1, r_2, \dots, r_n)$, let $\beta_\pi = \beta_1^{r_1}\beta_2^{r_2}\cdots\beta_n^{r_n}$. Show that

$$\exp\Big(\sum_{n=1}^{\infty} \frac{\beta_n}{n!} z^n\Big) = 1 + \sum_{n=1}^{\infty} \Big(\sum_{\pi\in\mathcal{P}(n)} \beta_\pi\Big)\frac{z^n}{n!}.$$

Exercise 12. Let $\sum_{n=1}^{\infty} \beta_n z^n$ be a formal power series. Using the power series expansion for $\log(1+x)$, show that

$$\log\Big(1+\sum_{n=1}^{\infty} \beta_n z^n\Big) = \sum_{n=1}^{\infty} \sum_{1\cdot r_1+\cdots+n\cdot r_n=n} (-1)^{r_1+\cdots+r_n-1}(r_1+\cdots+r_n-1)!\,\frac{\beta_1^{r_1}\cdots\beta_n^{r_n}}{r_1\cdots r_n!} z^n.$$

Use this to prove equation (1.2).

Exercise 13. (i) Let $\sum_{n=1}^{\infty} \alpha_n \frac{z^n}{n!}$ be a formal power series. Show that

$$\log\Big(1+\sum_{n=1}^{\infty} \alpha_n \frac{z^n}{n!}\Big) = \sum_{n=1}^{\infty} \Big(\sum_{\pi\in\mathcal{P}(n)} (-1)^{\#(\pi)-1}(\#(\pi)-1)!\,\alpha_\pi\Big)\frac{z^n}{n!}.$$

(ii) Let

$$k_n = \sum_{\pi\in\mathcal{P}(n)} (-1)^{\#(\pi)-1}(\#(\pi)-1)!\,\alpha_\pi.$$

Use the result of Exercise 11 to show that

$$\alpha_n = \sum_{\pi \in \mathcal{P}(n)} k_\pi .$$

Exercise 14. Suppose ν is a probability measure with moments $\{\alpha_n\}_n$ of all orders and let $\{k_n\}_n$ be its sequence of cumulants. Show that

$$\alpha_n = \sum_{\pi \in \mathcal{P}(n)} k_\pi \qquad \text{and} \qquad k_n = \sum_{\pi \in \mathcal{P}(n)} (-1)^{\#(\pi)-1} (\#(\pi) - 1)! \, \alpha_\pi .$$

Chapter 2
The Free Central Limit Theorem and Free Cumulants

Recall from Chapter 1 that if $(\mathcal{A}, \varphi)$ is a non-commutative probability space and $\mathcal{A}_1, \ldots, \mathcal{A}_s$ are subalgebras of $\mathcal{A}$ which are free with respect to φ, then freeness gives us in principle a rule by which we can evaluate $\varphi(a_1 a_2 \cdots a_k)$ for any alternating word in random variables $a_1, a_2, \ldots, a_k$. Thus we can in principle calculate all mixed moments for a system of free random variables. However, we do not yet have any concrete idea of the structure of this factorization rule. This situation will be greatly clarified by the introduction of *free cumulants*. Classical cumulants appeared in Chapter 1, where we saw that they are intimately connected with the combinatorial notion of set partitions. Our free cumulants will be linked in a similar way to the lattice of non-crossing set partitions; the latter were introduced in combinatorics by Kreweras [113]. We will motivate the appearance of free cumulants and non-crossing partition lattices in free probability theory by examining in detail a proof of the central limit theorem by the method of moments.

The combinatorial approach to free probability was initiated by Speicher in [159, 161], in order to get alternative proofs for the free central limit theorem and the main properties of the R-transform, which had been treated before by Voiculescu in [176, 177] by more analytic tools. Nica showed a bit later in [135] how this combinatorial approach connects in general to Voiculescu's operator-theoretic approach in terms of creation and annihilation operators on the full Fock space. The combinatorial path was pursued much further by Nica and Speicher; for more details on this, we refer to the standard reference [137].

2.1 The classical and free central limit theorems

Our setting is that of a non-commutative probability space $(\mathcal{A}, \varphi)$ and a sequence $(a_i)_{i \in \mathbb{N}} \subset \mathcal{A}$ of centred and identically distributed random variables. This means that $\varphi(a_i) = 0$ for all $i \geq 1$ and that $\varphi(a_i^n) = \varphi(a_j^n)$ for any $i, j, n \geq 1$. We assume that our random variables a_i, $i \geq 1$ are either classically independent or freely independent as defined in Chapter 1. Either form of independence gives us a *factorization rule* for calculating mixed moments in the random variables.

J.A. Mingo, R. Speicher, *Free Probability and Random Matrices*,
Fields Institute Monographs 35, DOI 10.1007/978-1-4939-6942-5_2

For $k \geq 1$, set

$$S_k := \frac{1}{\sqrt{k}}(a_1 + \cdots + a_k). \tag{2.1}$$

The Central Limit Theorem is a statement about the limit distribution of the random variable S_k in the large k limit. Let us begin by reviewing the kind of convergence we shall be considering.

Recall that given a real-valued random variable X on a probability space, we have a probability measure μ_X on $\mathbb{R}$, called the *distribution* of X. The distribution of X is defined by the equation

$$\mathrm{E}(f(X)) = \int f(t)\, d\mu_X(t) \text{ for all } f \in C_b(\mathbb{R}) \tag{2.2}$$

where $C_b(\mathbb{R})$ is the C^*-algebra of all bounded continuous functions on $\mathbb{R}$. We say that a probability measure μ on $\mathbb{R}$ is *determined by it moments* if μ has moments $\{\alpha_k\}_k$ of all orders and μ is the only probability measure on $\mathbb{R}$ with moments $\{\alpha_k\}_k$. If the moment generating function of μ has a positive radius of convergence, then μ is determined by its moments (see Billingsley [41, Theorem 30.1]).

Exercise 1. Show that a compactly supported measure is determined by its moments.

A more general criterion is the Carleman condition (see Akhiezer [3, p. 85]) which says that a measure μ is determined by its moments $\{\alpha_k\}_k$ if we have $\sum_{k\geq 1}(\alpha_{2k})^{-1/(2k)} = \infty$.

Exercise 2. Using the Carleman condition, show that the Gaussian measure is determined by its moments.

A sequence of probability measures $\{\mu_n\}_n$ on $\mathbb{R}$ is said to converge *weakly* to μ if $\{\int f\, d\mu_n\}_n$ converges to $\int f\, d\mu$ for all $f \in C_b(\mathbb{R})$. Given a sequence $\{X_n\}_n$ of real-valued random variables, we say that $\{X_n\}_n$ *converges in distribution* (or *converges in law)* if the probability measures $\{\mu_{X_n}\}_n$ converge weakly.

If we are working in a non-commutative probability space $(\mathcal{A}, \varphi)$, we call an element a of A a *non-commutative random variable*. Given such an a, we may define μ_a by $\int p\, d\mu_a = \varphi(p(a))$ for all polynomials $p \in \mathbb{C}[x]$. At this level of generality, we may not be able to define $\int f\, d\mu_a$ for all functions $f \in C_b(\mathbb{R})$, so we call the linear functional $\mu_a : \mathbb{C}[x] \to \mathbb{C}$ the *algebraic distribution* of a, even if it is not a probability measure. However when it is clear from the context we shall just call μ_a the distribution of a. Note that if a is a self-adjoint element of a C^*-algebra and φ is positive and has norm 1, then μ_a extends from $\mathbb{C}[x]$ to $C_b(\mathbb{R})$ and thus μ_a becomes a probability measure on $\mathbb{R}$.

Definition 1. Let $(\mathcal{A}_k, \varphi_k)$, for $k \in \mathbb{N}$, and $(\mathcal{A}, \varphi)$ be non-commutative probability spaces.

1) Let $(b_k)_{k\in\mathbb{N}}$ be a sequence of non-commutative random variables with $b_k \in \mathcal{A}_k$, and let $b \in \mathcal{A}$. We say that b_k *converges in distribution* to b, denoted by $b_k \xrightarrow{\text{distr}} b$, if

$$\lim_{k\to\infty} \varphi_k(b_k^n) = \varphi(b^n) \tag{2.3}$$

for any fixed $n \in \mathbb{N}$.

2) More generally, let I be an index set. For each $i \in I$, let $b_k^{(i)} \in \mathcal{A}_k$ for $k \in \mathbb{N}$ and $b^{(i)} \in \mathcal{A}$. We say that $(b_k^{(i)})_{i\in I}$ *converges in distribution* to $(b^{(i)})_{i\in I}$, denoted by $(b_k^{(i)})_{i\in I} \xrightarrow{\text{distr}} (b^{(i)})_{i\in I}$, if

$$\lim_{k\to\infty} \varphi_k(b_k^{(i_1)} \cdots b_k^{(i_n)}) = \varphi(b^{(i_1)} \cdots b^{(i_n)}) \tag{2.4}$$

for all $n \in \mathbb{N}$ and all $i_1, \ldots, i_n \in I$.

Note that this definition is neither weaker nor stronger than weak convergence of the corresponding distributions. For real-valued random variables, the convergence in (2.3) is sometimes called convergence in moments. However there is an important case where the two conditions coincide. If we have a sequence of probability measures $\{\mu_k\}_k$ on $\mathbb{R}$, each having moments of all orders and a probability measure μ determined by its moments, such that for every n we have $\int t^n \, d\mu_k(t) \to \int t^n \, d\mu$ as $k \to \infty$, then $\{\mu_k\}_k$ converges weakly to μ (see Billingsley [41, Theorem 30.2]). To see that weak convergence does not imply convergence in moments, consider the sequence $\{\mu_k\}_k$ where $\mu_k = (1 - 1/k)\delta_0 + (1/k)\delta_k$ and δ_k is the probability measure with an atom at k of mass 1.

Exercise 3. Show that $\{\mu_k\}_k$ converges weakly to δ_0 but that we do not have convergence in moments.

We want to make a statement about convergence in distribution of the random variables $(S_k)_{k\in\mathbb{N}}$ from (2.1) (which all come from the same underlying non-commutative probability space). Thus we need to do a moment calculation. Let $[k] = \{1, \ldots, k\}$ and $[n] = \{1, \ldots, n\}$. We have

$$\varphi(S_k^n) = \frac{1}{k^{n/2}} \sum_{r:[n]\to[k]} \varphi(a_{r_1} \cdots a_{r_n}).$$

It turns out that the fact that the random variables $a_1, \ldots, a_k$ are independent and identically distributed makes the task of calculating this sum less complex than it initially appears. The key observation is that because of (classical or free) independence of the a_i's and the fact that they are identically distributed, the value of $\varphi(a_{r_1} \cdots a_{r_n})$ depends not on all details of the multi-index r, but just on the

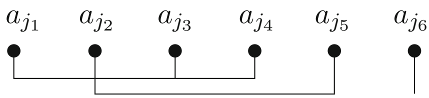


Fig. 2.1 Suppose $j_1 = j_3 = j_4$ and $j_2 = j_5$ but $\{j_1, j_2, j_6\}$ are distinct. Then $\ker(j) = \{(1,3,4),(2,5),(6)\}$

information where the indices are the same and where they are different. Let us recall some notation from the proof of Theorem 1.1.

Notation 2. *Let $i = (i_1, \ldots, i_n)$ be a multi-index. Then its* kernel*, denoted by* $\ker i$*, is that partition in $\mathcal{P}(n)$ whose blocks correspond exactly to the different values of the indices (Fig. 2.1),*

$$k \text{ and } l \text{ are in the same block of } \ker i \iff i_k = i_l.$$

Lemma 3. *With this notation we have that* $\ker i = \ker j$ *implies* $\varphi(a_{i_1} \cdots a_{i_n}) = \varphi(a_{j_1} \cdots a_{j_n})$.

Proof: To see this note first that $\ker i = \ker j$ implies that the i-indices can be obtained from the j-indices by the application of some permutation σ, i.e. $(j_1, \ldots, j_n) = (\sigma(i_1), \ldots, \sigma(i_n))$. We know that the random variables $a_1, \ldots, a_k$ are (classically or freely) independent. This means that we have a factorization rule for calculating mixed moments in $a_1, \ldots, a_k$ in terms of the moments of individual a_i's. In particular this means that $\varphi(a_{i_1} \cdots a_{i_n})$ can be written as some expression in moments $\varphi(a_i^r)$, while $\varphi(a_{j_1} \cdots a_{j_n})$ can be written as that same expression except with $\varphi(a_i^r)$ replaced by $\varphi(a_{\sigma(i)}^r)$. However, since our random variables all have the same distribution, then $\varphi(a_i^r) = \varphi(a_{\sigma(i)}^r)$ for any i, j, and thus $\varphi(a_{i_1} \cdots a_{i_n}) = \varphi(a_{j_1} \cdots a_{j_n})$. □

Let us denote the common value of $\varphi(a_{i_1} \cdots a_{i_n})$ for all i with $\ker i = \pi$, for some $\pi \in \mathcal{P}(n)$, by $\varphi(\pi)$. Consequently, we have

$$\varphi(S_k^n) = \frac{1}{k^{n/2}} \sum_{\pi \in \mathcal{P}(n)} \varphi(\pi) \cdot |\{i : [n] \to [k] \mid \ker i = \pi\}|.$$

It is not difficult to see that

$$\#\{i : [n] \to [k] \mid \ker i = \pi\} = k(k-1)\cdots(k - \#(\pi) + 1)$$

because we have k choices for the first block of π, $k-1$ choices for the second block of π, and so on until the last block where we have $k - \#(\pi) + 1$.

Then what we have proved is that

$$\varphi(S_k^n) = \frac{1}{k^{n/2}} \sum_{\pi \in \mathcal{P}(n)} \varphi(\pi) \cdot k(k-1)\cdots(k-\#(\pi)+1).$$

The great advantage of this expression over what we started with is that the number of terms does not depend on k. Thus we are in a position to take the limit as $k \to \infty$, provided we can effectively estimate each term of the sum.

Our first observation is the most obvious one, namely, we have

$$k(k-1)\cdots(k-\#(\pi)+1) \sim k^{\#(\pi)} \qquad \text{as } k \to \infty.$$

Next observe that if π has a block of size 1, then we will have $\varphi(\pi) = 0$. Indeed suppose that $\pi = \{V_1, \ldots, V_m, \ldots, V_s\} \in \mathcal{P}(n)$ with $V_m = \{l\}$ for some $l \in [n]$. Then we will have

$$\varphi(\pi) = \varphi(a_{j_1} \cdots a_{j_{l-1}} a_{j_l} a_{j_{l+1}} \cdots a_{j_n})$$

where $\ker(j) = \pi$ and thus $j_l \notin \{j_1, \ldots, j_{l-1}, j_{l+1}, \ldots, j_n\}$. Hence we can write $\varphi(\pi) = \varphi(b a_{j_l} c)$, where $b = a_{j_1} \cdots a_{j_{l-1}}$ and $c = a_{j_{l+1}} \cdots a_{j_n}$ and thus

$$\varphi(\pi) = \varphi(b a_{j_l} c) = \varphi(a_{j_l})\varphi(bc) = 0,$$

since a_{j_l} is (classically or freely) independent of $\{b, c\}$. (For the free case, this factorization was considered in Equation (1.13) in the last chapter. In the classical case, it is obvious, too.) Of course, for this part of the argument, it is crucial that we assume our variables a_i to be centred.

Thus the only partitions which contribute to the sum are those with blocks of size at least 2. Note that such a partition can have at most $n/2$ blocks. Now,

$$\lim_{k\to\infty} \frac{k^{\#(\pi)}}{k^{n/2}} = \begin{cases} 1, & \text{if } \#(\pi) = n/2 \\ 0, & \text{if } \#(\pi) < n/2 \end{cases}.$$

Hence the only partitions which contribute to the sum in the $k \to \infty$ limit are those with *exactly* $n/2$ blocks, i.e. partitions each of whose blocks has size 2. Such partitions are called *pairings*, and the set of pairings is denoted $\mathcal{P}_2(n)$.

Thus we have shown that

$$\lim_{k\to\infty} \varphi(S_k^n) = \sum_{\pi \in \mathcal{P}_2(n)} \varphi(\pi).$$

Note that in particular if n is odd, then $\mathcal{P}_2(n) = \emptyset$, so that the odd limiting moments vanish. In order to determine the even limiting moments, we must distinguish between the setting of classical independence and free independence.

2.1.1 Classical central limit theorem

In the case of classical independence, our random variables commute and factorize completely with respect to φ. Thus if we denote by $\varphi(a_i^2) = \sigma^2$ the common variance of our random variables, then for any pairing $\pi \in \mathcal{P}_2(n)$ we have $\varphi(\pi) = \sigma^n$. Thus we have

$$\lim_{k\to\infty} \varphi(S_k^n) = \sum_{\pi \in \mathcal{P}_2(n)} \sigma^n = \begin{cases} \sigma^n (n-1)(n-3)\dots 5\cdot 3\cdot 1, & \text{if } n \text{ even} \\ 0, & \text{if } n \text{ odd} \end{cases}.$$

From Section 1.1, we recognize these as exactly the moments of a Gaussian random variable of mean 0 and variance σ^2. Since by Exercise 2 the normal distribution is determined by its moments, and hence our convergence in moments is the same as the classical convergence in distribution, we get the following form of the classical central limit theorem: if $(a_i)_{i\in\mathbb{N}}$ are classically independent random variables which are identically distributed with $\varphi(a_i) = 0$ and $\varphi(a_i^2) = \sigma^2$, and having all moments, then S_k converges in distribution to a Gaussian random variable with mean 0 and variance σ^2. Note that one can see the derivation above also as a proof of the Wick formula for Gaussian random variables if one takes the central limit theorem for granted.

2.1.2 Free central limit theorem

Now we want to deal with the case where the random variables are freely independent. In this case, $\varphi(\pi)$ will not be the same for all pair partitions $\pi \in \mathcal{P}_2(2n)$ (we focus on the even moments now because we already know that the odd ones are zero). Let's take a look at some examples:

$$\begin{aligned}
\varphi(\{(1,2),(3,4)\}) &= \varphi(a_1a_1a_2a_2) = \varphi(a_1^2)\varphi(a_2^2) = \sigma^4 \\
\varphi(\{(1,4),(2,3)\}) &= \varphi(a_1a_2a_2a_1) = \varphi(a_1^2)\varphi(a_2^2) = \sigma^4 \\
\varphi(\{(1,3),(2,4)\}) &= \varphi(a_1a_2a_1a_2) = 0.
\end{aligned}$$

The last equality is just from the definition of freeness, because $a_1a_2a_1a_2$ is an alternating product of centred free variables.

In general, we will get $\varphi(\pi) = \sigma^{2n}$ if we can successively remove neighbouring pairs of identical random variables in the word corresponding to π so that we end with a single pair (see Fig. 2.2); if we cannot we will have $\varphi(\pi) = 0$ as in the example $\varphi(a_1a_2a_1a_2) = 0$ above. Thus the only partitions that give a non-zero contribution are the *non-crossing* ones (see [137, p. 122] for details). Non-crossing pairings were encountered already in Chapter 1, where we denoted the set of non-crossing pairings by $NC_2(2n)$. Then we have as our free central limit theorem that

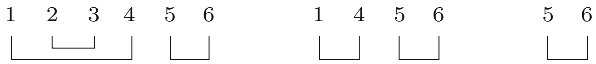


Fig. 2.2 We start with the pairing $\{(1,4),(2,3),(5,6)\}$ and remove the pair $(2,3)$ of adjacent elements (middle figure). Next we remove the pair $(1,4)$ of adjacent elements. We are then left with a single pair; so the pairing must have been non-crossing to start with

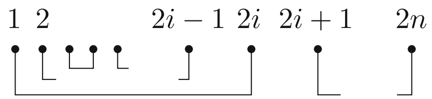


Fig. 2.3 We have C_{i-1} possible pairings on $[2, 2i-1]$ and C_{n-i} possible pairings on $[2i+1, 2n]$

$$\lim_{k\to\infty} \varphi(S_k^{2n}) = \sigma^{2n} \cdot |NC_2(2n)|.$$

In Chapter 1 we already mentioned that the cardinality $C_n := |NC_2(2n)|$ is given by the *Catalan numbers*. We want now to elaborate on the proof of this claim.

A very simple method is to show that the pairings are in a bijective correspondence with the Dyck paths; by using André's reflection principle, one finds that there are $\binom{2n}{n} - \binom{2n}{n-1} = \frac{1}{n+1}\binom{2n}{n}$ such paths (see [137, Prop. 2.11] for details).

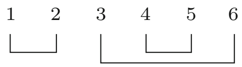
Our second method for counting non-crossing pairings is to find a simple recurrence which they satisfy. The idea is to look at the block of a pairing which contains the number 1. In order for the pairing to be non-crossing, 1 must be paired with some even number in the set $[2n]$, else we would necessarily have a crossing. Thus 1 must be paired with $2i$ for some $i \in [n]$. Now let i run through all possible values in $[n]$, and count for each the number of non-crossing pairings that contain this pair, as in the diagram (Fig. 2.3).

In this way we see that the cardinality C_n of $NC_2(2n)$ must satisfy the recurrence relation

$$C_n = \sum_{i=1}^{n} C_{i-1} C_{n-i}, \tag{2.5}$$

with initial condition $C_0 = 1$. One can then check using a generating function that the Catalan numbers satisfy this recurrence; hence $C_n = \frac{1}{n+1}\binom{2n}{n}$.

Exercise 4. Let $f(z) = \sum_{n=0}^{\infty} C_n z^n$ be the generating function for $\{C_n\}_n$, where $C_0 = 1$ and C_n satisfies the recursion (2.5).



1	3	4
2	5	6

Fig. 2.4 In the bijection between $NC_2(6)$ and 2×3 standard Young tableaux, the pairing $\{(1,2),(3,6),(4,5)\}$ gets mapped to the tableaux on the right

(*i*) Show that $1 + zf(z)^2 = f(z)$.
(*ii*) Show that f is also the power series for $\frac{1-\sqrt{1-4z}}{2z}$.
(*iii*) Show that $C_n = \frac{1}{n+1}\binom{2n}{n}$.

We can also prove directly that $C_n = \frac{1}{n+1}\binom{2n}{n}$ by finding a bijection between $NC_2(2n)$ and some standard set of objects which we can see directly is enumerated by the Catalan numbers. A reasonable choice for this "canonical" set is the collection of $2 \times n$ standard Young tableaux. A standard Young tableaux of shape $2 \times n$ is a filling of the squares of a $2 \times n$ grid with the numbers $1, \ldots, 2n$ which is strictly increasing in each of the two rows and each of the n columns. The number of these standard Young tableaux is very easy to calculate, using a famous and fundamental result known as the *hook-length formula* [167, Vol. 2, Corollary 7.21.6]. The hook-length formula tells us that the number of standard Young tableaux on the $2 \times n$ rectangle is

$$\frac{(2n)!}{(n+1)!n!} = \frac{1}{n+1}\binom{2n}{n}. \tag{2.6}$$

Thus we will have proved that $|NC_2(2n)| = \frac{1}{n+1}\binom{2n}{n}$ if we can bijectively associate to each pair partition $\pi \in NC_2(2n)$ a standard Young tableaux on the $2 \times n$ rectangular grid. This is very easy to do. Simply take the "left-halves" of each pair in π and write them in increasing order in the cells of the first row. Then take the "right-halves" of each pair of π and write them in increasing order in the cells of the second row. Figure 2.4 shows the bijection between $NC_2(6)$ and standard Young tableaux on the 2×3 rectangle.

Definition 4. A self-adjoint random variable s with odd moments $\varphi(s^{2n+1}) = 0$ and even moments $\varphi(s^{2n}) = \sigma^{2n} C_n$, where C_n is the n-th Catalan number and $\sigma > 0$ is a constant, is called a *semi-circular element* of variance σ^2. In the case $\sigma = 1$, we call it the *standard* semi-circular element.

The argument we have just provided gives us the *free central limit theorem.*

Theorem 5. *If $(a_i)_{i \in \mathbb{N}}$ are self-adjoint, freely independent, and identically distributed with $\varphi(a_i) = 0$ and $\varphi(a_i^2) = \sigma^2$, then S_k converges in distribution to a semi-circular element of variance σ^2 as $k \to \infty$.*

This free central limit theorem was proved as one of the first results in free probability theory by Voiculescu already in [176]. His proof was much more operator theoretic; the proof presented here is due to Speicher [159] and was the first hint at a relation between free probability theory and the combinatorics of non-crossing partitions. (An early concrete version of the free central limit theorem, before the notion of freeness was isolated, appeared also in the work of Bożejko [43] in the context of convolution operators on free groups.)

Recall that in Chapter 1 it was shown that for a random matrix X_N chosen from $N \times N$ GUE we have that

$$\lim_{N\to\infty} E[\mathrm{tr}(X_N^n)] = \begin{cases} 0, & \text{if } n \text{ odd} \\ C_{n/2}, & \text{if } n \text{ even} \end{cases} \tag{2.7}$$

so that a GUE random matrix is a semi-circular element in the limit of large matrix size, $X_N \xrightarrow{\text{distr}} s$.

We can also define a family of semi-circular random variables.

Definition 6. Suppose $(\mathcal{A}, \varphi)$ is a $*$-probability space. A self-adjoint family $(s_i)_{i\in I} \subset \mathcal{A}$ is called a *semi-circular family* of covariance $C = (c_{ij})_{i,j\in I}$ if $C \geq 0$ and for any $n \geq 1$ and any n-tuple $i_1, \dots, i_n \in I$ we have

$$\varphi(s_{i_1} \cdots s_{i_n}) = \sum_{\pi\in NC_2(n)} \varphi_\pi[s_{i_1}, \dots, s_{i_n}],$$

where

$$\varphi_\pi[s_{i_1}, \dots, s_{i_n}] = \prod_{(p,q)\in\pi} c_{i_p i_q}.$$

If C is diagonal, then $(s_i)_{i\in I}$ is a free semi-circular family.

This is the free analogue of Wick's formula. In fact, using this language and our definition of convergence in distribution from Definition 1, it follows directly from Lemma 1.9 that if $X_1, \dots, X_r$ are matrices chosen independently from GUE, then, in the large N limit, they converge in distribution to a semi-circular family $s_1, \dots, s_r$ of covariance $c_{ij} = \delta_{ij}$.

Exercise 5. Show that if $\{x_1, \dots, x_n\}$ is a semi-circular family and $A = (a_{ij})$ is an invertible matrix with real entries, then $\{y_1, \dots, y_n\}$ is a semi-circular family where $y_i = \sum_j a_{ij} x_j$.

Exercise 6. Let $\{x_1, \dots, x_n\}$ be a semi-circular family such that for all i and j we have $\varphi(x_i x_j) = \varphi(x_j x_i)$. Show that by diagonalizing the covariance matrix we can find an orthogonal matrix $O = (o_{ij})$ such that $\{y_1, \dots, y_n\}$ is a free semi-circular family where $y_i = \sum_j o_{ij} x_j$.

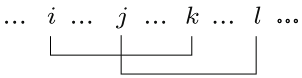


Fig. 2.5 A crossing in a partition

Exercise 7. Formulate and prove a multidimensional version of the free central limit theorem.

2.2 Non-crossing partitions and free cumulants

We begin by recalling some relevant definitions concerning non-crossing partitions from Section 1.8.

Definition 7. A partition $\pi \in \mathcal{P}(n)$ is called *non-crossing* if there do not exist numbers $i, j, k, l \in [n]$ with $i < j < k < l$ such that i and k are in the same block of π and j and l are in the same block of π, but i and j are not in the same block of π. The collection of all non-crossing partitions of $[n]$ was denoted $NC(n)$.

Figure 2.5 should make it clear what a *crossing* in a partition is; a non-crossing partition is a partition with no crossings.

Note that $\mathcal{P}(n)$ is partially ordered by

$$\pi_1 \leq \pi_2 \iff \text{each block of } \pi_1 \text{ is contained in a block of } \pi_2. \tag{2.8}$$

We also say that π_1 is a *refinement* of $\pi_2.NC(n)$ is a subset of $\mathcal{P}(n)$ and inherits this partial order, so $NC(n)$ is an induced sub-poset of $\mathcal{P}(n)$. In fact both are lattices; they have well-defined *join* $\vee$ and *meet* $\wedge$ operations (though the join of two non-crossing partitions in $NC(n)$ does not necessarily agree with their join when viewed as elements of $\mathcal{P}(n)$). Recall that the join $\pi_1 \vee \pi_2$ in a lattice is the smallest σ with the property that $\sigma \geq \pi_1$ and $\sigma \geq \pi_2$ and that the meet $\pi_1 \wedge \pi_2$ is the largest σ with the property that $\sigma \leq \pi_1$ and $\sigma \leq \pi_2$.

We now define the important *free cumulants* of a non-commutative probability space (A, φ). They were introduced by Speicher in [161]. For other notions of cumulants and the relation between them, see [11, 74, 117, 153].

Definition 8. Let $(\mathcal{A}, \varphi)$ be a non-commutative probability space. The corresponding *free cumulants* $\kappa_n : \mathcal{A}^n \to \mathbb{C}$ $(n \geq 1)$ are defined inductively in terms of moments by the *moment-cumulant formula*

$$\varphi(a_1 \cdots a_n) = \sum_{\pi \in NC(n)} \kappa_\pi(a_1, \ldots, a_n), \tag{2.9}$$

where, by definition, if $\pi = \{V_1, \dots, V_r\}$, then

$$\kappa_\pi(a_1, \dots, a_n) = \prod_{\substack{V \in \pi \\ V=(i_1,\dots,i_l)}} \kappa_l(a_{i_1}, \dots, a_{i_l}). \tag{2.10}$$

Remark 9. In Equation (2.10) and below, we always mean that the elements $i_1, \dots, i_l$ of V are in increasing order. Note that Equation (2.9) has a formulation using Möbius inversion which we might call the *cumulant-moment formula*. To present this we need the moment version of Equation (2.10). For a partition $\pi \in \mathcal{P}(n)$ with $\pi = \{V_1, \dots, V_r\}$, we set

$$\varphi_\pi(a_1, \dots, a_n) = \prod_{\substack{V \in \pi \\ V=(i_1,\dots,i_l)}} \varphi(a_{i_1} \cdots a_{i_l}). \tag{2.11}$$

We also need the Möbius function μ for $NC(n)$ (see [137, Lecture 10]). Then our cumulant-moment relation can be written

$$\kappa_n(a_1, \dots, a_n) = \sum_{\pi \in NC(n)} \mu(\pi, 1_n)\varphi_\pi(a_1, \dots, a_n). \tag{2.12}$$

One could use Equation (2.12) as the definition of free cumulants; however for practical calculations Equation (2.9) is usually easier to work with.

Example 10. (1) For $n = 1$, we have $\varphi(a_1) = \kappa_1(a_1)$, and thus

$$\kappa_1(a_1) = \varphi(a_1). \tag{2.13}$$

(2) For $n = 2$, we have

$$\varphi(a_1 a_2) = \kappa_{\{(1,2)\}}(a_1, a_2) + \kappa_{\{(1),(2)\}}(a_1, a_2) = \kappa_2(a_1, a_2) + \kappa_1(a_1)\kappa_1(a_2).$$

Since we know from the $n = 1$ calculation that $\kappa_1(a_1) = \varphi(a_1)$, this yields

$$\kappa_2(a_1, a_2) = \varphi(a_1 a_2) - \varphi(a_1)\varphi(a_2). \tag{2.14}$$

(3) For $n = 3$, we have

$$\begin{aligned}
\varphi(a_1 a_2 a_3) &= \kappa_{\{(1,2,3)\}}(a_1, a_2, a_3) + \kappa_{\{(1,2),(3)\}}(a_1, a_2, a_3) + \kappa_{\{(1),(2,3)\}}(a_1, a_2, a_3) \\
&\quad + \kappa_{\{(1,3),(2)\}}(a_1, a_2, a_3) + \kappa_{\{(1),(2),(3)\}}(a_1, a_2, a_3) \\
&= \kappa_3(a_1, a_2, a_3) + \kappa_2(a_1, a_2)\kappa_1(a_3) + \kappa_2(a_2, a_3)\kappa_1(a_1) \\
&\quad + \kappa_2(a_1, a_3)\kappa_1(a_2) + \kappa_1(a_1)\kappa_1(a_2)\kappa_1(a_3).
\end{aligned}$$

Thus we find that

$$\kappa_3(a_1, a_2, a_3) = \varphi(a_1a_2a_3) - \varphi(a_1)\varphi(a_2a_3) \\ - \varphi(a_2)\varphi(a_1a_3) - \varphi(a_3)\varphi(a_1a_2) + 2\varphi(a_1)\varphi(a_2)\varphi(a_3). \tag{2.15}$$

These three examples outline the general procedure of recursively defining κ_n in terms of the mixed moments. It is easy to see that κ_n is an n-linear function.

Exercise 8. (*i*) Show the following: if φ is a trace, then the cumulant κ_n is, for each $n \in \mathbb{N}$, invariant under cyclic permutations, i.e. for all $a_1, \ldots, a_n \in \mathcal{A}$, we have

$$\kappa_n(a_1, a_2, \ldots, a_n) = \kappa_n(a_2, \ldots, a_n, a_1).$$

(*ii*) Let us assume that all moments with respect to φ are invariant under all permutations of the entries, i.e. that we have for all $n \in \mathbb{N}$ and all $a_1, \ldots, a_n \in \mathcal{A}$ and all $\sigma \in S_n$ that $\varphi(a_{\sigma(1)} \cdots a_{\sigma(n)}) = \varphi(a_1 \cdots a_n)$. Is it then true that also the free cumulants κ_n ($n \in \mathbb{N}$) are invariant under all permutations?

Let us also point out how the definition appears when $a_1 = \cdots = a_n = a$, i.e. when all the random variables are the same. Then we have

$$\varphi(a^n) = \sum_{\pi \in NC(n)} \kappa_\pi(a, \ldots, a).$$

Thus if we write $\alpha_n^a := \varphi(a^n)$ and $\kappa_\pi^a := \kappa_\pi(a, \ldots, a)$, this reads

$$\alpha_n^a = \sum_{\pi \in NC(n)} \kappa_\pi^a. \tag{2.16}$$

Note the similarity to Equation (1.3) for classical cumulants.

Since the Catalan number is the number of non-crossing pairings of $[2n]$ as well as the number of non-crossing partitions of $[n]$, we can use Equation (2.16) to show that the cumulants of the standard semi-circle law are all 0 except $\kappa_2 = 1$.

Exercise 9. Use Equation (2.16) to show that for the standard semi-circle law all cumulants are 0, except κ_2 which equals 1.

As another demonstration of the simplifying power of the moment-cumulant formula (2.16), let us use the formula to find a simple expression for the moments and free cumulants of the *Marchenko-Pastur law*. This is a probability measure on $\mathbb{R}^+ \cup \{0\}$ that is as fundamental as the semi-circle law (see Section 4.5). Let $0 < c < \infty$ be a positive real number. For each c we shall construct a probability measure ν_c. Set $a = (1 - \sqrt{c})^2$ and $b = (1 + \sqrt{c})^2$. For $c \geq 1$, ν_c has as support the interval $[a, b]$ and the density $\sqrt{(b - x)(x - a)}/(2\pi x)$; that is

$$d\nu_c(x) = \frac{\sqrt{(b - x)(x - a)}}{2\pi x} dx.$$

For $0 < c < 1$, ν_c has the same density on $[a,b]$ and in addition has an atom at 0 of mass $1-c$; thus

$$d\nu_c(x) = (1-c)\delta_0 + \frac{\sqrt{(b-x)(x-a)}}{2\pi x}dx.$$

Note that when $c = 1$, $a = 0$ and the density has a "pole" of order 1/2 at 0 and thus is still integrable.

Exercise 10. In this exercise we shall show that ν_c is a probability measure for all c. Let $R = -x^2 + (a+b)x - ab$, and then write

$$\frac{\sqrt{R}}{x} = \frac{R}{x\sqrt{R}} = \frac{1}{2}\frac{-2x+(a+b)}{\sqrt{R}} + \frac{1}{2}\frac{a+b}{\sqrt{R}} - \frac{ab}{x\sqrt{R}}.$$

(i) Show that the integral of the first term on $[a,b]$ is 0.
(ii) Using the substitution $t = (x-(1+c))/\sqrt{c}$, show that the integral of the second term over $[a,b]$ is $\pi(a+b)/2$.
(iii) Let $u = (b-a)/(2ab)$, $v = (b+a)/(2ab)$ and $t = u^{-1}(v - x^{-1})$. With this substitution show that the integral of the third term over $[a,b]$ is $-\pi\sqrt{ab}$.
(iv) Using the first three parts, show that ν_c is a probability measure.

Definition 11. The *Marchenko-Pastur distribution* is the law with distribution ν_c with $0 < c < \infty$. We shall see in Exercise 11 that all free cumulants of ν_c are equal to c. By analogy with the classical cumulants of the Poisson distribution, ν_c is also called the *free Poisson law* (of rate c). We should also note that we have chosen a different normalization than that used by other authors in order to make the cumulants simple; see Remark 12 and Exercise 12 below.

Exercise 11. In this exercise we shall find the moments and free cumulants of the Marchenko-Pastur law.

(i) Let α_n be the n^{th} moment. Use the substitution $t = (x-(1+c))/\sqrt{c}$ to show that

$$\alpha_n = \sum_{k=0}^{[(n-1)/2]} \frac{1}{k+1}\binom{n-1}{2k}\binom{2k}{k}(1+c)^{n-2k-1}c^{1+k}.$$

(ii) Expand the expression $(1+c)^{n-2k-1}$ to obtain that

$$\alpha_n = \sum_{k=0}^{[(n-1)/2]}\sum_{l=k}^{n-k-1} \frac{(n-1)!}{k!\,(k+1)!\,(l-k)!\,(n-k-l-1)!}c^{l+1}.$$

(*iii*) Interchange the order of summation and use Vandermonde convolution ([79, (5.23)]) to show that

$$\alpha_n = \sum_{l=1}^{n} \frac{c^l}{n} \binom{n}{l-1} \binom{n}{l}.$$

(*iv*) Finally use the fact ([137, Cor. 9.13]) that $\frac{1}{n}\binom{n}{l-1}\binom{n}{l}$ is the number of non-crossing partitions of $[n]$ with l blocks to show that

$$\alpha_n = \sum_{\pi \in NC(n)} c^{\#(\pi)}.$$

Use this formula to show that $\kappa_n = c$ for all $n \geq 1$.

Remark 12. Given $y > 0$, let $a' = (1 - \sqrt{y})^2$ and $b' = (1 + \sqrt{y})^2$. Let ρ_y be the probability measure on $\mathbb{R}$ given by $\sqrt{(b'-t)(t-a')}/(2\pi y t)\, dt$ on $[a', b']$ when $y \leq 1$ and $(1 - y^{-1})\delta_0 + \sqrt{(b'-t)(t-a')}/(2\pi y t)\, dt$ on $\{0\} \cup [a', b']$ when $y > 1$. As above δ_0 is the Dirac mass at 0. This might be called the standard form of the Marchenko-Pastur law. In the exercise below, we shall see that ρ_y is related to ν_c in a simple way and the cumulants of ρ_y are not as simple as those of ν_c.

Exercise 12. Show that by setting $c = 1/y$ and making the substitution $t = x/c$ we have

$$\int x^k \, d\nu_c(x) = c^k \int t^k \, d\rho_y(t).$$

Show that the free cumulants of ρ_y are given by $\kappa_n = c^{1-n}$.

There is a combinatorial formula by Krawczyk and Speicher [111] for expanding cumulants whose arguments are products of random variables. For example, consider the expansion of $\kappa_2(a_1 a_2, a_3)$. This can be written as

$$\kappa_2(a_1 a_2, a_3) = \kappa_3(a_1, a_2, a_3) + \kappa_1(a_1)\kappa_2(a_2, a_3) + \kappa_2(a_1, a_3)\kappa_1(a_2). \tag{2.17}$$

A more complicated example is given by:

$$\begin{aligned}
&\kappa_2(a_1 a_2, a_3 a_4) \\
&\quad = \kappa_4(a_1, a_2, a_3, a_4) + \kappa_1(a_1)\kappa_3(a_2, a_3, a_4) + \kappa_1(a_2)\kappa_3(a_1, a_3, a_4) \\
&\qquad + \kappa_1(a_3)\kappa_3(a_1, a_2, a_4) + \kappa_1(a_4)\kappa_3(a_1, a_2, a_3) + \kappa_2(a_1, a_4)\kappa_2(a_2, a_3) \\
&\qquad + \kappa_2(a_1, a_3)\kappa_1(a_2)\kappa_1(a_4) + \kappa_2(a_1, a_4)\kappa_1(a_2)\kappa_1(a_3) \\
&\qquad + \kappa_1(a_1)\kappa_2(a_2, a_3)\kappa_1(a_4) + \kappa_1(a_1)\kappa_2(a_2, a_4)\kappa_1(a_3).
\end{aligned} \tag{2.18}$$

In general, the evaluation of a free cumulant with products of entries involves summing over all π which have the property that they connect all different product strings. Here is the precise formulation, for the proof we refer to [137, Theorem 11.12]. Note that this is the free counter part of the formula (1.16) for classical cumulants.

Theorem 13. *Suppose $n_1, \ldots, n_r$ are positive integers and $n = n_1 + \cdots + n_r$. Consider a non-commutative probability space $(\mathcal{A}, \varphi)$ and $a_1, a_2, \ldots, a_n \in \mathcal{A}$. Let*

$$A_1 = a_1 \cdots a_{n_1}, \quad A_2 = a_{n_1+1} \cdots a_{n_1+n_2}, \quad \ldots, \quad A_r = a_{n_1+\cdots+n_{r-1}+1} \cdots a_n.$$

Then

$$\kappa_r(A_1, \ldots, A_r) = \sum_{\substack{\pi \in NC(n) \\ \pi \vee \tau = 1_n}} \kappa_\pi(a_1, \ldots, a_n) \tag{2.19}$$

where the summation is over those $\pi \in NC(n)$ which connect the blocks corresponding to $A_1, \ldots, A_r$. More precisely, this means that $\pi \vee \tau = 1_n$ where

$$\tau = \{(1, \ldots, n_1), (n_1 + 1, \ldots, n_1 + n_2), \ldots, (n_1 + \cdots + n_{r-1} + 1, \ldots, n)\}$$

and $1_n = \{(1, 2, \ldots, n)\}$ is the partition with only one block.

Exercise 13. (*i*) Let $\tau = \{(1, 2), (3)\}$. List all $\pi \in NC(3)$ such that $\pi \vee \tau = 1_3$. Check that these are exactly the terms appearing on the right-hand side of Equation (2.17).

(*ii*) Let $\tau = \{(1, 2), (3, 4)\}$. List all $\pi \in NC(4)$ such that $\pi \vee \tau = 1_4$. Check that these are exactly the terms on the right-hand side of Equation (2.18).

The most important property of free cumulants is that we may characterize free independence by the vanishing of “mixed” cumulants. Let $(\mathcal{A}, \varphi)$ be a non-commutative probability space and $\mathcal{A}_1, \ldots, \mathcal{A}_s \subset \mathcal{A}$ unital subalgebras. A cumulant $\kappa_n(a_1, a_2, \ldots, a_n)$ is *mixed* if each a_i is in one of the subalgebras, but $a_1, a_2, \ldots, a_n$ do not all come from the same subalgebra.

Theorem 14. *The subalgebras $\mathcal{A}_1, \ldots, \mathcal{A}_s$ are free if and only if all mixed cumulants vanish.*

The proof of this theorem relies on formula (2.19) and on the following proposition which is a special case of Theorem 14. For the details of the proof of Theorem 14, we refer again to [137, Theorem 11.15].

Proposition 15. *Let $(\mathcal{A}, \varphi)$ be a non-commutative probability space and let κ_n, $n \geq 1$ be the corresponding free cumulants. For $n \geq 2$, $\kappa_n(a_1, \ldots, a_n) = 0$ if $1 \in \{a_1, \ldots, a_n\}$.*

Proof: We consider the case where the last argument a_n is equal to 1 and proceed by induction on n.

For $n = 2$,

$$\kappa_2(a, 1) = \varphi(a1) - \varphi(a)\varphi(1) = 0.$$

So the base step is done.

Now assume for the induction hypothesis that the result is true for all $1 \leq k < n$. We have that

$$\begin{aligned}\varphi(a_1 \cdots a_{n-1}1) &= \sum_{\pi \in NC(n)} \kappa_\pi(a_1, \ldots, a_{n-1}, 1) \\ &= \kappa_n(a_1, \ldots, a_{n-1}, 1) + \sum_{\substack{\pi \in NC(n) \\ \pi \neq 1_n}} \kappa_\pi(a_1, \ldots, a_{n-1}, 1).\end{aligned}$$

According to our induction hypothesis, a partition $\pi \neq 1_n$ can have $\kappa_\pi(a_1, \ldots, a_{n-1}, 1)$ different from zero only if (n) is a one-element block of π, i.e. $\pi = \sigma \cup \{(n)\}$ for some $\sigma \in NC(n-1)$. For such a partition, we have

$$\kappa_\pi(a_1, \ldots, a_{n-1}, 1) = \kappa_\sigma(a_1, \ldots, a_{n-1})\kappa_1(1) = \kappa_\sigma(a_1, \ldots, a_{n-1}),$$

hence

$$\begin{aligned}\varphi(a_1 \cdots a_{n-1}1) &= \kappa_n(a_1, \ldots, a_{n-1}, 1) + \sum_{\sigma \in NC(n-1)} \kappa_\sigma(a_1, \ldots, a_{n-1}) \\ &= \kappa_n(a_1, \ldots, a_{n-1}, 1) + \varphi(a_1 \cdots a_{n-1}).\end{aligned}$$

Since $\varphi(a_1 \cdots a_{n-1}1) = \varphi(a_1 \cdots a_{n-1})$, we have proved that $\kappa_n(a_1, \ldots, a_{n-1}, 1) = 0$. $\square$

Whereas Theorem 14 gives a useful characterization for the freeness of subalgebras, its direct application to the case of random variables would not yield a satisfying characterization in terms of the vanishing of mixed cumulants in the subalgebras generated by the variables. By invoking again the product formula for free cumulants, Theorem 13, it is quite straightforward to get the following much more useful characterization in terms of mixed cumulants of the variables.

Theorem 16. *Let $(\mathcal{A}, \varphi)$ be a non-commutative probability space. The random variables $a_1, \ldots, a_s \in \mathcal{A}$ are free if and only if all mixed cumulants of the $a_1, \ldots, a_s$ vanish. That is, $a_1, \ldots, a_s$ are free if and only if whenever we choose $i_1, \ldots, i_n \in \{1, \ldots, s\}$ in such a way that $i_k \neq i_l$ for some $k, l \in [n]$, then $\kappa_n(a_{i_1}, \ldots, a_{i_n}) = 0$.*

2.3 Products of free random variables

We want to understand better the calculation rule for mixed moments of free variables. Thus we will now derive the basic combinatorial description for such mixed moments.

Let $\{a_1, \ldots, a_r\}$ and $\{b_1, \ldots, b_r\}$ be free random variables, and consider

$$\varphi(a_1 b_1 a_2 b_2 \cdots a_r b_r) = \sum_{\pi \in NC(2r)} \kappa_\pi(a_1, b_1, a_2, b_2, \ldots, a_r, b_r).$$

Since the a's are free from the b's, we only need to sum over those partitions π which do not connect the a's with the b's. Each such partition may be written as $\pi = \pi_a \cup \pi_b$, where π_a denotes the blocks consisting of a's and π_b the blocks consisting of b's. Hence by the definition of free cumulants

$$\begin{aligned}\varphi(a_1 b_1 a_2 b_2 \cdots a_r b_r) &= \sum_{\pi_a \cup \pi_b \in NC(2r)} \kappa_{\pi_a}(a_1, \ldots, a_r) \cdot \kappa_{\pi_b}(b_1, \ldots, b_r) \\ &= \sum_{\pi_a \in NC(r)} \kappa_{\pi_a}(a_1, \ldots, a_r) \cdot \Bigg(\sum_{\substack{\pi_b \in NC(r) \\ \pi_a \cup \pi_b \in NC(2r)}} \kappa_{\pi_b}(b_1, \ldots, b_r) \Bigg).\end{aligned}$$

It is now easy to see that, for a given $\pi_a \in NC(r)$, there exists a biggest $\sigma \in NC(r)$ with the property that $\pi_a \cup \sigma \in NC(2r)$. This σ is called the *Kreweras complement* of π_a and is denoted by $K(\pi_a)$; see [137, Def. 9.21]. This $K(\pi_a)$ is given by connecting as many b's as possible in a non-crossing way without getting crossings with the blocks of π_a. The mapping K is an order-reversing bijection on the lattice $NC(r)$.

But then the summation condition on the internal sum above is equivalent to the condition $\pi_b \leq K(\pi_a)$. Summing κ_π over all $\pi \in NC(r)$ gives the corresponding r-th moment, which extends easily to

$$\sum_{\substack{\pi \in NC(r) \\ \pi \leq \sigma}} \kappa_\pi(b_1, \ldots, b_r) = \varphi_\sigma(b_1, \ldots, b_r),$$

where φ_σ denotes, in the same way as in κ_π, the product of moments along the blocks of σ; see Equation (2.11).

Thus we get as the final conclusion of our calculations that

$$\varphi(a_1 b_1 a_2 b_2 \cdots a_r b_r) = \sum_{\pi \in NC(r)} \kappa_\pi(a_1, \ldots, a_r) \cdot \varphi_{K(\pi)}(b_1, \ldots, b_r). \tag{2.20}$$

Let us consider some simple examples for this formula. For $r = 1$, there is only one $\pi \in NC(1)$, which is its own complement, and we get

$$\varphi(a_1 b_1) = \kappa_1(a_1)\varphi(b_1).$$

As $\kappa_1 = \varphi$, this gives the usual factorization formula

$$\varphi(a_1 b_1) = \varphi(a_1)\varphi(b_1).$$

For $r = 2$, there are two elements in $NC(2)$, $\vert\,\vert$ and $\sqcup$, and we have

$$K(\vert\,\vert) = \sqcup \qquad \text{and} \qquad K(\sqcup) = \vert\,\vert$$

and the formula above gives

$$\varphi(a_1 b_1 a_2 b_2) = \kappa_2(a_1, a_2)\varphi(b_1)\varphi(b_2) + \kappa_1(a_1)\kappa_1(a_2)\varphi(b_1 b_2).$$

With $\kappa_1(a) = \varphi(a)$ and $\kappa_2(a_1, a_2) = \varphi(a_1 a_2) - \varphi(a_1)\varphi(a_2)$, this reproduces formula (1.14).

The formula above is not symmetric between the a's and the b's (the former appear with cumulants, the latter with moments). Of course, one can also exchange the roles of a and b, in which case one ends up with

$$\varphi(a_1 b_1 a_2 b_2 \cdots a_r b_r) = \sum_{\pi \in NC(r)} \varphi_{K^{-1}(\pi)}(a_1, \ldots, a_r) \cdot \kappa_\pi(b_1, \ldots, b_r). \tag{2.21}$$

Note that K^2 is not the identity, but a cyclic rotation of π.

Formulas (2.20) and (2.21) are particularly useful when one of the sets of variables has simple cumulants, as is the case for semi-circular random variables $b_i = s$. Then only the second cumulants $\kappa_2(s, s) = 1$ are non-vanishing, i.e. in effect the sum is only over non-crossing pairings. Thus, if s is semi-circular and free from $\{a_1, \ldots, a_r\}$, then we have

$$\varphi(a_1 s a_2 s \cdots a_r s) = \sum_{\pi \in NC_2(r)} \varphi_{K^{-1}(\pi)}(a_1, \ldots, a_r). \tag{2.22}$$

Let us also note in passing that one can rewrite the Equations (2.20) and (2.21) above in the symmetric form (see [137, (14.4)])

$$\kappa_r(a_1 b_1, a_2 b_2, \ldots, a_r b_r) = \sum_{\pi \in NC(r)} \kappa_\pi(a_1, \ldots, a_r) \cdot \kappa_{K(\pi)}(b_1, \ldots, b_r). \tag{2.23}$$

2.4 Functional relation between moment series and cumulant series

Notice how much more efficient the result on the description of freeness in terms of cumulants is in checking freeness of random variables than the original definition of free independence. In the cumulant framework, we can forget about centredness and weaken "alternating" to "mixed". Also, the problem of adding two freely independent random variables becomes easy on the level of free cumulants. If $a, b \in (\mathcal{A}, \varphi)$ are free with respect to φ, then

$$\begin{aligned}\kappa_n^{a+b} &= \kappa_n(a+b, \dots, a+b)\\ &= \kappa_n(a, \dots, a) + \kappa_n(b, \dots, b) + (\text{mixed cumulants in } a, b)\\ &= \kappa_n^a + \kappa_n^b.\end{aligned}$$

Thus the problem of calculating moments is shifted to the relation between cumulants and moments. We already know that the moments are polynomials in the cumulants, according to the moment-cumulant formula (2.16), but we want to put this relationship into a framework more amenable to performing calculations.

For any $a \in \mathcal{A}$, let us consider formal power series in an indeterminate z defined by

$$M(z) = 1 + \sum_{n=1}^{\infty} \alpha_n^a z^n, \qquad \textit{moment series of } a$$

$$C(z) = 1 + \sum_{n=1}^{\infty} \kappa_n^a z^n, \qquad \textit{cumulant series of } a.$$

We want to translate the moment-cumulant formula (2.16) into a statement about the relationship between the moment and cumulant series.

Proposition 17. *The relation between the moment series $M(z)$ and the cumulant series $C(z)$ of a random variable is given by*

$$M(z) = C(zM(z)). \tag{2.24}$$

Proof: The idea is to sum first over the possibilities for the block of π containing 1, as in the derivation of the recurrence for C_n. Suppose that the first block of π looks like $V = \{1, v_2, \dots, v_s\}$, where $1 < v_1 < \dots < v_s \le n$. Then we build up the rest of the partition π out of smaller "nested" non-crossing partitions $\pi_1, \dots, \pi_s$ with $\pi_1 \in NC(\{2, \dots, v_2 - 1\})$, $\pi_2 \in NC(\{v_2 + 1, \dots, v_3 - 1\})$, etc. Hence if we denote $i_1 = |\{2, \dots, v_2 - 1\}|$, $i_2 = |\{v_2 + 1, \dots, v_3 - 1\}|$, etc., then we have

$$\begin{aligned}
\alpha_n &= \sum_{s=1}^{n} \sum_{\substack{i_1,\ldots,i_s \geq 0 \\ s+i_1+\cdots+i_s=n}} \sum_{\pi = V \cup \pi_1 \cup \cdots \cup \pi_s} \kappa_s \kappa_{\pi_1} \cdots \kappa_{\pi_s} \\
&= \sum_{s=1}^{n} \sum_{\substack{i_1,\ldots,i_s \geq 0 \\ s+i_1+\cdots+i_s=n}} \kappa_s \Big(\sum_{\pi_1 \in NC(i_1)} \kappa_{\pi_1} \Big) \cdots \Big(\sum_{\pi_s \in NC(i_s)} \kappa_{\pi_s} \Big) \\
&= \sum_{s=1}^{n} \sum_{\substack{i_1,\ldots,i_s \geq 0 \\ s+i_1+\cdots+i_s=n}} \kappa_s \alpha_{i_1} \cdots \alpha_{i_s}.
\end{aligned}$$

Thus we have

$$\begin{aligned}
1 + \sum_{n=1}^{\infty} \alpha_n z^n &= 1 + \sum_{n=1}^{\infty} \sum_{s=1}^{n} \sum_{\substack{i_1,\ldots,i_s \geq 0 \\ s+i_1+\cdots+i_s=n}} \kappa_s z^s \alpha_{i_1} z^{i_1} \ldots \alpha_{i_s} z^{i_s} \\
&= 1 + \sum_{s=1}^{\infty} \kappa_s z^s \Big(\sum_{i=0}^{\infty} \alpha_i z^i \Big)^s.
\end{aligned}$$

□

Now consider the *Cauchy transform* of a:

$$G(z) := \varphi\Big(\frac{1}{z-a}\Big) = \sum_{n=0}^{\infty} \frac{\varphi(a^n)}{z^{n+1}} = \frac{1}{z} M(1/z) \tag{2.25}$$

and the R-*transform* of a defined by

$$R(z) := \frac{C(z)-1}{z} = \sum_{n=0}^{\infty} \kappa_{n+1}^a z^n. \tag{2.26}$$

Also put $K(z) = R(z) + \frac{1}{z} = \frac{C(z)}{z}$. Then we have the relations

$$K(G(z)) = \frac{1}{G(z)} C(G(z)) = \frac{1}{G(z)} C\left(\frac{1}{z} M\left(\frac{1}{z}\right)\right) = \frac{1}{G(z)} z G(z) = z.$$

Note that M and C are in $\mathbb{C}[\![z]\!]$, the ring of formal power series in z, $G \in \mathbb{C}[\![\frac{1}{z}]\!]$, and $K \in \mathbb{C}(\!(z)\!)$, the ring of formal Laurent series in z, i.e. $zK(z) \in \mathbb{C}[\![z]\!]$. Thus $K \circ G \in \mathbb{C}(\!(\frac{1}{z})\!)$ and $G \circ K \in \mathbb{C}[\![z]\!]$. We then also have $G(K(z)) = z$.

Thus we recover the following theorem of Voiculescu, which is the main result on the R-transform. Voiculescu's original proof in [177] was much more operator theoretic. One should also note that this computational machinery for the R-transform was also found independently and about the same time by Woess [204, 205], Cartwright and Soardi [49], and McLaughlin [125], in a more restricted setting of random walks on free product of groups. Our presentation here is based on the approach of Speicher in [161].

Theorem 18. *For a random variable a, let $G_a(z)$ be its Cauchy transform, and define its R-transform $R_a(z)$ by*

$$G_a[R_a(z) + 1/z] = z. \tag{2.27}$$

Then, for a and b freely independent, we have

$$R_{a+b}(z) = R_a(z) + R_b(z). \tag{2.28}$$

Let us write, for a and b free, the above as:

$$z = G_{a+b}[R_{a+b}(z) + 1/z] = G_{a+b}[R_a(z) + R_b(z) + 1/z]. \tag{2.29}$$

If we now put $w := R_{a+b}(z) + 1/z$, then we have $z = G_{a+b}(w)$ and we can continue Equation (2.29) as:

$$G_{a+b}(w) = z = G_a[R_a(z) + 1/z] = G_a[w - R_b(z)] = G_a\big[w - R_b[G_{a+b}(w)]\big].$$

Thus we get the *subordination* functions ω_a and ω_b given by

$$\omega_a(z) = z - R_b[G_{a+b}(z)] \qquad \text{and} \qquad \omega_b(z) = z - R_a[G_{a+b}(z)]. \tag{2.30}$$

We have $\omega_a, \omega_b \in \mathbb{C}((\frac{1}{z}))$, so $G_a \circ \omega_a \in \mathbb{C}[[\frac{1}{z}]]$. These satisfy the subordination relations

$$G_{a+b}(z) = G_a[\omega_a(z)] = G_b[\omega_b(z)]. \tag{2.31}$$

We say that G_{a+b} is subordinate to both G_a and G_b. The name comes from the theory of univalent functions; see [65, Ch. 6] for a general discussion.

Exercise 14. Show that $\omega_a(z) + \omega_b(z) - 1/G_a(\omega_a(z)) = z$.

Exercise 15. Suppose we have formal Laurent series $\omega_a(z)$ and $\omega_b(z)$ in $\frac{1}{z}$ such that

$$G_a(\omega_a(z)) = G_b(\omega_b(z)) \quad \text{and} \quad \omega_a(z) + \omega_b(z) - 1/G_a(\omega_a(z)) = z. \tag{2.32}$$

Let G be the formal power series $G(z) = G_a(\omega_a(z))$ and $R(z) = G^{\langle -1 \rangle}(z) - z^{-1}$. ($G^{\langle -1 \rangle}$ denotes here the inverse under composition of G.) By replacing z by $G^{\langle -1 \rangle}(z)$ in the second equation of (2.32), show that $R(z) = R_a(z) + R_b(z)$. These equations can thus be used to define the distribution of the sum of two free random variables.

At the moment these are identities on the level of formal power series. In the next chapter, we will elaborate on their interpretation as identities of analytic functions; see Theorem 3.43.

2.5 Subordination and the non-commutative derivative

One might wonder about the relevance of the subordination formulation in (2.31). Since it has become more and more evident that the subordination formulation of free convolution is in many cases preferable to the (equivalent) description in terms of the R-transform, we want to give here some idea why subordination is a very natural concept in the context of free probability. When subordination appeared in this context first in papers of Voiculescu [181] and Biane [34], it was more an ad hoc construction – its real nature was only revealed later in the paper [190] of Voiculescu, where he related it to the non-commutative version of the derivative operation.

We will now introduce the basics of this non-commutative derivative; as before in this chapter, we will ignore all analytic questions and just deal with formal power series. In Chapter 8 we will have more to say about the analytic properties of the non-commutative derivatives.

Let $\mathbb{C}\langle x\rangle$ be the algebra of polynomials in the variable x. Then we define the *non-commutative derivative* ∂_x as a linear mapping $\partial_x : \mathbb{C}\langle x\rangle \to \mathbb{C}\langle x\rangle \otimes \mathbb{C}\langle x\rangle$ by the requirements that it satisfies the Leibniz rule

$$\partial_x(qp) = \partial_x(q) \cdot 1 \otimes p + q \otimes 1 \cdot \partial_x(p)$$

and by

$$\partial_x 1 = 0, \qquad \partial_x x = 1 \otimes 1.$$

This means that it is given more explicitly as the linear extension of

$$\partial_x x^n = \sum_{k=0}^{n-1} x^k \otimes x^{n-1-k}. \tag{2.33}$$

We can also (and will) extend this definition from polynomials to infinite formal power series.

Exercise 16. (i) Let, for some $z \in \mathbb{C}$ with $z \neq 0$, f be the formal power series

$$f(x) = \frac{1}{z-x} = \sum_{n=0}^{\infty} \frac{x^n}{z^{n+1}}.$$

Show that we have then $\partial_x f = f \otimes f$.

(ii) Let f be a formal power series in x with the property that $\partial_x f = f \otimes f$. Show that f must then be either zero or of the form $f(x) = 1/(z-x)$ for some $z \in \mathbb{C}$, with $z \neq 0$.

We will now consider polynomials and formal power series in two non-commuting variables x and y. In this context, we still have the notion of ∂_x

(and also of ∂_y), and now their character as "partial" derivatives becomes apparent. Namely, we define $\partial_x : \mathbb{C}\langle x, y\rangle \to \mathbb{C}\langle x, y\rangle \otimes \mathbb{C}\langle x, y\rangle$ by the requirements that it should be a derivation, i.e. satisfy the Leibniz rule, and by the prescriptions:

$$\partial_x x = 1 \otimes 1, \qquad \partial_x y = 0, \qquad \partial_x 1 = 0.$$

For a monomial $x_{i_1} \cdots x_{i_n}$ in x and y (where we put $x_1 := x$ and $x_2 := y$), this means explicitly

$$\partial_x x_{i_1} \cdots x_{i_n} = \sum_{k=1}^{n} \delta_{1 i_k} x_{i_1} \cdots x_{i_{k-1}} \otimes x_{i_{k+1}} \cdots x_{i_n}. \tag{2.34}$$

Again it is clear that we can extend this definition also to formal power series in non-commuting variables.

Let us note that we may define the derivation ∂_{x+y} on $\mathbb{C}\langle x + y\rangle$ exactly as we did ∂_x. Namely, $\partial_{x+y}(1) = 0$ and $\partial_{x+y}(x + y) = 1 \otimes 1$. Note that ∂_{x+y} can be extended to all of $\mathbb{C}\langle x, y\rangle$ but not in a unique way unless we specify another basis element. Since $\mathbb{C}\langle x + y\rangle \subset \mathbb{C}\langle x, y\rangle$, we may apply ∂_x to $\mathbb{C}\langle x + y\rangle$ and observe that $\partial_x(x + y) = 1 \otimes 1 = \partial_{x+y}(x + y)$. Thus

$$\partial_x (x + y)^n = \sum_{k=1}^{n} (x + y)^{k-1} \otimes (x + y)^{n-k} = \partial_{x+y}(x + y)^n.$$

Hence

$$\partial_x|_{\mathbb{C}\langle x+y\rangle} = \partial_{x+y}. \tag{2.35}$$

If we are given a polynomial $p(x, y) \in \mathbb{C}\langle x, y\rangle$, then we will also consider $\mathrm{E}_x[p(x, y)]$, the conditional expectation of $p(x, y)$ onto a function of just the variable x, which should be the best approximation to p among such functions. There is no algebraic way of specifying what best approximation means; we need a state φ on the $*$-algebra generated by self-adjoint elements x and y for this. Given such a state, we will require that the difference between $p(x, y)$ and $\mathrm{E}_x[p(x, y)]$ cannot be detected by functions of x alone; more precisely, we ask that

$$\varphi\big(q(x) \cdot \mathrm{E}_x[p(x, y)]\big) = \varphi\big(q(x) \cdot p(x, y)\big) \tag{2.36}$$

for all $q \in \mathbb{C}\langle x\rangle$. If we are going from the polynomials $\mathbb{C}\langle x, y\rangle$ over to the Hilbert space completion $L^2(x, y, \varphi)$ with respect to the inner product given by $\langle f, g\rangle := \varphi(g^* f)$, then this amounts just to an orthogonal projection from the space $L^2(x, y, \varphi)$ onto the subspace $L^2(x, \varphi)$ generated by polynomials in the variable x. (Let us assume that φ is positive and faithful so that we get an inner product.) Thus, on the Hilbert space level, the existence and uniqueness of $\mathrm{E}_x[p(x, y)]$ are clear. In general, though, it might not be the case that the projection of a polynomial in x and

y is a polynomial in x – it will just be an L^2-function. If we assume, however, that x and y are free, then we claim that this projection maps polynomials to polynomials. In fact for this construction to work at the algebraic level we only need assume that $\varphi|_{\mathbb{C}\langle x\rangle}$ is non-degenerate as this shows that E_x is well defined by (2.36). It is clear from Equation (2.36) that $\varphi(\mathrm{E}_x(a)) = \varphi(a)$ for all $a \in \mathbb{C}\langle x, y\rangle$.

Let us consider some examples. Assume that x and y are free. Then it is clear that we have

$$\mathrm{E}_x[x^n y^m] = x^n \varphi(y^m)$$

and more generally

$$\mathrm{E}_x[x^{n_1} y^m x^{n_2}] = x^{n_1+n_2}\varphi(y^m).$$

It is not so clear what $\mathrm{E}_x[yxyx]$ might be. Before giving the general rule, let us make some simple observations.

Exercise 17. Let $\mathcal{A}_1 = \mathbb{C}\langle x\rangle$ and $\mathcal{A}_2 = \mathbb{C}\langle y\rangle$ with x and y free and $\varphi|_{\mathcal{A}_1}$ non-degenerate.

(*i*) Show that $\mathrm{E}_x[\mathring{\mathcal{A}}_2] = 0$.
(*ii*) For $\alpha_1, \dots, \alpha_n \in \{1,2\}$ with $\alpha_1 \neq \cdots \neq \alpha_n$ and $n \geq 2$, show that $\mathrm{E}_x[\mathring{\mathcal{A}}_{\alpha_1} \cdots \mathring{\mathcal{A}}_{\alpha_n}] = 0$.

Exercise 18. Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be as in Exercise 17. Since $\mathcal{A}_1$ and $\mathcal{A}_2$ are free, we can use Equation (1.12) from Exercise 1.9 to write

$$\mathcal{A}_1 \vee \mathcal{A}_2 = \mathcal{A}_1 \oplus \mathring{\mathcal{A}}_2 \oplus \sum_{n\geq 2}^{\oplus} \sum_{\alpha_1\neq\cdots\neq\alpha_n}^{\oplus} \mathring{\mathcal{A}}_{\alpha_1}\mathring{\mathcal{A}}_{\alpha_2}\cdots\mathring{\mathcal{A}}_{\alpha_n}.$$

We have just shown that if E_x is a linear map satisfying Equation (2.36), then E_x is the identity on the first summand and 0 on all remaining summands. Show that by defining E_x this way we get the existence of a linear mapping from $\mathcal{A}_1 \vee \mathcal{A}_2$ to $\mathcal{A}_1$ satisfying Equation (2.36). An easy consequence of this is that for $q_1(x), q_2(x) \in \mathbb{C}\langle x\rangle$ and $p(x,y) \in \mathbb{C}\langle x, y\rangle$ we have $\mathrm{E}_x[q_1(x)p(x,y)q_2(x)] = q_1(x)\mathrm{E}_x[p(x,y)]q_2(x)$.

Let $a_1 = y^{n_1}, a_2 = y^{n_2}$ and $b = x^{m_1}$. To compute $\mathrm{E}_x(y^{n_1}x^{m_1}y^{n_2})$ we follow the same centring procedure used to compute $\varphi(a_1ba_2)$ in Section 1.12. From Exercise 17 we see that

$$\begin{aligned}
\mathrm{E}_x[a_1ba_2] &= \mathrm{E}_x[\mathring{a}_1ba_2] + \varphi(a_1)b\varphi(a_2) \\
&= \mathrm{E}_x[\mathring{a}_1\mathring{b}a_2] + \varphi(\mathring{a}_1a_2)\varphi(b) + \varphi(a_1)b\varphi(a_2) \\
&= \varphi(\mathring{a}_1a_2)\varphi(b) + \varphi(a_1)b\varphi(a_2) \\
&= \varphi(a_1a_2)\varphi(b) - \varphi(a_1)\varphi(b)\varphi(a_2) + \varphi(a_1)b\varphi(a_2).
\end{aligned}$$

Thus

$$\mathrm{E}_x[y^{n_1}x^{m_1}y^{n_2}x^{m_2}] = \varphi(y^{n_1+n_2})\varphi(x^{m_1})x^{m_2} + \varphi(y^{n_1})x^{m_1}\varphi(y^{n_2})x^{m_2} \\ - \varphi(y^{n_1})\varphi(x^{m_1})\varphi(y^{n_2})x^{m_2}.$$

The following theorem (essentially in the work [34] of Biane) gives the general recipe for calculating such expectations. As usual the formulas are simplified by using cumulants. To give the rule, we need the following bit of notation. Given $\sigma \in \mathcal{P}(n)$ and $a_1, \dots, a_n \in \mathcal{A}$, we define $\tilde{\varphi}_\sigma(a_1, \dots, a_n)$ in the same way as φ_σ in Equation (2.11) except we do not apply φ to the last block, i.e. the block containing n. For example, if $\sigma = \{(1,3,4),(2,6),(5)\}$, then $\tilde{\varphi}_\sigma(a_1,a_2,a_3,a_4,a_5,a_6) = \varphi(a_1a_3a_4)\varphi(a_5)a_2a_6$. More explicitly, for $\sigma = \{V_1, \dots, V_s\} \in NC(r)$ with $r \in V_s$, we put

$$\tilde{\varphi}_\sigma(a_1, \dots, a_r) = \varphi\Big(\prod_{i_1 \in V_1} a_{i_1}\Big) \cdots \varphi\Big(\prod_{i_{s-1} \in V_{s-1}} a_{i_{s-1}}\Big) \cdot \prod_{i_s \in V_s} a_{i_s}.$$

Theorem 19. *Let x and y be free. Then for $r \geq 1$ and $n_1, m_1, \dots, n_r, m_r \geq 0$, we have*

$$E_x[y^{n_1}x^{m_1} \cdots y^{n_r}x^{m_r}] = \sum_{\pi \in NC(r)} \kappa_\pi(y^{n_1}, \dots, y^{n_r}) \cdot \tilde{\varphi}_{K(\pi)}(x^{m_1}, \dots, x^{m_r}). \tag{2.37}$$

Let us check that this agrees with our previous calculation of $\mathrm{E}_x[y^{n_1}x^{m_1}y^{n_2}x^{m_2}]$.

$$\begin{aligned} &\mathrm{E}_x[y^{n_1}x^{m_1}y^{n_2}x^{m_2}] \\ &\quad = \kappa_{\{(1,2)\}}(y^{n_1}, y^{n_2}) \cdot \tilde{\varphi}_{\{(1),(2)\}}(x^{m_1}, x^{m_2}) + \kappa_{\{(1),(2)\}}(y^{n_1}, y^{n_2}) \cdot \tilde{\varphi}_{\{(1,2)\}}(x^{m_1}, x^{m_2}) \\ &\quad = \kappa_2(y^{n_1}, y^{n_2})\varphi(x^{m_1})x^{m_2} + \kappa_1(y^{n_1})\kappa_1(y^{n_2})x^{m_1+m_2} \\ &\quad = \big(\varphi(y^{n_1+n_2}) - \varphi(y^{n_1})\varphi(y^{n_2})\big)\varphi(x^{m_1}) \cdot x^{m_2} + \varphi(y^{n_1})\varphi(y^{n_2}) \cdot x^{m_1+m_2}. \end{aligned}$$

The proof of the theorem is outlined in the exercise below.

Exercise 19. (*i*) Given $\pi \in NC(n)$, let π' be the non-crossing partition of $[n'] = \{0, 1, 2, 3, \dots, n\}$ obtained by joining 0 to the block of π containing n. For $a_0, a_1, \dots, a_n \in \mathcal{A}$, show that $\varphi_{\pi'}(a_0, a_1, a_2, \dots, a_n) = \varphi(a_0\tilde{\varphi}_\pi(a_1, \dots, a_n))$.

(*ii*) Suppose that $\mathcal{A}_1, \mathcal{A}_2 \subset \mathcal{A}$ are unital subalgebras of $\mathcal{A}$ which are free with respect to the state φ. Let $x_0, x_1, \dots, x_n \in \mathcal{A}_1$ and $y_1, y_2, \dots, y_n \in \mathcal{A}_2$. Show that

$$\varphi(x_0y_1x_1y_2x_2 \cdots y_nx_n) = \sum_{\pi \in NC(n)} \kappa_\pi(y_1, \dots, y_n)\varphi_{K(\pi)'}(x_0, x_1, \dots, x_n).$$

Prove Theorem 19 by showing that with the expression given in (2.37) one has for all $m \geq 0$

$$\varphi\big(x^m \cdot \mathrm{E}_x[y^{n_1}x^{m_1}\cdots y^{n_r}x^{m_r}]\big) = \varphi\big(x^m \cdot y^{n_1}x^{m_1}\cdots y^{n_r}x^{m_r}\big).$$

Exercise 20. Use the method of Exercise 19 to work out $\mathrm{E}_x[x^{m_1}y^{n_1}\cdots x^{m_r}y^{n_r}]$.

By linear extension of Equation (2.37), one can thus get the projection onto one variable x of any non-commutative polynomial or formal power series in two free variables x and y. We now want to identify the projection of resolvents in $x + y$. To achieve this we need a crucial intertwining relation between the partial derivative and the conditional expectation.

Lemma 20. *Suppose φ is a state on $\mathbb{C}\langle x, y\rangle$ such that x and y are free and $\varphi|_{\mathbb{C}\langle x\rangle}$ is non-degenerate. Then*

$$\mathrm{E}_x \otimes \mathrm{E}_x \circ \partial_{x+y}|_{\mathbb{C}\langle x+y\rangle} = \partial_x \circ \mathrm{E}_x|_{\mathbb{C}\langle x+y\rangle}. \tag{2.38}$$

Proof: We let $\mathcal{A}_1 = \mathbb{C}\langle x\rangle$ and $\mathcal{A}_2 = \mathbb{C}\langle y\rangle$. We use the decomposition from Exercise 1.9

$$\mathcal{A}_1 \vee \mathcal{A}_2 \ominus \mathcal{A}_1 = \mathring{\mathcal{A}}_2 \oplus \sum_{n\geq 2}^{\oplus} \sum_{\alpha_1\neq\cdots\neq\alpha_n}^{\oplus} \mathring{\mathcal{A}}_{\alpha_1}\cdots\mathring{\mathcal{A}}_{\alpha_n}$$

and examine the behaviour of $\mathrm{E}_x \otimes \mathrm{E}_x \circ \partial_x$ on each summand. We know that ∂_x is 0 on $\mathring{\mathcal{A}}_2$ by definition. For $n \geq 2$

$$\begin{aligned}
&\mathrm{E}_x \otimes \mathrm{E}_x \circ \partial_x(\mathring{\mathcal{A}}_{\alpha_1}\cdots\mathring{\mathcal{A}}_{\alpha_n})\\
&\subseteq \sum_{k=1}^{n} \delta_{1,\alpha_k}\mathrm{E}_x(\mathring{\mathcal{A}}_{\alpha_1}\cdots\mathring{\mathcal{A}}_{\alpha_{k-1}}(\mathbb{C}1 \oplus \mathring{\mathcal{A}}_{\alpha_k})) \otimes \mathrm{E}_x((\mathbb{C}1 \oplus \mathring{\mathcal{A}}_{\alpha_k})\mathring{\mathcal{A}}_{\alpha_{k+1}}\cdots\mathring{\mathcal{A}}_{\alpha_n}).
\end{aligned}$$

By Exercise 17, in each term, one or both of the factors is 0. Thus $\mathrm{E}_x \otimes \mathrm{E}_x \circ \partial_x|_{\mathcal{A}_1\vee\mathcal{A}_2\ominus\mathcal{A}_1} = 0$. Hence

$$\mathrm{E}_x \otimes \mathrm{E}_x \circ \partial_x|_{\mathcal{A}_1\vee\mathcal{A}_2} = \mathrm{E}_x \otimes \mathrm{E}_x \circ \partial_x \circ \mathrm{E}_x|_{\mathcal{A}_1\vee\mathcal{A}_2} = \partial_x \circ \mathrm{E}_x|_{\mathcal{A}_1\vee\mathcal{A}_2},$$

and then by Equation (2.35) we have

$$\mathrm{E}_x \otimes \mathrm{E}_x \circ \partial_{x+y}|_{\mathbb{C}\langle x+y\rangle} = \mathrm{E}_x \otimes \mathrm{E}_x \circ \partial_x|_{\mathbb{C}\langle x+y\rangle} = \partial_x \circ \mathrm{E}_x|_{\mathbb{C}\langle x+y\rangle}.$$

$\square$

Theorem 21. *Let x and y be free. For every $z \in \mathbb{C}$ with $z \neq 0$, there exists a $w \in \mathbb{C}$ such that*

$$\mathrm{E}_x\left[\frac{1}{z-(x+y)}\right] = \frac{1}{w-x}. \tag{2.39}$$

In other words, the best approximation for a resolvent in $x+y$ by a function of x is again a resolvent.

By applying the state φ to both sides of (2.39), one obtains the subordination for the Cauchy transforms, and thus it is clear that the w from above must agree with the subordination function from (2.31), $w = \omega(z)$.

Proof: We put

$$f(x, y) := \frac{1}{z-(x+y)}.$$

By Exercise 16, part (i), we know that $\partial_{x+y} f = f \otimes f$. By Lemma 20 we have that for functions g of $x+y$

$$\partial_x \mathrm{E}_x[g(x+y)] = \mathrm{E}_x \otimes \mathrm{E}_x[\partial_{x+y} g(x+y)]. \tag{2.40}$$

By applying (2.40) to f, we obtain

$$\partial_x \mathrm{E}_x[f] = \mathrm{E}_x \otimes \mathrm{E}_x[\partial_{x+y} f] = \mathrm{E}_x \otimes \mathrm{E}_x[f \otimes f] = \mathrm{E}_x[f] \otimes \mathrm{E}_x[f].$$

Thus, by the second part of Exercise 16, we know that $\mathrm{E}_x[f]$ is a resolvent in x and we are done. $\square$

Chapter 3
Free Harmonic Analysis

In this chapter we shall present an approach to free probability based on analytic functions. At the end of the previous chapter, we defined the Cauchy transform of a random variable a in an algebra $\mathcal{A}$ with a state φ to be the formal power series $G(z) = \frac{1}{z}M(\frac{1}{z})$ where $M(z) = 1 + \sum_{n\geq 1} \alpha_n z^n$ and $\alpha_n = \varphi(a^n)$ are the moments of a. Then $R(z)$, the R-transform of a, was defined to be the formal power series $R(z) = \sum_{n\geq 1} \kappa_n z^{n-1}$ determined by the moment-cumulant relation which we have shown to be equivalent to the equations

$$G\big(R(z) + 1/z\big) = z = 1/G(z) + R(G(z)). \tag{3.1}$$

If a is a self-adjoint element of a unital C^*-algebra $\mathcal{A}$ with a state φ, then there is a spectral measure ν on $\mathbb{R}$ such that the moments of a are the same as the moments of the probability measure ν. We can then define the analytic function

$$G(z) = \varphi((z-a)^{-1}) = \int_{\mathbb{R}} (z-t)^{-1} d\nu(t)$$

on the complex upper half plane, $\mathbb{C}^+$. One can then consider the relation between the formal power series G obtained from the moment-generating function and the analytic function G obtained from the spectral measure. It turns out that on the exterior of a disc containing the support of ν, the formal power series converges to the analytic function, and the R-transform becomes an analytic function on an open set containing 0 whose power series expansion is the formal power series $\sum_{n\geq 1} \kappa_n z^{n-1}$ given in the previous chapter.

When ν does not have all moments, there is no formal power series; this corresponds to a being an unbounded self-adjoint operator affiliated with $\mathcal{A}$. However, the Cauchy transform is always defined. Moreover, one can construct the R-transform of ν, analytic on some open set, satisfying equation (3.1) – although there may not be any free cumulants if ν has no moments. However, if ν does have moments, then the R-transform has cumulants given by an asymptotic expansion at 0.

J.A. Mingo, R. Speicher, *Free Probability and Random Matrices*,
Fields Institute Monographs 35, DOI 10.1007/978-1-4939-6942-5_3

If X and Y are classically independent random variables with distributions ν_X and ν_Y, then the distribution of $X+Y$ is the convolution, $\nu_X * \nu_Y$. We shall construct the free analogue, $\nu_X \boxplus \nu_Y$, of the classical convolution. $\nu_X \boxplus \nu_Y$ is called the *free additive convolution* of ν_X and ν_Y; it is the distribution of the sum $X+Y$ when X and Y are freely independent. Since X and Y do not commute, we cannot do this with functions as in the classical case. We shall do this on the level of probability measures.

We shall ultimately show that the R-transform exists for all probability measures. However, we shall first do this for compactly supported probability measures, then for probability measures with finite variance, and finally for arbitrary probability measures. This follows more or less the historical development. The compactly supported case was treated in [177] by Voiculescu. The case of finite variance was then treated by Maassen in [120]; this was an important intermediate step, as it promoted the use of the reciprocal Cauchy transform $F = 1/G$ and of the subordination function. The general case was then first treated by Bercovici and Voiculescu in [30] by operator algebraic methods; however, more recent alternative approaches, by Belinschi and Bercovici [18, 21] and by Chistyakov and Götze [53, 54], rely on the subordination formulation. Since this subordination approach seems to be analytically better controllable than the R-transform and also best suited for generalizations to the operator-valued case (see Chapter 10, in particular Section 10.4), we will concentrate in our presentation on this approach and try to give a streamlined and self-contained presentation.

3.1 The Cauchy transform

Definition 1. Let $\mathbb{C}^+ = \{z \in \mathbb{C} \mid \mathrm{Im}(z) > 0\}$ denote the complex upper half-plane and $\mathbb{C}^- = \{z \mid \mathrm{Im}(z) < 0\}$ denote the lower half-plane. Let ν be a probability measure on $\mathbb{R}$ and for $z \notin \mathbb{R}$ let

$$G(z) = \int_{\mathbb{R}} \frac{1}{z-t}\, d\nu(t);$$

G is the *Cauchy transform* of the measure ν.

Let us briefly check that the integral converges to an analytic function on $\mathbb{C}^+$.

Lemma 2. *G is an analytic function on $\mathbb{C}^+$ with range contained in $\mathbb{C}^-$.*

Proof: Since $|z-t|^{-1} \leq |\mathrm{Im}(z)|^{-1}$ and ν is a probability measure, the integral is always convergent. If $\mathrm{Im}(w) \neq 0$ and $|z-w| < |\mathrm{Im}(w)|/2$, then for $t \in \mathbb{R}$, we have

$$\left|\frac{z-w}{t-w}\right| < \frac{|\mathrm{Im}(w)|}{2} \cdot \frac{1}{|\mathrm{Im}(w)|} = \frac{1}{2},$$

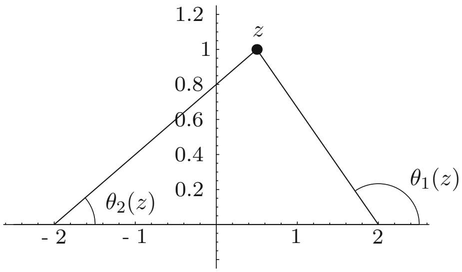


Fig. 3.1 We choose θ_1, the argument of $z-2$, to be such that $0 \le \theta_1 < 2\pi$. Similarly we choose θ_2, the argument of $z+2$, such that $0 \le \theta_2 < 2\pi$. Thus $\theta_1 + \theta_2$ is continuous on $\mathbb{C} \setminus [-2, \infty)$. However $e^{i(\theta_1+\theta_2)/2}$ is continuous on $\mathbb{C} \setminus [-2, 2]$ because $e^{i(0+0)/2} = 1 = e^{i(2\pi+2\pi)/2}$, so there is no jump as the half lines $(-\infty, -2]$ and $[2, \infty)$ are crossed

so the series $\sum_{n=0}^{\infty}((z-w)/(t-w))^n$ converges uniformly to $(t-w)/(t-z)$ on $|z-w| < |\mathrm{Im}(w)|/2$. Thus $(z-t)^{-1} = -\sum_{n=0}^{\infty}(t-w)^{-(n+1)}(z-w)^n$ on $|z-w| < |\mathrm{Im}(w)|/2$. Hence

$$G(z) = -\sum_{n=0}^{\infty}\left[\int_{\mathbb{R}}(t-w)^{-(n+1)}d\nu(t)\right](z-w)^n$$

is analytic on $|z-w| < |\mathrm{Im}(w)|/2$.

Finally note that for $\mathrm{Im}(z) > 0$, we have for $t \in \mathbb{R}$, $\mathrm{Im}((z-t)^{-1}) < 0$, and hence $\mathrm{Im}(G(z)) < 0$. Thus G maps $\mathbb{C}^+$ into $\mathbb{C}^-$. □

Exercise 1. (*i*) Let μ be the atomic probability measure with atoms at the real numbers $\{a_1, \ldots, a_n\}$ and let $\lambda_i = \mu(\{a_i\})$ be the mass of the atom at a_i. Find the Cauchy transform of μ.

(*ii*) Let ν be the Cauchy distribution, i.e. $d\nu(t) = \pi^{-1}(1+t^2)^{-1}\,dt$. Show that $G(z) = 1/(z+i)$.

In the next two exercises, we need to choose a branch of $\sqrt{z^2-4}$ for z in the upper half-plane, $\mathbb{C}^+$. We write $z^2-4 = (z-2)(z+2)$ and define each of $\sqrt{z-2}$ and $\sqrt{z+2}$ on $\mathbb{C}^+$. For $z \in \mathbb{C}^+$, let θ_1 be the angle between the x-axis and the line joining z to 2; and θ_2 the angle between the x-axis and the line joining z to -2. See Fig. 3.1. Then $z-2 = |z-2|e^{i\theta_1}$ and $z+2 = |z+2|e^{i\theta_2}$ and so we define $\sqrt{z^2-4}$ to be $|z^2-4|^{1/2}e^{i(\theta_1+\theta_2)/2}$.

Exercise 2. For $z = u + iv \in \mathbb{C}^+$ let $\sqrt{z} = \sqrt{|z|}e^{i\theta/2}$ where $0 < \theta < \pi$ is the argument of z. Show that

$$\mathrm{Re}(\sqrt{z}) = \sqrt{\frac{\sqrt{u^2+v^2}+u}{2}} \quad \text{and} \quad \mathrm{Im}(\sqrt{z}) = \sqrt{\frac{\sqrt{u^2+v^2}-u}{2}}.$$

Exercise 3. For $z \in \mathbb{C}^+$ show that

$$|\mathrm{Im}(z)| < \left|\mathrm{Im}\left(\sqrt{z^2-4}\right)\right| \quad \text{and} \quad \left|\mathrm{Re}\left(\sqrt{z^2-4}\right)\right| \leq |\mathrm{Re}(z)|;$$

with equality in the second relation only when $\mathrm{Re}(z) = 0$.

Exercise 4. In this exercise we shall compute the Cauchy transform of the *arc-sine* law using contour integration. Recall that the density of the arc-sine law on the interval $[-2, 2]$ is given by $d\nu(t) = 1/(\pi\sqrt{4-t^2})$. Let

$$G(z) = \frac{1}{\pi}\int_{-2}^{2} \frac{(z-t)^{-1}}{\sqrt{4-t^2}}\, dt.$$

(i) Make the substitution $t = 2\cos\theta$ for $0 \leq \theta \leq \pi$. Show that

$$G(z) = \frac{1}{2\pi}\int_0^{2\pi} (z - 2\cos\theta)^{-1}\, d\theta.$$

(ii) Make the substitution $w = e^{i\theta}$ and show that we can write G as the contour integral

$$G(z) = \frac{1}{2\pi i}\int_{\Gamma} \frac{1}{zw - w^2 - 1}\, dw$$

where $\Gamma = \{w \in \mathbb{C} \mid |w| = 1\}$.

(iii) Show that the roots of $zw - w^2 - 1 = 0$ are $w_1 = (z - \sqrt{z^2-4})/2$ and $w_2 = (z + \sqrt{z^2-4})/2$ and that $w_1 \in \mathrm{int}(\Gamma)$ and that $w_2 \notin \mathrm{int}(\Gamma)$, using the branch defined above.

(iv) Using the residue calculus, show that $G(z) = 1/\sqrt{z^2-4}$.

Exercise 5. In this exercise we shall compute the Cauchy transform of the semi-circle law using contour integration. Recall that the density of the semi-circle law on the interval $[-2, 2]$ is given by $d\nu(t) = (2\pi)^{-1}\sqrt{4-t^2}$. Let

$$G(z) = \frac{1}{2\pi}\int_{-2}^{2} \frac{\sqrt{4-t^2}}{z-t}\, dt.$$

(i) Make the substitution $t = 2\cos\theta$ for $0 \leq \theta \leq \pi$. Show that

$$G(z) = \frac{1}{4\pi}\int_0^{2\pi} \frac{4\sin^2\theta}{z - 2\cos\theta}\, d\theta.$$

(*ii*) Make the substitution $w = e^{i\theta}$ and show that we can write G as the contour integral

$$G(z) = \frac{1}{4\pi i}\int_\Gamma \frac{(w^2-1)^2}{w^2(w^2-zw+1)}\,dw$$

where $\Gamma = \{w \in \mathbb{C} \mid |w| = 1\}$.

(*iii*) Using the results from Exercise 3 and the residue calculus, show that

$$G(z) = \frac{z-\sqrt{z^2-4}}{2}, \tag{3.2}$$

using the branch defined above.

Exercise 6. In this exercise we shall compute the Cauchy transform of the Marchenko-Pastur law with parameter c using contour integration. We shall start by supposing that $c > 1$. Recall that the density of the Marchenko-Pastur law on the interval $[a,b]$ is given by $d\nu_c(t) = \sqrt{(b-t)(t-a)}/(2\pi t)dt$ with $a = (1-\sqrt{c})^2$ and $b = (1+\sqrt{c})^2$. Let

$$G(z) = \int_a^b \frac{\sqrt{(b-t)(t-a)}}{2\pi t(z-t)}dt.$$

(*i*) Make the substitution $t = 1 + 2\sqrt{c}\cos\theta + c$ for $0 \le \theta \le \pi$. Show that

$$G(z) = \frac{1}{4\pi}\int_0^{2\pi} \frac{4c\sin^2\theta}{(1+2\sqrt{c}\cos\theta + c)(z-1-2\sqrt{c}\cos\theta - c)}\,d\theta.$$

(*ii*) Make the substitution $w = e^{i\theta}$ and show that we can write G as the contour integral

$$G(z) = \frac{1}{4\pi i}\int_\Gamma \frac{(w^2-1)^2}{w(w^2+fw+1)(w^2-ew+1)}\,dw$$

where $\Gamma = \{w \in \mathbb{C} \mid |w| = 1\}$, $f = (1+c)/\sqrt{c}$, and $e = (z-(1+c))/\sqrt{c}$.

(*iii*) Using the results from Exercise 3 and the residue calculus, show that

$$G(z) = \frac{z+1-c-\sqrt{(z-a)(z-b)}}{2z}, \tag{3.3}$$

using the branch defined in the same way as with $\sqrt{z^2-4}$ above except a replaces -2 and b replaces 2.

Lemma 3. *Let G be the Cauchy transform of a probability measure ν. Then:*

$$\lim_{y\to\infty} iy\, G(iy) = 1 \quad \textit{and} \quad \sup_{y>0, x\in\mathbb{R}} y\, |G(x+iy)| = 1.$$

Proof: We have

$$y\,\mathrm{Im}(G(iy)) = \int_{\mathbb{R}} y\,\mathrm{Im}\Big(\frac{1}{iy-t}\Big)\, d\nu(t) = \int_{\mathbb{R}} \frac{-y^2}{y^2+t^2}\, d\nu(t)$$
$$= -\int_{\mathbb{R}} \frac{1}{1+(t/y)^2}\, d\nu(t) \to -\int_{\mathbb{R}} d\nu(t) = -1$$

as $y \to \infty$; since $(1+(t/y)^2)^{-1} \le 1$, we could apply Lebesgue's dominated convergence theorem.

We have

$$y\,\mathrm{Re}(G(iy)) = \int_{\mathbb{R}} \frac{-yt}{y^2+t^2}\, d\nu(t).$$

But for all $y > 0$ and for all t

$$\left|\frac{yt}{y^2+t^2}\right| \le \frac{1}{2},$$

and $|yt/(y^2+t^2)|$ converges to 0 as $y \to \infty$. Therefore $y\,\mathrm{Re}(G(iy)) \to 0$ as $y \to \infty$, again by the dominated convergence theorem. This gives the first equation of the lemma.

For $y > 0$ and $z = x + iy$,

$$y\,|G(z)| \le \int_{\mathbb{R}} \frac{y}{|z-t|}\, d\nu(t) = \int_{\mathbb{R}} \frac{y}{\sqrt{(x-t)^2+y^2}}\, d\nu(t) \le 1.$$

Thus $\sup_{y>0, x\in\mathbb{R}} y\, |G(x+iy)| \le 1$. By the first part, however, the supremum is 1. □

Another frequently used notation is to let $m(z) = \int (t-z)^{-1}\, d\nu(t)$. We have $m(z) = -G(z)$ and m is usually called the *Stieltjes transform* of ν. It maps $\mathbb{C}^+$ to $\mathbb{C}^+$.

Notation 4. *Let us recall the* Poisson kernel *from harmonic analysis. Let*

$$P(t) = \frac{1}{\pi}\frac{1}{1+t^2} \quad \textit{and} \quad P_\epsilon(t) = \epsilon^{-1}P(t\epsilon^{-1}) = \frac{1}{\pi}\frac{\epsilon}{t^2+\epsilon^2} \quad \textit{for } \epsilon > 0.$$

If ν_1 and ν_2 are two probability measures on $\mathbb{R}$, recall that their convolution is defined by $\nu_1 * \nu_2(E) = \int_{-\infty}^{\infty} \nu_1(E-t)\, d\nu_2(t)$ (see Rudin [151, Ex. 8.5]). If ν is a probability measure on $\mathbb{R}$ and $f \in L^1(\mathbb{R}, \nu)$, we can define $f * \nu$ by

$$f * \nu(t) = \int_{-\infty}^{\infty} f(t-s)\, d\nu(s).$$

Since P is bounded, we can form $P_\epsilon * \nu$ for any probability measure ν and any $\epsilon > 0$. Moreover P_ϵ is the density of a probability measure, namely a Cauchy distribution *with scale parameter ϵ. We shall denote this distribution by $\delta_{-i\epsilon}$.*

Remark 5. Note that $\delta_{-i\epsilon} * \nu$ is a probability measure with density

$$P_\epsilon * \nu(x) = -\frac{1}{\pi}\mathrm{Im}(G(x + i\epsilon)),$$

where G is the Cauchy transform of ν. It is a standard fact that $\delta_{-i\epsilon} * \nu$ converges weakly to ν as $\epsilon \to 0^+$. (Weak convergence is defined in Remark 12). Thus we can use the Cauchy transform to recover ν. In the next theorem, we write this in terms of the distribution functions of measures. In this form it is called the *Stieltjes inversion formula*.

Theorem 6. *Suppose ν is a probability measure on $\mathbb{R}$ and G is its Cauchy transform. For $a < b$ we have*

$$-\lim_{y\to 0^+}\frac{1}{\pi}\int_a^b \mathrm{Im}(G(x+iy))\,dx = \nu((a,b)) + \frac{1}{2}\nu(\{a,b\}).$$

If ν_1 and ν_2 are probability measures with $G_{\nu_1} = G_{\nu_2}$, then $\nu_1 = \nu_2$.

Proof: We have

$$\mathrm{Im}(G(x+iy)) = \int_{\mathbb{R}} \mathrm{Im}\Big(\frac{1}{x-t+iy}\Big)\,d\nu(t) = \int_{\mathbb{R}} \frac{-y}{(x-t)^2+y^2}d\nu(t).$$

Thus

$$\begin{aligned}\int_a^b \mathrm{Im}(G(x+iy))\,dx &= \int_{\mathbb{R}}\int_a^b \frac{-y}{(x-t)^2+y^2}\,dx\,d\nu(t)\\ &= -\int_{\mathbb{R}}\int_{(a-t)/y}^{(b-t)/y}\frac{1}{1+\tilde{x}^2}\,d\tilde{x}\,d\nu(t)\\ &= -\int_{\mathbb{R}}\Big[\tan^{-1}\Big(\frac{b-t}{y}\Big) - \tan^{-1}\Big(\frac{a-t}{y}\Big)\Big]\,d\nu(t),\end{aligned}$$

where we have let $\tilde{x} = (x-t)/y$.

So let $f(y,t) = \tan^{-1}((b-t)/y) - \tan^{-1}((a-t)/y)$ and

$$f(t) = \begin{cases} 0, & t \notin [a,b] \\ \pi/2, & t \in \{a,b\} \\ \pi, & t \in (a,b)\end{cases}.$$

Then $\lim_{y\to 0^+} f(y,t) = f(t)$, and, for all $y > 0$ and for all t, we have $|f(y,t)| \le \pi$. So by Lebesgue's dominated convergence theorem

$$\begin{aligned}\lim_{y\to 0^+}\int_a^b \mathrm{Im}(G(x+iy))\,dx &= -\lim_{y\to 0^+}\int_{\mathbb{R}} f(y,t)\,d\nu(t)\\ &= -\int_{\mathbb{R}} f(t)\,d\nu(t)\\ &= -\pi\Big(\nu((a,b)) + \frac{1}{2}\nu(\{a,b\})\Big).\end{aligned}$$

This proves the first claim.

Now assume that $G_{\nu_1} = G_{\nu_2}$. This implies, by the formula just proved, that $\nu_1((a,b)) = \nu_2((a,b))$ for all a and b which are atoms neither of ν_1 nor of ν_2. Since there are only countably many atoms of ν_1 and ν_2, we can write any interval (a,b) in the form $(a,b) = \cup_{n=1}^{\infty}(a+\epsilon_n, b-\epsilon_n)$ for a decreasing sequence $\epsilon \to 0^+$, such that all $a+\epsilon_n$ and all $b-\epsilon_n$ are atoms neither of ν_1 nor of ν_2. But then we get

$$\nu_1((a,b)) = \lim_{\epsilon_n\to 0^+}\nu_1((a+\epsilon_n, b-\epsilon_n)) = \lim_{\epsilon_n\to 0^+}\nu_2((a+\epsilon_n, b-\epsilon_n)) = \nu_2((a,b)).$$

This shows that ν_1 and ν_2 agree on all open intervals and thus are equal. □

Example 7 (The semi-circle distribution). As an example of Stieltjes inversion, let us take a familiar example and calculate its Cauchy transform using a generating function and then using only the Cauchy transform find the density by using Stieltjes inversion. The density of the semi-circle law $\nu := \mu_s$ is given by

$$d\nu(t) = \frac{\sqrt{4-t^2}}{2\pi}\,dt \qquad \text{on } [-2,2];$$

and the moments are given by

$$m_n = \int_{-2}^{2} t^n\,d\nu(t) = \begin{cases} 0, & n \text{ odd}\\ C_{n/2}, & n \text{ even}\end{cases},$$

where the C_n's are the Catalan numbers:

$$C_n = \frac{1}{n+1}\binom{2n}{n}.$$

Now let $M(z)$ be the moment-generating function

$$M(z) = 1 + C_1 z^2 + C_2 z^4 + \cdots$$

then

$$M(z)^2 = \sum_{m,n\geq 0} C_m C_n z^{2(m+n)} = \sum_{k\geq 0}\left(\sum_{m+n=k} C_m C_n\right) z^{2k}.$$

Now we saw in equation (2.5) that $\sum_{m+n=k} C_m C_n = C_{k+1}$, so

$$M(z)^2 = \sum_{k\geq 0} C_{k+1} z^{2k} = \frac{1}{z^2}\sum_{k\geq 0} C_{k+1} z^{2(k+1)}$$

and therefore

$$z^2 M(z)^2 = M(z) - 1 \qquad \text{or} \qquad M(z) = 1 + z^2 M(z)^2.$$

By replacing $M(z)$ by $z^{-1}G(1/z)$, we get that G satisfies the quadratic equation $zG(z) = 1 + G(z)^2$. Solving this we find that

$$G(z) = \frac{z \pm \sqrt{z^2-4}}{2}.$$

We use the branch of $\sqrt{z^2-4}$ defined before Exercise 2; however, we must choose the sign in front of the square root. By Lemma 3, we require that $\lim_{y\to\infty} iyG(iy) = 1$. Note that for $y > 0$, we have that, using our definition, $\sqrt{(iy)^2-4} = i\sqrt{y^2+4}$. Thus

$$\lim_{y\to\infty}(iy)\frac{iy - \sqrt{(iy)^2-4}}{2} = 1$$

and

$$\lim_{y\to\infty}(iy)\frac{iy + \sqrt{(iy)^2-4}}{2} = \infty.$$

Hence

$$G(z) = \frac{z - \sqrt{z^2-4}}{2}.$$

Of course, this agrees with the result in Exercise 5.

Returning to the equation $zG(z) = 1 + G(z)^2$, we see that $z = G(z) + 1/G(z)$, so $K(z) = z + 1/z$ and thus $R(z) = z$, i.e. all cumulants of the semi-circle law are 0 except κ_2, which equals 1, something we observed already in Exercise 2.9.

Now let us apply Stieltjes inversion to $G(z)$. We have

$$\operatorname{Im}\left(\sqrt{(x+iy)^2-4}\right) = \left|(x+iy)^2-4\right|^{1/2}\sin((\theta_1+\theta_2)/2)$$

$$\lim_{y\to 0^+}\operatorname{Im}\left(\sqrt{(x+iy)^2-4}\right) = \begin{cases} |x^2-4|^{1/2}\cdot 0 = 0, & |x|>2 \\ |x^2-4|^{1/2}\cdot 1 = \sqrt{4-x^2}, & |x|\le 2 \end{cases}$$

and thus

$$\lim_{y\to 0^+}\operatorname{Im}(G(x+iy)) = \lim_{y\to 0^+}\operatorname{Im}\left(\frac{x+iy-\sqrt{(x+iy)^2-4}}{2}\right)$$

$$= \begin{cases} 0, & |x|>2 \\ \dfrac{-\sqrt{4-x^2}}{2}, & |x|\le 2 \end{cases}.$$

Therefore

$$-\lim_{y\to 0^+}\frac{1}{\pi}\operatorname{Im}(G(x+iy)) = \begin{cases} 0, & |x|>2 \\ \dfrac{\sqrt{4-x^2}}{2\pi}, & |x|\le 2 \end{cases}.$$

Hence we recover our original density.

If G is the Cauchy transform of a probability measure, we cannot in general expect $G(z)$ to converge as z converges to $a \in \mathbb{R}$. It might be that $|G(z)| \to \infty$ as $z \to a$ or that G behaves as if it has an essential singularity at a. However $(z-a)G(z)$ always has a limit as $z \to a$ if we take a *non-tangential* limit. Let us recall the definition. Suppose $f : \mathbb{C}^+ \to \mathbb{C}$ and $a \in \mathbb{R}$, we say $\lim_{\sphericalangle z\to a} f(z) = b$ if for every $\theta > 0$, $\lim_{z\to a} f(z) = b$ when we restrict z to be in the cone

$$\{x+iy \mid y>0 \text{ and } |x-a| < \theta y\} \subset \mathbb{C}^+.$$

Proposition 8. *Suppose ν is a probability measure on $\mathbb{R}$ with Cauchy transform G. For all $a \in \mathbb{R}$, we have $\lim\limits_{\sphericalangle z\to a}(z-a)G(z) = \nu(\{a\})$.*

Proof: Let $\theta > 0$ be given. If $z = x+iy$ and $|x-a| < \theta y$, then for $t \in \mathbb{R}$, we have

$$\left|\frac{z-a}{z-t}\right|^2 = \frac{(x-a)^2+y^2}{(x-t)^2+y^2} = \frac{1+(\frac{x-a}{y})^2}{1+(\frac{x-t}{y})^2} \le 1+\left(\frac{x-a}{y}\right)^2 < 1+\theta^2.$$

Let $m = \nu(\{a\})$, δ_a the Dirac mass at a, and $\sigma = \nu - m\delta_a$. Then σ is a subprobability measure and so

$$|(z-a)G(z)-m| = \left| \int \frac{z-a}{z-t}\, d\sigma(t) \right| \le \int \left| \frac{z-a}{z-t} \right| d\sigma(t).$$

We have $|(z-a)/(z-t)| \to 0$ as $z \to a$ for all $t \neq a$. Since $\{a\}$ is a set of σ measure 0, we may apply the dominated convergence theorem to conclude that indeed $\lim_{\sphericalangle z \to a}(z-a)G(z) = m$. □

Let $f(z) = (z-a)G(z)$. Suppose f has an analytic extension to a neighbourhood of a then G has a meromorphic extension to a neighbourhood of a. If $m = \lim_{\sphericalangle z \to a} f(z) > 0$, then G has a simple pole at a with residue m, and ν has an atom of mass m at a. If $m = 0$, then G has an analytic extension to a neighbourhood of a.

Let us illustrate this with the example of the Marchenko-Pastur distribution with parameter c (see the discussion following Exercise 2.9). In that case we have $G(z) = (z+1-c-\sqrt{(z-a)(z-b)})/(2z)$; recall that $a = (1-\sqrt{c})^2$ and $b = (1+\sqrt{c})^2$. If we write this as $f(z)/z$ with $f(z) = (z+1-c-\sqrt{(z-a)(z-b)}\,)/2$, then we may (using the convention of Exercise 6 (*ii*)) extend f to be analytic on $\{z \mid \mathrm{Re}(z) < a\}$ by choosing $\pi/2 < \theta_1, \theta_2 < 3\pi/2$. With this convention we have $f(0) = 1-c$ when $c < 1$ and $f(0) = 0$ when $c > 1$. Note that this is exactly the weight of the atom at 0.

For many probability measures arising in free probability, G has a meromorphic extension to a neighbourhood of a given point a. This is due to two results. The first is a theorem of Greenstein [80, Thm. 1.2] which states that G can be continued analytically to an open set containing the interval (a,b) if and only if the restriction of ν to (a,b) is absolutely continuous with respect to Lebesgue measure and that the density is real analytic. The second is a theorem of Belinschi [19, Thm. 4.1] which states that the free additive convolution (see § 3.5) of two probability measures (provided neither is a Dirac mass) has no continuous singular part and the density is real analytic whenever positive and finite. This means that for such measures G has a meromorphic extension to a neighbourhood of every point where the density is positive on some open set containing the point.

Remark 9. The proof of the next theorem depends on a fundamental result of R. Nevanlinna which provides an integral representation for an analytic function from $\mathbb{C}^+$ to $\mathbb{C}^+$. It is the upper half-plane version of a better known theorem about the harmonic extension of a measure on the boundary of the open unit disc to its interior. Suppose that $\varphi : \mathbb{C}^+ \to \mathbb{C}^+$ is analytic, then the theorem of Nevanlinna asserts that there is a unique finite positive Borel measure σ on $\mathbb{R}$ and real numbers α and β, with $\beta \ge 0$ such that for $z \in \mathbb{C}^+$

$$\varphi(z) = \alpha + \beta z + \int_{\mathbb{R}} \frac{1+tz}{t-z}\, d\sigma(t).$$

This integral representation is achieved by mapping the upper half-plane to the open unit disc $\mathbb{D}$, via $\xi = (iz+1)/(iz-1)$, and then defining ψ on $\mathbb{D}$ by $\psi(\xi) = -i\varphi(z) = -i\varphi\big(i(1+\xi)/(1-\xi)\big)$ and obtaining an analytic function ψ mapping the open unit disc, $\mathbb{D}$, into the complex right half-plane. In the disc version of the problem, we must find a real number β' and a positive measure σ' on $\partial\mathbb{D} = [0, 2\pi]$ such that

$$\psi(z) = i\beta' + \int_0^{2\pi} \frac{e^{it} + z}{e^{it} - z}\, d\sigma'(t).$$

The measure σ' is then obtained as a limit using the Helly selection principle (see, e.g. Lukacs [119, Thm. 3.5.1]). This representation is usually attributed to Herglotz. The details can be found in Akhiezer and Glazman [4, Ch. VI, §59], Rudin [151, Thm. 11.9], or Hoffman [99, p. 34].

The next theorem answers the question as to which analytic functions from $\mathbb{C}^+$ to $\mathbb{C}^-$ are the Cauchy transform of a positive Borel measure.

Theorem 10. *Suppose $G : \mathbb{C}^+ \to \mathbb{C}^-$ is analytic and $\limsup_{y\to\infty} y|G(iy)| = c < \infty$. Then there is a unique positive Borel measure ν on $\mathbb{R}$ such that*

$$G(z) = \int_{\mathbb{R}} \frac{1}{z-t} d\nu(t) \qquad \textit{and} \qquad \nu(\mathbb{R}) = c.$$

Proof: By the remark above, applied to $-G$, there is a unique finite positive measure σ on $\mathbb{R}$ such that $G(z) = \alpha + \beta z + \int (1+tz)/(z-t)\, d\sigma(t)$ with $\alpha \in \mathbb{R}$ and $\beta \le 0$.

Considering first the real part of $iyG(iy)$, we get that for all $y > 0$ large enough

$$2c \ge \mathrm{Re}(iyG(iy)) = y^2\Big(-\beta + \int \frac{1+t^2}{y^2+t^2}\, d\sigma(t)\Big).$$

Since both $-\beta$ and $\int (1+t^2)/(y^2+t^2)\, d\sigma(t)$ are non-negative, the right-hand term can only stay bounded if $\beta = 0$. Thus for all $y > 0$ sufficiently large

$$\int \frac{1+t^2}{1+(t/y)^2}\, d\sigma(t) \le 2c.$$

Thus by the monotone convergence theorem $\int (1+t^2)\, d\sigma(t) \le 2c$ and so σ has a second moment.

From the imaginary part of $iyG(iy)$, we get that for all $y > 0$ sufficiently large

$$y \left| \alpha + \int_{\mathbb{R}} \frac{t(y^2-1)}{t^2+y^2} d\sigma(t) \right| \le 2c,$$

which implies that

$$\alpha = -\lim_{y\to\infty} \int_{\mathbb{R}} \frac{t(y^2-1)}{t^2+y^2}\, d\sigma(t).$$

Since $|(y^2-1)/(t^2+y^2)| \le 1$ for $y \ge 1$ and since σ has a second (and hence also a first) moment, we can apply the dominated convergence theorem and conclude that

$$\alpha = -\lim_{y\to\infty}\int_{\mathbb{R}} t\frac{1-y^{-2}}{1+(t/y)^2}\,d\sigma(t) = -\int_{\mathbb{R}} t\,d\sigma(t).$$

Hence

$$G(z) = \int_{\mathbb{R}}\left(-t+\frac{1+tz}{z-t}\right)d\sigma(t) = \int_{\mathbb{R}}\frac{1}{z-t}(1+t^2)d\sigma(t) = \int_{\mathbb{R}}\frac{1}{z-t}d\nu(t),$$

where we have put $\nu(E) := \int_E (1+t^2)d\sigma(t)$. This ν is a finite measure since σ has a second moment. So G is the Cauchy transform of the positive Borel measure ν. Since the imaginary part of $iyG(iy)$ tends to 0, by Lemma 3, and the real part is positive, we have

$$c = \limsup_{y\to\infty}|iyG(iy)| = \lim_{y\to\infty}\operatorname{Re}(iyG(iy)) = \int(1+t^2)\,d\sigma(t) = \nu(\mathbb{R}).$$

□

Remark 11. Recall that in Definition 2.11, we defined the Marchenko-Pastur law via the density ν_c on $\mathbb{R}$. We then showed in Exercise 2.11 the free cumulants of ν_c are given by $\kappa_n = c$ for all $n \geq 1$. We can also approach the Marchenko-Pastur distribution from the other direction; namely, start with the free cumulants and derive the density using Theorems 6 and 10.

If we assume that $\kappa_n = c$ for all $n \geq 1$ and $0 < c < \infty$, then $R(z) = c/(1-z)$ and so by the reverse of equation (2.27)

$$\frac{1}{G(z)} + R(G(z)) = z \tag{3.4}$$

we conclude that G satisfies the quadratic equation

$$\frac{1}{G(z)} + \frac{c}{1-G(z)} = z.$$

So using our previous notation: $a = (1-\sqrt{c})^2$ and $b = (1+\sqrt{c})^2$, we have

$$G(z) = \frac{z+1-c\pm\sqrt{(z-a)(z-b)}}{2z}.$$

As in Exercise 6, we choose the branch of the square root defined by $\sqrt{(z-a)(z-b)} = \sqrt{|(z-a)(z-b)|}\,e^{i(\theta_1+\theta_2)/2}$, where $0 < \theta_1, \theta_2 < \pi$ and θ_1 is the argument of $z-b$ and θ_2 is the argument of $z-a$. This gives us an analytic function on $\mathbb{C}^+$. To choose the sign in front of $\sqrt{(z-a)(z-b)}$, we take our lead from Theorem 6.

Exercise 7. Show the following.

(i) When $z = iy$ with $y > 0$ we have

$$\lim_{y\to\infty} z \cdot \frac{z+1-c-\sqrt{(z-a)(z-b)}}{2z} = 1;$$

(ii) and

$$\lim_{y\to\infty} z \cdot \frac{z+1-c+\sqrt{(z-a)(z-b)}}{2z} = \infty.$$

(iii) For $z \in \mathbb{C}^+$ show that

$$\frac{z+1-c-\sqrt{(z-a)(z-b)}}{2z} \notin \mathbb{R}.$$

This forces the sign, so now we let

$$G(z) = \frac{z+1-c-\sqrt{(z-a)(z-b)}}{2z} \qquad \text{for } z \in \mathbb{C}^+.$$

This is our candidate for the Cauchy transform of a probability measure. Since $G(\mathbb{C}^+)$ is an open connected subset of $\mathbb{C} \setminus \mathbb{R}$, we have that $G(\mathbb{C}^+)$ is contained in either $\mathbb{C}^+$ or $\mathbb{C}^-$. Part (i) of Exercise 7 shows that $G(\mathbb{C}^+) \subset \mathbb{C}^-$. So by Theorem 10, there is a probability measure on $\mathbb{R}$ for which G is the Cauchy transform.

Exercise 8. As was done in Example 7, show by Stieltjes inversion that the probability measure of which G is the Cauchy transform is ν_c.

Exercise 9. Let a and b be real numbers with $b \le 0$. Let $G(z) = (z - a - ib)^{-1}$. Show that G is the Cauchy transform of a probability measure, δ_{a+ib}, which has a density and find its density using Stieltjes inversion. Let ν be a probability measure on $\mathbb{R}$ with Cauchy transform G. Show that, $\widetilde{G}$, the Cauchy transform of $\delta_{a+ib} * \nu$ is the function $\widetilde{G}(z) = G(z - (a+ib))$. Here $*$ denotes the classical convolution, c.f. Notation 4.

Note that though G looks like the Cauchy transform of the complex constant random variable $a+ib$, it is shown here that it is actually the Cauchy transform of an (unbounded) real random variable. To be clear, we have defined Cauchy transforms only for real-valued random variables, i.e. probability measures on $\mathbb{R}$.

Remark 12. If $\{\nu_n\}_n$ is a sequence of finite Borel measures on $\mathbb{R}$, we say that $\{\nu_n\}_n$ *converges weakly* to the measure ν if for every $f \in C_b(\mathbb{R})$ (the continuous bounded functions on $\mathbb{R}$) we have $\lim_n \int f(t)\, d\nu_n(t) = \int f(t)\, d\nu(t)$. We say that $\{\nu_n\}_n$ *converges vaguely* to ν if for every $f \in C_0(\mathbb{R})$ (the continuous functions on $\mathbb{R}$ vanishing at infinity) we have $\lim_n \int f(t)\, d\nu_n(t) = \int f(t)\, d\nu(t)$. Weak convergence implies vague convergence but not conversely. However, if all ν_n and ν are probability measures, then the vague convergence of $\{\nu_n\}_n$ to ν does imply that $\{\nu_n\}_n$ converges weakly to ν [55, Thm. 4.4.2]. If $\{\nu_n\}_n$ is a sequence of probability measures converging weakly to ν, then the corresponding sequence of Cauchy transforms, $\{G_n\}_n$, converges pointwise to the Cauchy transform of ν, as for fixed $z \in \mathbb{C}^+$, the function $t \mapsto (z-t)^{-1}$ is a continuous function on $\mathbb{R}$, bounded by $|z-t|^{-1} \leq (\mathrm{Im}(z))^{-1}$. The following theorem gives the converse.

Theorem 13. *Suppose that $\{\nu_n\}_n$ is a sequence of probability measures on $\mathbb{R}$ with G_n the Cauchy transform of ν_n. Suppose $\{G_n\}_n$ converges pointwise to G on $\mathbb{C}^+$. If* $\lim_{y\to\infty} iy\,G(iy) = 1$, *then there is a unique probability measure ν on $\mathbb{R}$ such that $\nu_n \to \nu$ weakly, and $G(z) = \int (z-t)^{-1} d\nu(t)$.*

Proof: $\{G_n\}_n$ is uniformly bounded on compact subsets of $\mathbb{C}^+$ (as we have $|G(z)| \leq |\mathrm{Im}(z)|^{-1}$ for the Cauchy transform of any probability measure), so by Montel's theorem $\{G_n\}_n$ is relatively compact in the topology of uniform convergence on compact subsets of $\mathbb{C}^+$, thus, in particular, $\{G_n\}_n$ has a subsequence which converges uniformly on compact subsets of $\mathbb{C}^+$ to an analytic function, which must be G. Thus G is analytic. Now for $z \in \mathbb{C}^+$, $G(z) \in \overline{\mathbb{C}^-}$. Also for each $n \in \mathbb{N}$, $x \in \mathbb{R}$ and $y > 0$, $y\,|G_n(x+iy)| \leq 1$. Thus $\forall x \in \mathbb{R}$, $\forall y \geq 0$, $y\,|G(x+iy)| \leq 1$. So in particular, G is non-constant. If for some $z \in \mathbb{C}^+$, $\mathrm{Im}(G(z)) = 0$ then by the minimum modulus principle G would be constant. Thus G maps $\mathbb{C}^+$ into $\mathbb{C}^-$. Hence by Theorem 10 there is a unique finite measure ν such that $G(z) = \int_{\mathbb{R}} (z-t)^{-1} d\nu(t)$ and $\nu(\mathbb{R}) \leq 1$. Since, by assumption, $\lim_{y\to\infty} iyG(iy) = 1$ we have by Theorem 10 that $\nu(\mathbb{R}) = 1$ and thus ν is a probability measure.

Now by the Helly selection theorem, there is a subsequence $\{\nu_{n_k}\}_k$ converging vaguely to some measure $\tilde{\nu}$. For fixed z the function $t \mapsto (z-t)^{-1}$ is in $C_0(\mathbb{R})$. Thus for $\mathrm{Im}(z) > 0$, $G_{n_k}(z) = \int_{\mathbb{R}} (z-t)^{-1} d\nu_{n_k}(t) \to \int_{\mathbb{R}} (z-t)^{-1} d\tilde{\nu}(t)$. Therefore $G(z) = \int (z-t)^{-1}\, d\tilde{\nu}(t)$ i.e. $\nu = \tilde{\nu}$. Thus $\{\nu_{n_k}\}_k$ converges vaguely to ν. Since ν is a probability measure, $\{\nu_{n_k}\}_k$ converges weakly to ν. So all weak cluster points of $\{\nu_n\}_n$ are ν and thus the whole sequence $\{\nu_n\}_n$ converges weakly to ν. □

Note that we needed the assumption $\lim_{y\to\infty} yG(iy) = -i$ in order to ensure that the limit measure ν is indeed a probability measure. In general, without this assumption, one might lose mass in the limit, and one has only the following statement.

Corollary 14. *Suppose $\{\nu_n\}_n$ is a sequence of probability measures on $\mathbb{R}$ with Cauchy transforms $\{G_n\}_n$. If $\{G_n\}_n$ converges pointwise on $\mathbb{C}^+$, then there is a finite positive Borel measure ν with $\nu(\mathbb{R}) \leq 1$ such that $\{\nu_n\}$ converges vaguely to ν.*

Exercise 10. Identify ν_n and ν for the sequence of Cauchy transforms which are given by $G_n(z) = 1/(z-n)$.

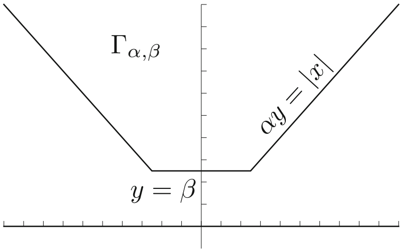


Fig. 3.2 The Stolz angle $\Gamma_{\alpha,\beta}$

3.2 Moments and asymptotic expansions

We saw in Lemma 3 that $zG(z)$ approaches 1 as z approaches ∞ in $\mathbb{C}^+$ along the imaginary axis. Thus $zG(z)-1$ approaches 0 as $z \in \mathbb{C}^+$ tends to ∞. Quantifying how fast $zG(z) - 1$ approaches 0 will be useful in showing that near ∞, G is univalent and thus has an inverse. If our measure has moments, then we get an asymptotic expansion for the Cauchy transform.

Notation 15. *Let $\alpha > 0$ and let $\Gamma_\alpha = \{x + iy \mid \alpha y > |x|\}$ and for $\beta > 0$ let $\Gamma_{\alpha,\beta} = \{z \in \Gamma_\alpha \mid \mathrm{Im}(z) > \beta\}$. See Fig. 3.2. Note that for $z \in \mathbb{C}^+$, we have $z \in \Gamma_\alpha$ if and only if $\sqrt{1+\alpha^2}\,\mathrm{Im}(z) > |z|$.*

Definition 16. If $\alpha > 0$ is given and f is a function on Γ_α, we say $\lim_{z\to\infty, z\in\Gamma_\alpha} f(z) = c$ to mean that for every $\epsilon > 0$ there is $\beta > 0$ so that $|f(z) - c| < \epsilon$ for $z \in \Gamma_{\alpha,\beta}$. If this holds for every α, we write $\lim_{\sphericalangle z\to\infty} f(z) = c$. When it is clear from the context, we shall abbreviate this to $\lim_{z\to\infty} f(z) = c$. We call Γ_α a *Stolz angle* and $\Gamma_{\alpha,\beta}$ a *truncated Stolz angle*. To show convergence in a Stolz angle, it is sufficient to show convergence along a sequence $\{z_n\}_n$ in Γ_α tending to infinity. Hence the usual rules for sums and products of limits apply.

We extend this definition to the case $c = \infty$ as follows. If for every $\alpha_1 < \alpha_2$ and every $\beta_2 > 0$ there is β_1 such that $f(\Gamma_{\alpha_1,\beta_1}) \subset \Gamma_{\alpha_2,\beta_2}$, we say $\lim\limits_{\sphericalangle z\to\infty} f(z) = \infty$.

Exercise 11. Show that if $\lim_{\sphericalangle z\to\infty} f(z)/z = 1$, then $\lim_{\sphericalangle z\to\infty} f(z) = \infty$.

In the following exercises, G will be the Cauchy transform of the probability measure ν.

Exercise 12. Let ν be a probability measure on $\mathbb{R}$ and $\alpha > 0$. In this exercise we will consider limits as $z \to \infty$ with $z \in \Gamma_\alpha$. Show that:

(i) for $z \in \Gamma_\alpha$ and $t \in \mathbb{R}$, $|z - t| \geq |t|/\sqrt{1+\alpha^2}$;
(ii) for $z \in \Gamma_\alpha$ and $t \in \mathbb{R}$, $|z - t| \geq |z|/\sqrt{1+\alpha^2}$;
(iii) $\lim_{z\to\infty} \int_{\mathbb{R}} t/(z-t)\, d\nu(t) = 0$;
(iv) $\lim_{z\to\infty} zG(z) = 1$.

Exercise 13. Let $F : \mathbb{C}^+ \to \mathbb{C}^+$ be analytic and let

$$F(z) = a + bz + \int_{\mathbb{R}} \frac{1 + tz}{t - z}\, d\sigma(t)$$

be its Nevanlinna representation with a real and $b \geq 0$. Then for all $\alpha > 0$, we have $\lim_{z\to\infty} F(z)/z = b$ for $z \in \Gamma_\alpha$.

Exercise 14. Let ν be a probability measure on $\mathbb{R}$. Suppose ν has a moment of order n, i.e. $\int_{\mathbb{R}} |t|^n\, d\nu(t) < \infty$. Let $\alpha_1, \dots, \alpha_n$ be the first n moments of ν. Let $\alpha > 0$ be given. As in Exercise 12, all limits as $z \to \infty$ will be assumed to be in a Stolz angle as in Notation 15.

(*i*) Show that

$$\lim_{z\to\infty} \int_{\mathbb{R}} \left| \frac{t^{n+1}}{z - t} \right| d\nu(t) = 0.$$

(*ii*) Show that

$$\lim_{z\to\infty} z^{n+1} \left(G(z) - \left(\frac{1}{z} + \frac{\alpha_1}{z^2} + \frac{\alpha_2}{z^3} + \cdots + \frac{\alpha_n}{z^{n+1}} \right) \right) = 0.$$

Exercise 15. Suppose that $\alpha > 0$ and ν is a probability measure on $\mathbb{R}$ and that for some $n > 0$ there are real numbers $\alpha_1, \alpha_2, \dots, \alpha_{2n}$ such that as $z \to \infty$ in Γ_α

$$\lim_{z\to\infty} z^{2n+1} \left(G(z) - \left(\frac{1}{z} + \frac{\alpha_1}{z^2} + \cdots + \frac{\alpha_{2n}}{z^{2n+1}} \right) \right) = 0.$$

Show that ν has a moment of order $2n$, i.e. $\int_{\mathbb{R}} t^{2n}\, d\nu(t) < \infty$ and that $\alpha_1, \alpha_2, \dots, \alpha_{2n}$ are the first $2n$ moments of ν.

3.3 Analyticity of the R-transform: compactly supported measures

We now turn to the problem of showing that the R-transform is always an analytic function; recall that in Chapter 2 the R-transform was defined as a formal power series, satisfying Equation (2.26). The more we assume about the probability measure ν, the better behaved is $R(z)$.

Indeed, if ν is a compactly supported measure, supported in the interval $[-r, r]$, then $R(z)$ is analytic on the disc with centre 0 and radius $1/(6r)$. Moreover the coefficients in the power series expansion $R(z) = \kappa_1 + \kappa_2 z + \kappa_3 z^2 + \cdots$ are exactly the free cumulants introduced in Chapter 2.

If ν has variance σ^2, then $R(z)$ is analytic on a disc with centre $-i/(4\sigma)$ and radius $1/(4\sigma)$ (see Theorem 26). Note that 0 is on the boundary of this disc so ν may fail to have any free cumulants beyond the second. However if ν has moments of all orders, then $R(z)$ has an asymptotic expansion at 0, and the coefficients in this expansion are the free cumulants of ν.

The most general situation is when ν is not assumed to have any moments. Then $R(z)$ is analytic on a wedge $\Delta_{\alpha,\beta} = \{z^{-1} \mid z \in \Gamma_{\alpha,\beta}\}$ in the lower half-plane with 0 at its vertex (see Theorem 33).

Consider now first the case that ν is a compactly supported probability measure on $\mathbb{R}$. Then ν has moments of all orders. We will show that the Cauchy transform of ν is univalent on the exterior of a circle centred at the origin. We can then solve the equation $G(R(z)+1/z) = z$ for $R(z)$ to obtain a function R, analytic on the interior of a disc centred at the origin and with power series given by the free cumulants of ν. The precise statements are given in the next theorem.

Theorem 17. *Let ν be a probability measure on $\mathbb{R}$ with support contained in the interval $[-r, r]$ and let G be its Cauchy transform. Then*

(*i*) *G is univalent on $\{z \mid |z| > 4r\}$;*
(*ii*) *$\{z \mid 0 < |z| < 1/(6r)\} \subset \{G(z) \mid |z| > 4r\}$;*
(*iii*) *there is a function R, analytic on $\{z \mid |z| < 1/(6r)\}$ such that $G(R(z)+1/z) = z$ for $0 < |z| < 1/(6r)$;*
(*iv*) *if $\{\kappa_n\}_n$ are the free cumulants of ν, then, for $|z| < 1/(6r)$, $\sum_{n\geq 1} \kappa_n z^{n-1}$ converges to $R(z)$.*

Proof: Let $\{\alpha_n\}_n$ be the moments of ν and let $\alpha_0 = 1$. Note that $|\alpha_n| \leq \int |t|^n \, d\nu(t) \leq r^n$. Let

$$f(z) = G\big(1/z\big) = z \int \frac{1}{1-tz} \, d\nu(t).$$

For $|z| < 1/r$ and $t \in \operatorname{supp}(\nu)$, $|zt| < 1$ and the series $\sum (zt)^n$ converges uniformly on $\operatorname{supp}(\nu)$ and thus $\sum_{n\geq 0} \alpha_n z^{n+1}$ converges uniformly to $f(z)$ on compact subsets of $\{z \mid |z| < 1/r\}$. Hence $\sum_{n\geq 0} \alpha_n z^{-(n+1)}$ converges uniformly to $G(z)$ on compact subsets of $\{z \mid |z| > r\}$.

Suppose $|z_1|, |z_2| < r^{-1}$. Then

$$\begin{aligned} \left| \frac{f(z_1) - f(z_2)}{z_1 - z_2} \right| &\geq \operatorname{Re}\left(\frac{f(z_1) - f(z_2)}{z_1 - z_2} \right) \\ &= \operatorname{Re} \int_0^1 \frac{d}{dt} \left[\frac{f(z_1 + t(z_2 - z_1))}{z_2 - z_1} \right] dt \\ &= \int_0^1 \operatorname{Re}\big(f'(z_1 + t(z_2 - z_1))\big) \, dt. \end{aligned}$$

And

$$\begin{aligned}
\mathrm{Re}(f'(z)) &= \mathrm{Re}(1 + 2z\alpha_1 + 3z^2\alpha_2 + \cdots) \\
&\geq 1 - 2\,|z|\,r - 3\,|z|^2\,r^2 - \cdots \\
&= 2 - (1 + 2(|z|\,r) + 3(|z|\,r)^2 + \cdots) \\
&= 2 - \frac{1}{(1 - |z|\,r)^2}.
\end{aligned}$$

For $|z| < (4r)^{-1}$, we have

$$\mathrm{Re}(f'(z)) \geq 2 - \frac{1}{(1 - 1/4)^2} = \frac{2}{9}.$$

Hence for $|z_1|, |z_2| < (4r)^{-1}$, we have $|f(z_1) - f(z_2)| \geq 2|z_1 - z_2|/9$. In particular, f is univalent on $\{z \mid |z| < (4r)^{-1}\}$. Hence G is univalent on $\{z \mid |z| > 4r\}$. This proves (i).

For any curve Γ in $\mathbb{C}$ and any w not on Γ let $\mathrm{Ind}_\Gamma(w) = \int_\Gamma (z - w)^{-1} dz/(2\pi i)$ be the index of w with respect to Γ (or the winding number of Γ around w). Now, as $f(0) = 0$, the only solution to $f(z) = 0$ for $|z| < (4r)^{-1}$ is $z = 0$. Let Γ be the curve $\{z \mid |z| = (4r)^{-1}\}$ and $f(\Gamma) = \{f(z) \mid z \in \Gamma\}$ be the image of Γ under f. By the argument principle

$$\mathrm{Ind}_{f(\Gamma)}(0) = \frac{1}{2\pi i}\int_\Gamma \frac{f'(z)}{f(z)}\,dz = 1.$$

Also for $|z| < (4r)^{-1}$

$$\begin{aligned}
|f(z)| &= |z|\,|1 + \alpha_1 z + \alpha_2 z^2 + \cdots| \\
&\geq |z|\,(2 - (1 + r|z| + r^2|z|^2 + \cdots)) \\
&= |z|\left(2 - \frac{1}{1 - r\,|z|}\right) \\
&\geq |z|\left(2 - \frac{1}{1 - 1/4}\right) \\
&= \frac{2}{3}|z|.
\end{aligned}$$

Thus for $|z| = (4r)^{-1}$, we have $|f(z)| \geq (6r)^{-1}$. Hence $f(\Gamma)$ lies outside the circle $|z| = (6r)^{-1}$ and thus $\{z \mid |z| < (6r)^{-1}\}$ is contained in the connected component of $\mathbb{C}\backslash f(\Gamma)$ containing 0. So for $w \in \{z \mid |z| < (6r)^{-1}\}$, $\mathrm{Ind}_{f(\Gamma)}(w) = \mathrm{Ind}_{f(\Gamma)}(0) = 1$, as the index is constant on connected components of the complement of $f(\Gamma)$. Hence

$$1 = \mathrm{Ind}_{f(\Gamma)}(w) = \frac{1}{2\pi i}\int_\Gamma \frac{f'(z)}{f(z) - w}\,dz,$$

so again by the argument principle there is exactly one $|z|$ with $z < (4r)^{-1}$ such that $f(z) = w$. Hence

$$\{z \mid |z| < (6r)^{-1}\} \subset \{f(z) \mid |z| < (4r)^{-1}\}$$

and thus

$$\{z \mid 0 < |z| < (6r)^{-1}\} \subset \{G(z) \mid |z| > 4r\}.$$

This proves (*ii*).

Let $f^{\langle -1 \rangle}$ be the inverse of f on $\{z \mid |z| < (6r)^{-1}\}$. Then $f^{\langle -1 \rangle}(0) = 0$ and $f^{\langle -1 \rangle\prime}(0) = 1/f'(0) = 1$, so $f^{\langle -1 \rangle}$ has a simple zero at 0. Let K be the meromorphic function on $\{z \mid |z| < (6r)^{-1}\}$ given by $K(z) = 1/f^{\langle -1 \rangle}(z)$. Then K has a simple pole at 0 with residue 1. Hence $R(z) = K(z) - 1/z$ is holomorphic on $\{z \mid |z| < (6r)^{-1}\}$, and for $0 < |z| < (6r)^{-1}$

$$G(R(z) + 1/z) = G(K(z)) = f\left(\frac{1}{K(z)}\right) = f(f^{\langle -1 \rangle})(z) = z.$$

This proves (*iii*).

Let $C(z) = 1 + zR(z) = zK(z)$. Then C is analytic on $\{z \mid |z| < (6r)^{-1}\}$ and so has a power series expansion $\sum_{n \geq 0} \tilde{\kappa}_n z^n$, with $\tilde{\kappa}_0 = 1$. We shall have proved (*iv*) if we can show that for all $n \geq 1$, $\tilde{\kappa}_n = \kappa_n$ where $\{\kappa_n\}_n$ are the free cumulants of ν.

Recall that $f(0) = 0$, so $M(z) := f(z)/z = \sum_{n \geq 0} \alpha_n z^n$ is analytic on the set $\{z \mid |z| < r^{-1}\}$. For z such that $|z| < (4r)^{-1}$ and $|f(z)| < (6r)^{-1}$ we have

$$C(f(z)) = f(z)K(f(z)) = \frac{f(z)}{f^{\langle -1 \rangle}(f(z))} = \frac{f(z)}{z} = M(z). \tag{3.5}$$

Fix $p \geq 1$. Then we may write

$$M(z) = 1 + \sum_{l=1}^{p} \alpha_l z^l + o(z^p),$$

$$C(z) = 1 + \sum_{l=1}^{p} \tilde{\kappa}_l z^l + o(z^p),$$

and

$$(f(z))^l = \Big(\sum_{m=1}^{p} \alpha_{m-1} z^m\Big)^l + o(z^p).$$

Hence

$$C(f(z)) = 1 + \sum_{l=1}^{p} \tilde{\kappa}_l \Big(\sum_{m=1}^{p} \alpha_{m-1} z^m\Big)^l + o(z^p).$$

Thus by equation (3.5)

$$1 + \sum_{l=1}^{p} \alpha_l z^l = 1 + \sum_{l=1}^{p} \tilde{\kappa}_l \Big(\sum_{m=1}^{p} \alpha_{m-1} z^m \Big)^l + o(z^p).$$

However this is exactly the relation between $\{\alpha_n\}_n$ and $\{\kappa_n\}_n$ found at the end of the proof of Proposition 2.17. Given $\{\alpha_n\}_n$ there are unique κ_n's that satisfy this relation, so we must have $\tilde{\kappa}_n = \kappa_n$ for all n. This proves (iv). □

Remark 18. A converse to Theorem 17 was found by F. Benaych-Georges [27]. Namely, if a probability measure ν has an R-transform which is analytic on an open set containing 0 and for all $k \geq 0$, the k^{th} derivative $R^{(k)}(0)$ is a real number, then ν has compact support. Note that for the Cauchy distribution $R^{(k)}(0) = 0$ for $k \geq 1$ but $R(0)$ is not real.

3.4 Measures with finite variance

In the previous section, we showed that if ν has compact support, then the R-transform of ν is analytic on an open disc containing 0. If we assume that ν has finite variance but make no assumption about the support, then we can still conclude that the equation $G(R(z) + 1/z) = z$ has an analytic solution on an open disc in the lower half-plane. The main problem is again demonstrating the univalence of G, which is accomplished by a winding number argument.

We have already seen in Exercise 12 that $zG(z) - 1 \to 0$ as $z \to \infty$ in some Stolz angle Γ_α. Let $G_1(z) = z - 1/G(z)$. Then $G_1(z)/z \to 0$, i.e. $G_1(z) = o(z)$. If ν has a first moment α_1, then $z^2(G(z) - (1/z + \alpha_1/z^2)) \to 0$, and we may write

$$z^2 \left(G(z) - \Big(\frac{1}{z} + \frac{\alpha_1}{z^2}\Big) \right) = zG(z)(G_1(z) - \alpha_1) + \alpha_1(zG(z) - 1).$$

Thus $G_1(z) \to \alpha_1$. Suppose ν has a second moment α_2 then

$$z^3 \left(G(z) - \Big(\frac{1}{z} + \frac{\alpha_1}{z^2} + \frac{\alpha_2}{z^3}\Big) \right) \to 0$$

or equivalently

$$G(z) = \frac{1}{z} + \frac{\alpha_1}{z^2} + \frac{\alpha_2}{z^3} + o\Big(\frac{1}{z^3}\Big)$$

and thus

$$G_1(z) = z - \frac{1}{\frac{1}{z} + \frac{\alpha_1}{z^2} + \frac{\alpha_2}{z^3} + o(\frac{1}{z^3})} = \alpha_1 + \frac{\alpha_2 - \alpha_1^2}{z} + o\Big(\frac{1}{z}\Big). \tag{3.6}$$

The next lemma shows that G_1 maps $\mathbb{C}^+$ to $\mathbb{C}^-$. We shall work with the function $F = 1/G$. It will be useful to establish some properties of F (Lemmas 19 and 20) and then show that these properties characterize the reciprocals of Cauchy transforms of measures of finite variance (Lemma 21).

Lemma 19. *Let ν be a probability measure on $\mathbb{R}$ and G its Cauchy transform. Let $F(z) = 1/G(z)$. Then F maps $\mathbb{C}^+$ to $\mathbb{C}^+$ and $\mathrm{Im}(z) \leq \mathrm{Im}(F(z))$ for $z \in \mathbb{C}^+$, with equality for some z only if ν is a Dirac mass.*

Proof: $\mathrm{Im}(G(z)) = -\mathrm{Im}(z) \int |z-t|^{-2}\, d\nu(t)$, so

$$\frac{\mathrm{Im}(F(z))}{\mathrm{Im}(z)} = \frac{-\mathrm{Im}(G(z))}{\mathrm{Im}(z)} \frac{1}{|G(z)|^2} = \frac{\int |z-t|^{-2}\, d\nu(t)}{|G(z)|^2}.$$

So our claim reduces to showing that $|G(z)|^2 \leq \int |z-t|^{-2}\, d\nu(t)$. However, by the Cauchy-Schwartz inequality

$$\left| \int \frac{1}{z-t}\, d\nu(t) \right|^2 \leq \int 1^2\, d\nu(t) \int \left| \frac{1}{z-t} \right|^2 d\nu(t) = \int \left| \frac{1}{z-t} \right|^2 d\nu(t),$$

with equality only if $t \mapsto (z-t)^{-1}$ is ν-almost constant, i.e. ν is a Dirac mass. This completes the proof. □

Lemma 20. *Let ν be a probability measure with finite variance σ^2 and let $G_1(z) = z - 1/G(z)$, where G is the Cauchy transform of ν. Then there is a probability measure ν_1 such that*

$$G_1(z) = \alpha_1 + \sigma^2 \int \frac{1}{z-t}\, d\nu_1(t)$$

where α_1 is the mean of ν.

Proof: If $\sigma^2 = 0$, then ν is a Dirac mass, thus $G_1(z) = \alpha_1$ and the assertion is trivially true. So let us assume that $\sigma^2 \neq 0$. $G_1(z) = z - 1/G(z)$ is analytic on $\mathbb{C}^+$ and by the previous lemma $G_1(\mathbb{C}^+) \subset \mathbb{C}^-$. Let α_1 and α_2 be the first and second moment of ν, respectively. Clearly, we also have $(G_1 - \alpha_1)(\mathbb{C}^+) \subset \mathbb{C}^-$ and, by equation (3.6), $\lim_{z\to\infty} z(G_1(z) - \alpha_1) = \alpha_2 - \alpha_1^2 = \sigma^2 > 0$. Thus by Theorem 10 there is a probability measure ν_1 such that

$$G_1(z) - \alpha_1 = \sigma^2 \int \frac{1}{z-t}\, d\nu_1(t).$$

□

Lemma 21. *Suppose that $F : \mathbb{C}^+ \to \mathbb{C}^+$ is analytic and there is $C > 0$ such that for $z \in \mathbb{C}^+$, $|F(z) - z| \leq C/\mathrm{Im}(z)$. Then there is a probability measure ν with mean 0 and variance $\sigma^2 \leq C$ such that $1/F$ is the Cauchy transform of ν. Moreover σ^2 is the smallest C such that $|F(z) - z| \leq C/\mathrm{Im}(z)$.*

Proof: Let $G(z) = 1/F(z)$. Then $G : \mathbb{C}^+ \to \mathbb{C}^-$ is analytic and

$$\left|1 - \frac{1}{zG(z)}\right| = \frac{|F(z) - z|}{|z|} \leq \frac{C}{|z|\mathrm{Im}(z)}.$$

Hence $\lim_{z\to\infty} zG(z) = 1$ in any Stolz angle. Thus by Theorem 10 there is a probability measure ν such that G is the Cauchy transform of ν. Now

$$\int \frac{y^2}{y^2+t^2} t^2 \, d\nu(t) = y^2 \left[- \int \frac{y^2}{y^2+t^2} d\nu(t) + 1 \right] = y\,\mathrm{Im}\big[iy\,G(iy)\,(F(iy) - iy)\big].$$

Also, allowing that both sides might equal ∞, we have by the monotone convergence theorem that

$$\int t^2 \, d\nu(t) = \lim_{y\to\infty} \int \frac{y^2}{y^2+t^2} t^2 \, d\nu(t).$$

However

$$|y\mathrm{Im}\big[iyG(iy)\,(F(iy) - iy)\big]| \leq \frac{y\,|iy\,G(iy)|\,C}{\mathrm{Im}(iy)} = C\,|iy\,G(iy)|,$$

thus $\int t^2 \, d\nu(t) \leq C$, and so ν has a second, and thus also a first, moment. Also

$$\int \frac{y^2}{y^2+t^2} t \, d\nu(t) = -y^2 \mathrm{Re}(G(iy)) = -\mathrm{Re}\big[iyG(iy)(F(iy) - iy)\big].$$

Since $iyG(iy) \to 1$ and $|F(iy) - iy| \leq C/y$, we see that the first moment of ν is 0, also by the monotone convergence theorem.

We now have that $\sigma^2 \leq C$. The inequality $|z - F(z)| \leq C/\mathrm{Im}(z)$ precludes ν being a Dirac mass other than δ_0. For $\nu = \delta_0$, we have $F(z) = z$, and then the minimal C is clearly $0 = \sigma^2$. Hence we can restrict to $\nu \neq \delta_0$, hence to ν not being a Dirac mass. Thus by Lemma 19 we have for $z \in \mathbb{C}^+$ that $z - F(z) \in \mathbb{C}^-$. By equation (3.6), $\lim_{z\to\infty} z(z - F(z)) = \sigma^2$ in any Stolz angle. Hence by Theorem 10 there is a probability measure $\tilde{\nu}$ such that $z - F(z) = \sigma^2 \int (z-t)^{-1} \, d\tilde{\nu}(t)$. Hence

$$|z - F(z)| \leq \sigma^2 \int \frac{1}{|z-t|} \, d\tilde{\nu}(t) \leq \sigma^2 \int \frac{1}{\mathrm{Im}(z)} \, d\tilde{\nu}(t) = \frac{\sigma^2}{\mathrm{Im}(z)}.$$

This proves the last claim. □

Exercise 16. Suppose ν has a fourth moment and we write

$$G(z) = \frac{1}{z} + \frac{\alpha_1}{z^2} + \frac{\alpha_2}{z^3} + \frac{\alpha_3}{z^4} + \frac{\alpha_4}{z^5} + o\left(\frac{1}{z^5}\right).$$

Show that

$$z - \frac{1}{G(z)} = \alpha_1 + \frac{\beta_0}{z} + \frac{\beta_1}{z^2} + \frac{\beta_2}{z^3} + o\Big(\frac{1}{z^3}\Big)$$

where

$$\beta_0 = \alpha_2 - \alpha_1^2 \qquad \beta_1 = \alpha_3 - 2\alpha_1\alpha_2 + \alpha_1^3 \qquad \beta_2 = \alpha_4 - 2\alpha_1\alpha_3 - \alpha_2^2 + 3\alpha_1^2\alpha_2 - \alpha_1^4$$

and thus conclude that the probability measure ν_1 of Lemma 20 has the second moment β_2/β_0.

Remark 22. We have seen that if ν has a second moment, then we may write

$$G(z) = \frac{1}{(z-\alpha_1) - (\alpha_2 - \alpha_1^2)\int \frac{1}{z-t}\, d\nu_1(t)} = \frac{1}{(z-a_1) - b_1 \int \frac{1}{z-t}\, d\nu_1(t)}$$

where ν_1 is a probability measure on $\mathbb{R}$, and $a_1 = \alpha_1$, $b_1 = \alpha_2 - \alpha_1^2$. If ν has a fourth moment, then ν_1 will have a second moment, and we may repeat our construction to write

$$\int \frac{1}{z-t}\, d\nu_1(t) = \frac{1}{(z-a_2) - b_2 \int \frac{1}{z-t}\, d\nu_2(t)}$$

for some probability measure ν_2, where $a_2 = (\alpha_3 - 2\alpha_1\alpha_2 + \alpha_1^3)/(\alpha_2 - \alpha_1^2)$ and $b_2 = (\alpha_2\alpha_4 + 2\alpha_1\alpha_2\alpha_3 - \alpha_2^3 - \alpha_1^2\alpha_4 - \alpha_3^2)/(\alpha_2 - \alpha_1^2)^2$. Thus

$$G(z) = \cfrac{1}{z - a_1 - \cfrac{b_1}{z - a_2 - b_2 \int \frac{1}{z-t}\, d\nu_2(t)}}.$$

If ν has moments of all orders $\{\alpha_n\}_n$, then the Cauchy transform of ν has a continued fraction expansion (often called a J-fraction because of the connection with Jacobi matrices).

$$G(z) = \cfrac{1}{z - a_1 - \cfrac{b_1}{z - a_2 - \cfrac{b_2}{z - a_3 - \cdots}}}.$$

The coefficients $\{a_n\}_n$ and $\{b_n\}_n$ are obtained from the moments $\{\alpha_n\}_n$ as follows. Let A_n be the $(n+1) \times (n+1)$ Hankel matrix

$$A_n = \begin{bmatrix} 1 & \alpha_1 & \cdots & \alpha_n \\ \alpha_1 & \alpha_2 & \cdots & \alpha_{n+1} \\ \vdots & \vdots & & \vdots \\ \alpha_n & \alpha_{n+1} & \cdots & \alpha_{2n} \end{bmatrix}$$

and $\tilde{A}_{n-1}$ be the $n \times n$ matrix obtained from A_n by deleting the last row and second last column and $\tilde{A}_0 = (\alpha_1)$. Then let $\Delta_{-1} = 1$, $\Delta_n = \det(A_n)$, $\tilde{\Delta}_{-1} = 0$, and $\tilde{\Delta}_n = \det(\tilde{A}_n)$. By Hamburger's theorem (see Shohat and Tamarkin [157, Thm. 1.2]), we have that for all n, $\Delta_n \geq 0$. Then $b_1 b_2 \cdots b_n = \Delta_n / \Delta_{n-1}$ and

$$a_1 + a_2 + \cdots + a_n = \tilde{\Delta}_{n-1}/\Delta_{n-1},$$

or equivalently $b_n = \Delta_{n-2}\Delta_n/\Delta_{n-1}^2$ and

$$a_n = \tilde{\Delta}_{n-1}/\Delta_{n-1} - \tilde{\Delta}_{n-2}/\Delta_{n-2}.$$

If for some n, $\Delta_n = 0$ then we only get a finite continued fraction.

Notation 23. *For $\beta > 0$ let $\mathbb{C}_\beta^+ = \{z \mid \mathrm{Im}(z) > \beta\}$.*

Lemma 24. *Suppose $F : \mathbb{C}^+ \to \mathbb{C}^+$ is analytic and there is $\sigma > 0$ such that for $z \in \mathbb{C}^+$ we have $|z - F(z)| \leq \sigma^2/\mathrm{Im}(z)$. Then*

(i) $\mathbb{C}_{2\sigma}^+ \subset F(\mathbb{C}_\sigma^+)$;
(ii) for each $w \in \mathbb{C}_{2\sigma}^+$, there is a unique $z \in \mathbb{C}_\sigma^+$ such that $F(z) = w$.

Hence there is an analytic function, $F^{\langle -1 \rangle}$, defined on $\mathbb{C}_{2\sigma}^+$ such that $F(F^{\langle -1 \rangle}(w)) = w$. Moreover for $w \in \mathbb{C}_{2\sigma}^+$

(iii) $\mathrm{Im}(F^{\langle -1 \rangle}(w)) \leq \mathrm{Im}(w) \leq 2\,\mathrm{Im}(F^{\langle -1 \rangle}(w))$, and
(iv) $|F^{\langle -1 \rangle}(w) - w| \leq 2\sigma^2/\mathrm{Im}(w)$.

Proof: Suppose $\mathrm{Im}(w) > 2\sigma$. If $|z - w| = \sigma$ then

$$\mathrm{Im}(z) \geq \mathrm{Im}(w - i\sigma) = \mathrm{Im}(w) - \sigma > 2\sigma - \sigma = \sigma.$$

Let C be the circle with centre w and radius σ. Then $C \subset \mathbb{C}_\sigma^+$. For $z \in C$ we have

$$|(F(z) - w) - (z - w)| = |F(z) - z| \leq \frac{\sigma^2}{\mathrm{Im}(z)} < \sigma = |z - w|.$$

Thus by Rouché's theorem there is a unique $z \in \mathrm{int}(C)$ with $F(z) = w$. This proves (i).

If $z' \in \mathbb{C}_\sigma^+$ and $F(z') = w$ then

$$|w - z'| = |F(z') - z'| \leq \frac{\sigma^2}{\mathrm{Im}(z')} < \sigma,$$

so $z' \in \mathrm{int}(C)$ and hence $z = z'$. This proves (ii). We define $F^{\langle -1 \rangle}(w) = z$.

By Lemma 21, $1/F$ is the Cauchy transform of a probability measure with finite variance. Thus by Lemma 19 we have that $\operatorname{Im}(F(z)) \geq \operatorname{Im}(z)$ thus $\operatorname{Im}(F^{\langle -1\rangle}(w)) \leq \operatorname{Im}(w)$. On the other hand, by replacing σ in (i) by $\beta > \sigma$, one has for $w \in \mathbb{C}^+_{2\beta}$ that $\operatorname{Im}(F^{\langle -1\rangle}(w)) > \beta$. By letting 2β approach $\operatorname{Im}(w)$, we get that $\operatorname{Im}(F^{\langle -1\rangle}(w)) \geq \frac{1}{2}\operatorname{Im}(w)$. This proves (iii).

For $w \in \mathbb{C}^+_{2\sigma}$ let $z = F^{\langle -1\rangle}(w) \in \mathbb{C}^+_\sigma$, then by (iii)

$$|F^{\langle -1\rangle}(w) - w| = |z - w| = |F(z) - z| \leq \frac{\sigma^2}{\operatorname{Im}(z)} \leq \frac{2\sigma^2}{\operatorname{Im}(w)}.$$

This proves (iv). □

Theorem 25. *Let ν be a probability measure on $\mathbb{R}$ with first and second moments α_1 and α_2. Let $G(z) = \int (z-t)^{-1}\, d\nu(t)$ be the Cauchy transform of ν and $\sigma^2 = \alpha_2 - \alpha_1^2$ be the variance of ν. Let $F(z) = 1/G(z)$, then $|F(z) + \alpha_1 - z| \leq \sigma^2/\operatorname{Im}(z)$. Moreover there is an analytic function $G^{\langle -1\rangle}$ defined on $\{z \mid |z + i(4\sigma)^{-1}| < (4\sigma)^{-1}\}$ such that $G(G^{\langle -1\rangle}(z)) = z$.*

Proof: Let $\widetilde{G}(z) = G(z + \alpha_1)$. Then $\widetilde{G}$ is the Cauchy transform of a centred probability measure. Let $\widetilde{F}(z) = 1/\widetilde{G}$ then $\widetilde{F} : \mathbb{C}^+ \to \mathbb{C}^+$. By Lemma 20 there is a probability measure $\tilde{\nu}$ such that $z - \widetilde{F}(z) = \sigma^2 \int (z-t)^{-1}\, d\tilde{\nu}(t)$. Thus

$$|z - \widetilde{F}(z)| \leq \int \frac{\sigma^2}{|z-t|}\, d\tilde{\nu}(t) \leq \int \frac{\sigma^2}{\operatorname{Im}(z)}\, d\tilde{\nu}(t) = \frac{\sigma^2}{\operatorname{Im}(z)}.$$

Then $|F(z) + \alpha_1 - z| \leq \sigma^2/\operatorname{Im}(z)$.

If we apply Lemma 24, we get an inverse for $\widetilde{F}$ on $\{z \mid \operatorname{Im}(z) > 2\sigma\}$. Note that $|z + i(4\sigma)^{-1}| < (4\sigma)^{-1}$ if and only if $\operatorname{Im}(1/z) > 2\sigma$. Since $G(z) = 1/\tilde{F}(z - \alpha_1)$ we let $G^{\langle -1\rangle}(z) = \widetilde{F}^{\langle -1\rangle}(1/z) + \alpha_1$ for $|z + i(4\sigma)^{-1}| < (4\sigma)^{-1}$. Then

$$G\left(G^{\langle -1\rangle}(z)\right) = G\left(\widetilde{F}^{\langle -1\rangle}(1/z) + \alpha_1\right) = \widetilde{G}\left(\widetilde{F}^{\langle -1\rangle}(1/z)\right) = \frac{1}{\widetilde{F}\left(\widetilde{F}^{\langle -1\rangle}(1/z)\right)} = z.$$

□

In the next theorem, we show that with the assumption of finite variance σ^2, we can find an analytic function R which solves the equation $G(R(z) + 1/z) = z$ on the open disc with centre $-i(4\sigma)^{-1}$ and radius $(4\sigma)^{-1}$. This is the *R-transform* of the measure.

Theorem 26. *Let ν be a probability measure with variance σ^2. Then on the open disc with centre $-i(4\sigma)^{-1}$ and radius $(4\sigma)^{-1}$ there is an analytic function $R(z)$ such that $G(R(z) + 1/z) = z$ for $|z + i(4\sigma)^{-1}| < (4\sigma)^{-1}$ where G the Cauchy transform of ν.*

Proof: Let $G^{\langle -1\rangle}$ be the inverse provided by Theorem 25 and $R(z) = G^{\langle -1\rangle}(z) - 1/z$. Then $G(R(z) + 1/z) = G(G^{\langle -1\rangle}(z)) = z$. □

One should note that the statements and proofs of Theorems 25 and 26, interpreted in the right way, remain also valid for the degenerated case $\sigma^2 = 0$,

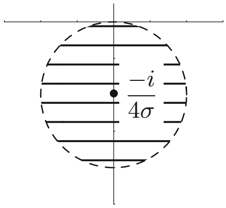


Fig. 3.3 If a probability measure ν has finite variance σ^2, then the R-transform of ν is analytic on a disc in the lower half-plane with centre $-i(4\sigma)^{-1}$ and passing through 0

where ν is a Delta mass. Then $G^{\langle -1\rangle}$ and R are defined on the whole lower half-plane $\mathbb{C}_-$; actually, for $\nu = \delta_{\alpha_1}$ we have $R(z) = \alpha_1$.

3.5 The free additive convolution of probability measures with finite variance

One of the main ideas of free probability is that if we have two self-adjoint operators a_1 and a_2 in a unital C^*-algebra with state φ and if a_1 and a_2 are free with respect to φ, then we can find the moments of $a_1 + a_2$ from the moments of a_1 and a_2 according to a universal rule. Since a_1, a_2 and $a_1 + a_2$ are all bounded self-adjoint operators there are probability measures ν_1, ν_2, and ν such that for $i = 1, 2$

$$\varphi(a_i^k) = \int t^k \, d\nu_i(t) \qquad \text{and} \qquad \varphi((a_1 + a_2)^k) = \int t^k \, d\nu(t).$$

We call ν the *free additive convolution* of ν_1 and ν_2 and denote it $\nu_1 \boxplus \nu_2$. An important observation is that because of the universal rule, $\nu_1 \boxplus \nu_2$ only depends on ν_1 and ν_2 and not on the operators a_1 and a_2 used to construct it. For bounded operators we also know that the free additive convolution can be described by the additivity of their R-transforms. We shall show in this section how to construct $\nu_1 \boxplus \nu_2$ without assuming that the measures have compact support and thus without using Banach algebra techniques. As we have seen in the last section, we can still define an R-transform by analytic means (at least for the case of finite variance); the idea is then of course to define $\nu = \nu_1 \boxplus \nu_2$ by prescribing the R-transform of ν as the sum of the R-transforms of ν_1 and ν_2. However, it is then not at all obvious that there actually exists a probability measure with this prescribed R-transform. In order to see that this is indeed the case, we have to reformulate our description in terms of the R-transform in a subordination form, as already alluded to in (2.31) at the end of the last chapter.

Recall that the R-transform in the compactly supported case satisfied the equation $G(R(z) + 1/z) = z$ for $|z|$ sufficiently small. So letting $F(z) = 1/G(z)$, this becomes $F(R(z) + 1/z) = z^{-1}$. For $|z|$ sufficiently small $G^{\langle -1\rangle}(z)$ is defined, and hence also $F^{\langle -1\rangle}(z^{-1})$; then for such z we have

$$R(z) = F^{\langle -1\rangle}(z^{-1}) - z^{-1}. \tag{3.7}$$

If ν_1 and ν_2 are compactly supported with Cauchy transforms G_1 and G_2 and corresponding F_1, F_2, R_1, and R_2, then we have the equation $R(z) = R_1(z) + R_2(z)$ for $|z|$ small; this implies

$$F^{\langle -1 \rangle}\left(z^{-1}\right) - z^{-1} = F_1^{\langle -1 \rangle}\left(z^{-1}\right) - z^{-1} + F_2^{\langle -1 \rangle}\left(z^{-1}\right) - z^{-1}.$$

If we let $w_1 = F_1^{\langle -1 \rangle}(z^{-1})$, $w_2 = F_2^{\langle -1 \rangle}(z^{-1})$, and $w = F^{\langle -1 \rangle}(z^{-1})$, this equation becomes

$$w - F(w) = w_1 - F_1(w_1) + w_2 - F_2(w_2).$$

Since $z^{-1} = F(w) = F_1(w_1) = F_2(w_2)$, we can write this as

$$F_1(w_1) = F_2(w_2) \qquad \text{and} \qquad w = w_1 + w_2 - F_1(w_1). \tag{3.8}$$

We shall now show, given two probability measures ν_1 and ν_2 with finite variance, we can construct a probability measure ν with finite variance such that $R = R_1 + R_2$, the R-transforms of ν, ν_1, and ν_2, respectively.

Given $w \in \mathbb{C}^+$, we shall show in Lemma 27 that there are w_1 and w_2 in $\mathbb{C}^+$ such that (3.8) holds. Then we define F by $F(w) = F_1(w_1)$ and show that $1/F$ is the Cauchy transform of a probability measure of finite variance. This measure will then be the free additive convolution of ν_1 and ν_2. Moreover the maps $w \mapsto w_1$ and $w \mapsto w_2$ will be the subordination maps of equation (2.31).

We need the notion of the degree of an analytic function which we summarize in the exercise below.

Exercise 17. Let X be a Riemann surface and $f : X \to \mathbb{C}$ an analytic map. Let us recall the definition of the *multiplicity* of f at z_0 in X (see, e.g. Miranda [133, Ch. II, Def. 4.2]). There is $m \geq 0$ and a chart $(\mathcal{U}, \varphi)$ of z_0 such that $\varphi(z_0) = 0$ and $f(\varphi^{\langle -1 \rangle}(z)) = z^m + f(z_0)$ for z in $\varphi(\mathcal{U})$. We set $\mathrm{mult}(f, z_0) = m$. For each z in $\mathbb{C}$, we define the *degree* of f at z, denoted $\deg_f(z)$, by

$$\deg_f(z) = \sum_{w \in f^{\langle -1 \rangle}(z)} \mathrm{mult}(f, w).$$

It is a standard theorem that if X is compact, then $\deg_f$ is constant (see, e.g. Miranda [133, Ch. II, Prop. 4.8]).

(i) Adapt the proof in the compact case to show that if X is not necessarily compact but f is proper, i.e. if the inverse image of a compact set is compact, then $\deg_f$ is constant.

(ii) Suppose that $F_1, F_2 : \mathbb{C}^+ \to \mathbb{C}^+$ are analytic and $F_i'(z) \neq 0$ for $z \in \mathbb{C}^+$ and $i = 1, 2$. Let $X = \{(z_1, z_2) \in \mathbb{C}^+ \times \mathbb{C}^+ \mid F_1(z_1) = F_2(z_2)\}$. Give X the structure of a complex manifold so that $(z_1, z_2) \mapsto F_1(z_1)$ is analytic.

(*iii*) Suppose F_1, F_2, and X are as in (*ii*) and in addition there are σ_1 and σ_2 such that for $i = 1, 2$ and $z \in \mathbb{C}^+$ we have $|z - F_i(z)| \leq \sigma_i^2/\mathrm{Im}(z)$. Show that $\theta : X \to \mathbb{C}$ given by $\theta(z_1, z_2) = z_1 + z_2 - F_1(z_1)$ is a proper map.

Lemma 27. *Suppose F_1 and F_2 are analytic maps from $\mathbb{C}^+$ to $\mathbb{C}^+$ and that there is $r > 0$ such that for $z \in \mathbb{C}^+$ and $i = 1, 2$, we have $|F_i(z) - z| \leq r^2/\mathrm{Im}(z)$. Then for each $z \in \mathbb{C}^+$ there is a unique pair $(z_1, z_2) \in \mathbb{C}^+ \times \mathbb{C}^+$ such that*

(*i*) $F_1(z_1) = F_2(z_2)$, *and*
(*ii*) $z_1 + z_2 - F_1(z_1) = z$.

Proof: Note that, by Lemma 21, our assumptions imply that, for $i = 1, 2$, $1/F_i$ is the Cauchy transform of some probability measure and thus, by Lemma 19, we also know that it satisfies $\mathrm{Im}(z) \leq \mathrm{Im}(F_i(z))$.

We first assume that $z \in \mathbb{C}^+_{4r}$. If (z_1, z_2) satisfies (*i*) and (*ii*),

$$\mathrm{Im}(z_1) = \mathrm{Im}(z) + \mathrm{Im}(F_2(z_2) - z_2) \geq \mathrm{Im}(z).$$

Likewise $\mathrm{Im}(z_2) \geq \mathrm{Im}(z)$. Hence, if we are to find a solution to (*i*) and (*ii*), we shall find it in $\mathbb{C}^+_{4r} \times \mathbb{C}^+_{4r}$. By Lemma 24, F_1 and F_2 are invertible on $\mathbb{C}^+_{2r}$. Thus to find a solution to (*i*) and (*ii*), it is sufficient to find $u \in \mathbb{C}^+_{2r}$ such that

$$F_1^{\langle -1 \rangle}(u) + F_2^{\langle -1 \rangle}(u) - u = z \tag{3.9}$$

and then let $z_1 = F_1^{\langle -1 \rangle}(u)$ and $z_2 = F_2^{\langle -1 \rangle}(u)$. Thus we must show that for every $z \in \mathbb{C}^+_{4r}$, there is a unique $u \in \mathbb{C}^+_{2r}$ satisfying equation (3.9).

Let C be the circle with centre z and radius $2r$. Then $C \subset \mathbb{C}^+_{2r}$ and for $u \in C$ we have by Lemma 24

$$|F_1^{\langle -1 \rangle}(u) - u| + |F_2^{\langle -1 \rangle}(u) - u| \leq \frac{4r^2}{\mathrm{Im}(u)} < \frac{4r^2}{2r} = 2r.$$

Hence

$$\Big|(z - u) - \big[z - u - (F_1^{\langle -1 \rangle}(u) - u) - (F_2^{\langle -1 \rangle}(u) - u)\big]\Big|$$
$$\leq \Big|F_1^{\langle -1 \rangle}(u) - u\Big| + \Big|F_2^{\langle -1 \rangle}(u) - u\Big| < 2r = |z - u|.$$

Thus by Rouché's theorem, there is a unique $u \in \mathrm{int}(C)$ such that

$$z - u = \Big(F_1^{\langle -1 \rangle}(u) - u\Big) + \Big(F_2^{\langle -1 \rangle}(u) - u\Big).$$

If there is $u' \in \mathbb{C}^+_{2r}$ with

$$z - u' = \Big(F_1^{\langle -1 \rangle}(u') - u'\Big) + \Big(F_2^{\langle -1 \rangle}(u') - u'\Big),$$

then, again by Lemma 24,

$$|z-u'| = \left|\left(F_1^{\langle -1\rangle}(u')-u'\right)+\left(F_2^{\langle -1\rangle}(u')-u'\right)\right| < 2r$$

and thus $u' \in \operatorname{int}(C)$ and hence $u' = u$. Thus there is a unique $u \in \mathbb{C}^+_{2r}$ satisfying equation (3.9).

Let $X = \{(z_1,z_2) \mid F_1(z_1) = F_2(z_2)\}$ be the Riemann surface in Exercise 17 and $\theta(z_1,z_2) = z_1+z_2-F_1(z_1)$. We have just shown that for $z \in \mathbb{C}^+_{4r}$, $\deg_\theta(z) = 1$. But by Exercise 17, $\deg_\theta$ is constant on $\mathbb{C}^+$ so there is a unique solution to (*i*) and (*ii*) for all $z \in \mathbb{C}^+$. □

Exercise 18. Let ν be a probability measure with variance σ^2 and mean m. Let $\tilde{\nu}(E) = \nu(E+m)$. Show that $\tilde{\nu}$ is a probability measure with mean 0 and variance σ^2. Let G and $\tilde{G}$ be the corresponding Cauchy transforms. Show that we have $\tilde{G}(z) = G(z+m)$. Let R and $\tilde{R}$ be the corresponding R-transforms. Show that $R(z) = \tilde{R}(z)+m$ for $|z+i(4\sigma)^{-1}| < (4\sigma)^{-1}$.

Theorem 28. *Let ν_1 and ν_2 be two probability measures on $\mathbb{R}$ with finite variances and R_1 and R_2 be the corresponding R-transforms. Then there is a unique probability measure with finite variance, denoted $\nu_1 \boxplus \nu_2$, and called the* free additive convolution *of ν_1 and ν_2, such that the R-transform of $\nu_1 \boxplus \nu_2$ is R_1+R_2.*

Moreover the first moment of $\nu_1 \boxplus \nu_2$ is the sum of the first moments of ν_1 and ν_2 and the variance of $\nu_1 \boxplus \nu_2$ is the sum of the variances of ν_1 and ν_2.

Proof: By Exercise 18 we only have to prove the theorem in the case ν_1 and ν_2 are centred. Moreover there are probability measures ρ_1 and ρ_2 such that for $z \in \mathbb{C}^+$ and $i = 1,2$ we have $z - F_i(z) = \sigma_i^2 \int (z-t)^{-1}\, d\rho_i(t)$. By Lemma 27 for each z in $\mathbb{C}^+$ there is a unique pair $(z_1,z_2) \in \mathbb{C}^+ \times \mathbb{C}^+$ such that $F_1(z_1) = F_2(z_2)$ and $z_1+z_2-F_1(z_1) = z$. Define $F(z) = F_1(z_1)$. Let $X = \{(z_1,z_2) \mid F_1(z_1) = F_2(z_2)\}$ and $\theta : X \to \mathbb{C}^+$ be as in Exercise 17. In Lemma 27 we showed that θ is an analytic bijection, since $\deg(\theta) = 1$. Then $F = F_1 \circ \pi \circ \theta^{-1}$ where $\pi(z_1,z_2) = z_1$. Thus F is analytic on $\mathbb{C}^+$ and we have

$$z - F(z) = z_1 - F_1(z_1) + z_2 - F_2(z_2). \tag{3.10}$$

Since $\operatorname{Im}(F_1(z)) \ge \operatorname{Im}(z)$, we have $\operatorname{Im}(z) = \operatorname{Im}(z_2) + \operatorname{Im}(z_1 - F_1(z)) \le \operatorname{Im}(z_2)$. Likewise $\operatorname{Im}(z) \le \operatorname{Im}(z_1)$. Thus

$$|z-F(z)| = |z_1 - F_1(z_1) + z_2 - F_2(z_2)| \le \frac{\sigma_1^2}{\operatorname{Im}(z_1)} + \frac{\sigma_2}{\operatorname{Im}(z_2)} \le \frac{\sigma_1^2+\sigma_2^2}{\operatorname{Im}(z)}.$$

Therefore, by Lemma 21, $1/F$ is the Cauchy transform of a centred probability measure with variance $\sigma^2 \leq \sigma_1^2 + \sigma_2^2$. Thus there is, by Lemma 20, a probability measure ρ such that

$$z - F(z) = \sigma^2 \int \frac{1}{z-t}\, d\rho(t).$$

So by equation (3.10)

$$\sigma^2 \int \frac{1}{z-t}\, d\rho(t) = \sigma_1^2 \int \frac{1}{z_1 - t}\, d\rho_1(t) + \sigma_2^2 \int \frac{1}{z_2 - t}\, d\rho_2(t)$$

and hence

$$\sigma^2 \int \frac{z}{z-t}\, d\rho(t) = \sigma_1^2 \int \frac{z}{z_1 - t}\, d\rho_1(t) + \sigma_2^2 \int \frac{z}{z_2 - t}\, d\rho_2(t). \tag{3.11}$$

For $z = iy$, we have $z/F(z) \to \infty$ by Exercise 12 (*ii*). Also

$$\left| F_1^{\langle -1 \rangle}(F(z)) - F(z) \right| \leq \frac{2\sigma^2}{\operatorname{Im}(F(z))} \leq \frac{2\sigma^2}{\operatorname{Im}(z)}$$

by Lemma 24, parts (*iii*) and (*iv*). Thus

$$\frac{z_1}{z} = \frac{F_1^{\langle -1 \rangle}(F(z)) - F(z)}{z} + \frac{F(z)}{z}.$$

The first term goes to 0 and the second term goes to 1 as $y \to \infty$, hence $z_1/z \to 1$. Likewise $z_2/z \to 1$. Thus

$$\lim_{y\to\infty} \int \frac{iy}{z_1 - t}\, d\rho_1(t) = 1 \qquad \text{and likewise} \qquad \lim_{y\to\infty} \int \frac{iy}{z_2 - t}\, d\rho_2(t) = 1.$$

If we now take limits as $y \to \infty$ in equation (3.11), we get $\sigma^2 = \sigma_1^2 + \sigma_2^2$.

Let $D_r = \{z \mid |z + ir| < r\}$, then $D_{1/(4\sigma)} \subset D_{1/(4\sigma_1)} \cap D_{1/(4\sigma_2)}$. Let $z \in D_{1/(4\sigma)}$, then z^{-1} is in the domains of $F^{\langle -1 \rangle}$, $F_1^{\langle -1 \rangle}$, and $F_2^{\langle -1 \rangle}$. Now by Lemma 27, for $F^{\langle -1 \rangle}(z^{-1})$ find z_1 and z_2 in $\mathbb{C}^+$ so that $F_1(z_1) = F_2(z_2)$ and $F^{\langle -1 \rangle}(z^{-1}) = z_1 + z_2 - F_1(z_1)$. By the construction of F, we have $z^{-1} = F(F^{\langle -1 \rangle}(z^{-1})) = F_1(z_1) = F_2(z_2)$ and so $z_1 = F_1^{\langle -1 \rangle}(z^{-1})$ and $z_2 = F_2^{\langle -1 \rangle}(z^{-1})$. Thus the equation $F^{\langle -1 \rangle}(z^{-1}) = z_1 + z_2 - F_1(z_1)$ becomes

$$F^{\langle -1 \rangle}\left(z^{-1}\right) - z^{-1} = F_1^{\langle -1 \rangle}\left(z^{-1}\right) - z^{-1} + F_2^{\langle -1 \rangle}\left(z^{-1}\right) - z^{-1}.$$

Now recall the construction of the R-transform given by Theorem 26, reformulated as in (3.7) in terms of F: $R(z) = F^{\langle -1 \rangle}(z^{-1}) - z^{-1}$. Hence $R(z) = R_1(z) + R_2(z)$. $\square$

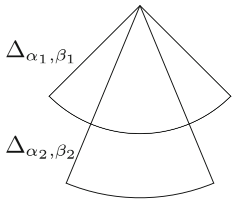


Fig. 3.4 Two wedges in $\mathbb{C}^-$: $\Delta_{\alpha_1,\beta_1}$ and $\Delta_{\alpha_2,\beta_2}$ with $\alpha_1 > \alpha_2$ and $\beta_1 > \beta_2$. We have $R^{(1)}$ on $\Delta_{\alpha_1,\beta_1}$ and $R^{(2)}$ on $\Delta_{\alpha_2,\beta_2}$ such that $R^{(1)}(z) = R^{(2)}(z)$ for $z \in \Delta_{\alpha_1,\beta_1} \cap \Delta_{\alpha_2,\beta_2}$. We shall denote the germ by R

3.6 The R-transform and free additive convolution of arbitrary measures

In this section we consider probability measures on $\mathbb{R}$ that may not have any moments. We first show that for all $\alpha > 0$, there is $\beta > 0$ so that the R-transform can be defined on the wedge $\Delta_{\alpha,\beta}$ in the lower half-plane:

$$\Delta_{\alpha,\beta} = \{z^{-1} \mid z \in \Gamma_{\alpha,\beta}\} = \left\{ w \in \mathbb{C}^- \mid |\mathrm{Re}(w)| < -\alpha\, \mathrm{Im}(w) \text{ and } |w + \frac{i}{2\beta}| < \frac{1}{2\beta} \right\}.$$

This means that the R-transform is a germ of analytic functions in that for each $\alpha > 0$, there is $\beta > 0$ and an analytic function R on $\Delta_{\alpha,\beta}$ such that whenever we are given another $\alpha' > 0$ for which there exists a $\beta' > 0$ and a second analytic function R' on $\Delta_{\alpha',\beta'}$, the two functions agree on $\Delta_{\alpha,\beta} \cap \Delta_{\alpha',\beta'}$. See Fig. 3.4.

Definition 29. Let ν be a probability measure on $\mathbb{R}$, and let G be the Cauchy transform of ν. We define the *R-transform* of ν as the germ of analytic functions on the domains $\Delta_{\alpha,\beta}$ satisfying equation (3.1). This means that for all $\alpha > 0$, there is $\beta > 0$ such that for all $z \in \Delta_{\alpha,\beta}$, we have $G(R(z) + 1/z) = z$ and for all $z \in \Gamma_{\alpha,\beta}$, we have $R(G(z)) + 1/G(z) = z$.

Remark 30. When ν is compactly supported, we can find a disc centred at 0 on which there is an analytic function satisfying equation (3.1). This was shown in Theorem 17. When ν has finite variance, we showed that there is a disc in $\mathbb{C}^-$ tangent to 0 and with centre on the imaginary axis (see Fig. 3.3) on which there is an analytic function satisfying equation (3.1). This was shown in Theorem 26. In the general case, we shall define $R(z)$ by the equation $R(z) = F^{\langle -1 \rangle}(z^{-1}) - z^{-1}$. The next two lemmas show that we can find a domain where this definition works.

Lemma 31. *Let F be the reciprocal of the Cauchy transform of a probability measure on $\mathbb{R}$. Suppose $0 < \alpha_1 < \alpha_2$. Then there is $\beta_0 > 0$ such that for all $\beta_2 \geq \beta_0$ and $\beta_1 \geq \beta_2(1 + \alpha_2 - \alpha_1)$,*

(*i*) *we have $\Gamma_{\alpha_1,\beta_1} \subseteq F(\Gamma_{\alpha_2,\beta_2})$*
(*ii*) *and $F^{\langle -1 \rangle}$ exists on $\Gamma_{\alpha_1,\beta_1}$, i.e. for each $w \in \Gamma_{\alpha_1,\beta_1}$ there is a unique $z \in \Gamma_{\alpha_2,\beta_2}$ such that $F(z) = w$.*

Proof: Let $\theta = \tan^{-1}\left(\alpha_1^{-1}\right) - \tan^{-1}\left(\alpha_2^{-1}\right)$. Choose $\epsilon > 0$ so that

$$\epsilon < \sin\theta = \frac{\alpha_2 - \alpha_1}{\sqrt{1+\alpha_1^2}\sqrt{1+\alpha_2^2}}.$$

Choose $\beta_0 > 0$ such that $|F(z) - z| < \epsilon|z|$ for $z \in \Gamma_{\alpha_2,\beta_0}$ (which is possible by Exercise 12). Let $\beta_2 \geq \beta_0$ and $\beta_1 \geq \beta_2(1 + \alpha_2 - \alpha_1)$.

Let us first show that for $w \in \Gamma_{\alpha_1,\beta_1}$ and for $z \in \partial\Gamma_{\alpha_2,\beta_2}$, we have $\epsilon|z| < |z - w|$.

If $z = \alpha_2 y + iy \in \partial\Gamma_{\alpha_2}$, then $|z - w|/|z| \geq \sin\theta > \epsilon$. If $z = x + i\beta_2 \in \partial\Gamma_{\alpha_2,\beta_2}$, then

$$|z-w| > \beta_1 - \beta_2 \geq \beta_2(\alpha_2 - \alpha_1) > \epsilon\beta_2\sqrt{1+\alpha_1^2}\sqrt{1+\alpha_2^2} \geq \epsilon|z|\sqrt{1+\alpha_1^2} > \epsilon|z|.$$

Thus for $w \in \Gamma_{\alpha_1,\beta_1}$ and $z \in \partial\Gamma_{\alpha_2,\beta_2}$, we have $\epsilon|z| < |z - w|$.

Now fix $w \in \Gamma_{\alpha_1,\beta_1}$ and let $r > |w|/(1-\epsilon)$. Thus for $z \in \{\tilde{z} \mid |\tilde{z}| = r\} \cap \Gamma_{\alpha_2,\beta_2}$ we have $|z - w| \geq r - |w| > \epsilon r = \epsilon|z|$. So let C be the curve

$$C := \left(\partial\Gamma_{\alpha_2,\beta_2} \cap \{\tilde{z}| \mid |\tilde{z}| \leq r\}\right) \cup \left(\{\tilde{z} \mid |\tilde{z}| = r\} \cap \Gamma_{\alpha_2,\beta_2}\right).$$

For $z \in C$, we have that $\epsilon|z| < |z - w|$. Thus for $z \in C$ we have

$$|(F(z) - w) - (z - w)| < \epsilon|z| < |z - w|.$$

So by Rouché's theorem, there is exactly one z in the interior of C such that $F(z) = w$. Since we can make r as large as we want, there is a unique $z \in \Gamma_{\alpha_2,\beta_2}$ such that $F(z) = w$. Hence F has an inverse on $\Gamma_{\alpha_1,\beta_1}$. □

Lemma 32. *Let F be the reciprocal of the Cauchy transform of a probability measure on $\mathbb{R}$. Suppose $0 < \alpha_1 < \alpha_2$. Then there is $\beta_0 > 0$ such that*

$$F(\Gamma_{\alpha_1,\beta_1}) \subseteq \Gamma_{\alpha_2,\beta_1} \qquad \textit{for all} \qquad \beta_1 \geq \beta_0.$$

Proof: Choose $1/2 > \epsilon > 0$ so that

$$\alpha_2 > \frac{\alpha_1 + \epsilon/\sqrt{1-\epsilon^2}}{1 - \alpha_1\epsilon/\sqrt{1-\epsilon^2}} > \alpha_1.$$

Then choose $\beta_0 > 0$ such that $|F(z) - z| < \epsilon|z|$ for $z \in \Gamma_{\alpha_1,\beta_0}$.

Suppose $\beta_1 \geq \beta_0$ and let $z \in \Gamma_{\alpha_1,\beta_1}$ with $\mathrm{Re}(z) \geq 0$, (the case $\mathrm{Re}(z) < 0$ is similar). Write $z = |z|e^{i\varphi}$. Then $\varphi > \tan^{-1}(\alpha_1^{-1})$. Write $F(z) = |F(z)|e^{i\psi}$. We have $|z^{-1}F(z) - 1| < \epsilon$. Thus $|\sin(\psi - \varphi)| < \epsilon$, so

$$\psi > \varphi - \sin^{-1}(\epsilon) > \tan^{-1}(\alpha_1^{-1}) - \sin^{-1}(\epsilon).$$

If $\psi \leq \pi/2$, then

$$\tan(\psi) > \tan\left(\tan^{-1}(\alpha_1^{-1}) - \sin^{-1}(\epsilon)\right) = \frac{\alpha_1^{-1} - \epsilon/\sqrt{1-\epsilon^2}}{1 + \alpha_1^{-1}\epsilon/\sqrt{1-\epsilon^2}} > \alpha_2^{-1}.$$

Thus $F(z) \in \Gamma_{\alpha_2}$.

Suppose $\psi \geq \pi/2$. Then we must show that $\pi - \psi > \tan^{-1}(\alpha_2^{-1})$ or equivalently that $\tan(\pi - \psi) > \alpha_2^{-1}$. Since $|\psi - \phi| < \sin^{-1}(\epsilon)$ and $\phi \leq \pi/2$ we must then have $\pi - \psi > \pi/2 - \sin^{-1}(\epsilon)$. Thus

$$\tan(\pi - \psi) > \tan(\pi/2 - \sin^{-1}(\epsilon)) = \sqrt{1-\epsilon^2}/\epsilon.$$

On the other hand,

$$\alpha_2 > \alpha_1 + \epsilon/\sqrt{1-\epsilon^2} > \epsilon/\sqrt{1-\epsilon^2},$$

so $\tan(\pi - \psi) > \alpha_2^{-1}$ as required. Thus in both cases $F(z) \in \Gamma_{\alpha_2}$.

Since we also have $\mathrm{Im}(F(z)) \geq \mathrm{Im}(z) > \beta_1$, we obtain $F(\Gamma_{\alpha_1,\beta_1}) \subseteq \Gamma_{\alpha_2,\beta_1}$. □

Theorem 33. *Let ν be a probability measure on $\mathbb{R}$ with Cauchy transform G and set $F = 1/G$. For every $\alpha > 0$, there is $\beta > 0$ so that $R(z) = F^{\langle -1\rangle}(z^{-1}) - z^{-1}$ is defined for $z \in \Delta_{\alpha,\beta}$ and such that we have:*

(i) $G(R(z) + 1/z) = z$ for $z \in \Delta_{\alpha,\beta}$ and
(ii) $R(G(z)) + 1/G(z) = z$ for $z \in \Gamma_{\alpha,\beta}$.

Proof: Let $F(z) = 1/G(z)$. Let $\alpha > 0$ be given and by Lemma 31 choose $\beta_0 > 0$ so that $F^{\langle -1\rangle}$ is defined on $\Gamma_{2\alpha,\beta_0}$. For $z \in \Delta_{2\alpha,\beta_0}$, $R(z)$ is thus defined and we have $G(R(z) + 1/z) = G(F^{\langle -1\rangle}(z^{-1})) = z$.

Now by Lemma 32, we may choose $\beta > \beta_0$ such that $F(\Gamma_{\alpha,\beta}) \subseteq \Gamma_{2\alpha,\beta}$. For $z \in \Gamma_{\alpha,\beta}$, we have $G(z) = 1/F(z) \in \Gamma_{2\alpha,\beta}^{-1} = \Delta_{2\alpha,\beta} \subseteq \Delta_{2\alpha,\beta_0}$ and so

$$R(G(z)) + 1/G(z) = F^{\langle -1\rangle}(F(z)) - F(z) + F(z) = z.$$

Since $\Delta_{\alpha,\beta} \subset \Delta_{2\alpha,\beta_0}$, we also have $G(R(z) + 1/z) = z$ for $z \in \Delta_{\alpha,\beta}$. □

Exercise 19. Let $w \in \mathbb{C}$ be such that $\mathrm{Im}(w) \leq 0$. Then we saw in Exercise 9 that $G(z) = (z - w)^{-1}$ is the Cauchy transform of a probability measure on $\mathbb{R}$. Show that the R-transform of this measure is $R(z) = w$. In this case R is defined on all of $\mathbb{C}$ even though the corresponding measure has no moments (when $\mathrm{Im}(w) < 0$).

Remark 34. We shall now show that given two probability measures ν_1 and ν_2 with R-transforms R_1 and R_2, respectively, we can find a third probability measure ν with Cauchy transform G and R-transform R such that $R = R_1 + R_2$. This means that for all $\alpha > 0$, there is $\beta > 0$ such that all three of R, R_1, and R_2 are defined on $\Delta_{\alpha,\beta}$ and for $z \in \Delta_{\alpha,\beta}$ we have $R(z) = R_1(z) + R_2(z)$. We shall denote ν by $\nu_1 \boxplus \nu_2$ and call it the *free additive convolution* of ν_1 and ν_2. Clearly, this extends then our definition for probability measures with finite variance from the last section.

When ν_1 is a Dirac mass at $a \in \mathbb{R}$, we can dispose of this case directly. An easy calculation shows that $R_1(z) = a$, c.f. Exercise 1. So $R(z) = a + R_2(z)$ and thus $G(z) = G_2(z-a)$. Thus $\nu(E) = \nu_2(E-a)$, c.f. Exercise 18. So for the rest of this section, we shall assume that neither ν_1 nor ν_2 is a Dirac mass.

There is another case that we can easily deal with. Suppose $\mathrm{Im}(w) < 0$. Let $\nu_1 = \delta_w$ be the probability measure with Cauchy transform $G_1(z) = (z-w)^{-1}$. This is the measure we discussed in Notation 4; see also Exercises 9 and 19. Then $R_1(z) = w$. Let ν_2 be any probability measure on $\mathbb{R}$. We let G_2 be the Cauchy transform of ν_2 and R_2 be its R-transform. So if $\nu_1 \boxplus \nu_2$ exists, its R-transform should be $R(z) = w + R_2(z)$. Let us now go back to the subordination formula (2.31) in Chapter 2. It says that if $\nu_1 \boxplus \nu_2$ exists, its Cauchy transform, G, should satisfy $G(z) = G_2(\omega_2(z))$ where $\omega_2(z) = z - R_1(G(z)) = z - w$. Now ω_2 maps $\mathbb{C}^+$ to $\mathbb{C}^+$ and letting $G = G_2 \circ \omega_2$ we have

$$\lim_{y\to\infty} iy\, G(iy) = 1.$$

So by Theorem 10, there is a measure, which we shall denote $\nu_1 \boxplus \nu_2$, of which G is the Cauchy transform, and thus the R-transform of this measure satisfies, by construction, the equation $R = R_1 + R_2$. Note that in this special case, we have $\delta_w \boxplus \nu_2 = \delta_w * \nu_2$ where $*$ means the classical convolution, because they have the same Cauchy transform $G(z) = G_2(z-w)$; see Notation 4, Theorem 6, and Exercise 9. We can also solve for ω_1 to conclude that $\omega_1(z) = F_2(z-w) + w$. For later reference we shall summarize this calculation in the theorem below.

Theorem 35. *Let $w = a + ib \in \overline{\mathbb{C}^-}$ and δ_w be the probability measure on $\mathbb{R}$ with density*

$$d\delta_w(t) = \frac{1}{\pi}\frac{-b}{b^2 + (t-a)^2}\, dt$$

when $b < 0$ and the Dirac mass at a when $b = 0$. Then for any probability measure ν, we have $\delta_w \boxplus \nu = \delta_w * \nu$.

In the remainder of this chapter, we shall define $\nu_1 \boxplus \nu_2$ in full generality; for this we will show that we can always find ω_1 and ω_2 satisfying (2.32).

Notation 36. *Let ν_1, ν_2 be probability measures on $\mathbb{R}$ with Cauchy transforms G_1 and G_2, respectively. Let $F_i(z) = 1/G_i(z)$ and $H_i(z) = F_i(z) - z$. The functions F_1, F_2, H_1, and H_2 are analytic functions that map the upper half-plane $\mathbb{C}^+$ to itself.*

Corollary 37. *Let F_1 and F_2 be as in Notation 36. Suppose $0 < \alpha_2 < \alpha_1$. Then there are $\beta_2 \geq \beta_0 > 0$ such that*

(i) $F_1^{\langle -1\rangle}$ is defined on $\Gamma_{\alpha_1,\beta_1}$ for any $\beta_1 \geq \beta_0$ with $F_1^{\langle -1\rangle}(\Gamma_{\alpha_1,\beta_1}) \subseteq \Gamma_{\alpha_1+1,\beta_1/2}$;
(ii) $F_2(\Gamma_{\alpha_2,\beta_2}) \subseteq \Gamma_{\alpha_1,\beta_0}$.

Proof: Let $\alpha = \alpha_1 + 1$. By Lemma 31 there is $\beta_0/2 > 0$ such that for all $\beta \geq \beta_0/2$ and $\beta_1 = \beta(1 + \alpha - \alpha_1) = 2\beta \geq \beta_0$, we have $\Gamma_{\alpha_1,\beta_1} \subseteq F_1(\Gamma_{\alpha,\beta})$ and $F_1^{\langle -1\rangle}$ exists on $\Gamma_{\alpha_1,\beta_1}$; thus $F_1^{\langle -1\rangle}(\Gamma_{\alpha_1,\beta_1}) \subseteq \Gamma_{\alpha_1+1,\beta_1/2}$. By Lemma 32 choose now $\beta_2 > 0$ (and also $\beta_2 \geq \beta_0$) so that $F_2(\Gamma_{\alpha_2,\beta_2}) \subseteq \Gamma_{\alpha_1,\beta_2} \subseteq \Gamma_{\alpha_1,\beta_0}$. □

Definition 38. For any $z, w \in \mathbb{C}^+$ let $g(z,w) = z + H_1(z + H_2(w))$. Then $g : \mathbb{C}^+ \times \mathbb{C}^+ \to \mathbb{C}^+$ is analytic. Let $g_z(w) = g(z,w)$.

Remark 39. Choose now some $\alpha_1 > \alpha_2 > 0$, and $\beta_2 \geq \beta_0 \geq 0$ according to Corollary 37. In the following we will also need to control $\mathrm{Im}(F_2(w) - w)$. Note that, by the fact that $F_2(w)/w \to 1$, for $w \to \infty$ in $\Gamma_{\alpha_2,\beta_2}$, we have, for any $\epsilon < 1$, $|F_2(w) - w| < \epsilon|w|$ for sufficiently large $w \in \Gamma_{\alpha_2,\beta_2}$. But then

$$0 \leq \mathrm{Im}(F_2(w) - w) \leq |F_2(w) - w| < \varepsilon|w| < \varepsilon\sqrt{1 + \alpha_2^2} \cdot \mathrm{Im}(w);$$

the latter inequality is from Notation 15 for $w \in \Gamma_{\alpha_2,\beta_2}$. By choosing $1/\varepsilon = 2\sqrt{1 + \alpha_2^2}$, we find thus a $\beta > 0$ (which we can take $\beta \geq \beta_2$) such that we have

$$\mathrm{Im}(F_2(w) - w) < \frac{1}{2}\mathrm{Im}(w) \qquad \text{for all } w \in \Gamma_{\alpha_2,\beta} \subseteq \Gamma_{\alpha_2,\beta_2}. \tag{3.12}$$

Consider now for $w \in \Gamma_{\alpha_2,\beta}$ the point $z = w + F_1^{\langle -1\rangle}(F_2(w)) - F_2(w)$. Since $F_2(w) \in \Gamma_{\alpha_1,\beta_0}$, this is well-defined. Furthermore, we have $\mathrm{Im}(F_2(w)) \geq \mathrm{Im}(w) > \beta \geq \beta_0$, and thus actually $F_2(w) \in \Gamma_{\alpha_1,\mathrm{Im}(w)}$, which then yields

$$F_1^{\langle -1\rangle}(w) \in \Gamma_{\alpha_1+1,\mathrm{Im}(w)/2} \qquad \text{i.e.} \qquad \mathrm{Im}(F_1^{\langle -1\rangle}(w)) > \frac{\mathrm{Im}(w)}{2}.$$

This together with (3.12) shows that we have $z \in \mathbb{C}^+$, whenever we choose $w \in \Gamma_{\alpha_2,\beta}$.

Lemma 40. *With α_2 and β as above, let $w \in \Gamma_{\alpha_2,\beta}$. Then*

$$z = w + F_1^{\langle -1\rangle}(F_2(w)) - F_2(w) \iff g(z,w) = w.$$

Proof: Suppose $z = w + F_1^{\langle -1\rangle}(F_2(w)) - F_2(w)$. By Remark 39 we have $z \in \mathbb{C}^+$.

$$\begin{aligned}
g(z,w) &= z + H_1(z + H_2(w)) \\
&= z + H_1(z + F_2(w) - w) \\
&= z + H_1(F_1^{\langle -1\rangle}(F_2(w))) \\
&= z + F_1(F_1^{\langle -1\rangle}(F_2(w))) - F_1^{\langle -1\rangle}(F_2(w)) \\
&= z + F_2(w) - F_1^{\langle -1\rangle}(F_2(w)) \\
&= w.
\end{aligned}$$

Suppose $g(z, w) = w$. Then

$$w = g(z, w) = w + F_1(z + F_2(w) - w) - F_2(w)$$

so

$$F_2(w) = F_1(z + F_2(w) - w)$$

thus

$$F_1^{\langle -1 \rangle}(F_2(w)) = z + F_2(w) - w$$

as required. □

Remark 41. By Lemma 40 the open set

$$\Omega = \left\{ w + F_1^{\langle -1 \rangle}(F_2(w)) - F_2(w) \mid w \in \Gamma_{\alpha_2,\beta} \right\} \subseteq \mathbb{C}^+$$

is such that for $z \in \Omega$, g_z has a fixed point in $\mathbb{C}^+$ (even in $\Gamma_{\alpha_2,\beta}$). Our goal is to show that for every $z \in \mathbb{C}^+$, there is w such that $g_z(w) = w$ and that w is an analytic function of z.

Exercise 20. In the next proof, we will use the following simple part of the Denjoy-Wolff Theorem. Suppose $f : \mathbb{D} \to \mathbb{D}$ is a non-constant holomorphic function on the unit disc $\mathbb{D} := \{z \in \mathbb{C} \mid |z| < 1\}$ and it is not an automorphism of $\mathbb{D}$ (i.e. not of the form $\lambda(z - \alpha)/(1 - \bar{\alpha}z)$ for some $\alpha \in \mathbb{D}$ and $\lambda \in \mathbb{C}$ with $|\lambda| = 1$). If there is a $z_0 \in \mathbb{D}$ with $f(z_0) = z_0$, then for all $z \in \mathbb{D}$, $f^{\circ n}(z) \to z_0$. In particular, the fixed point is unique.

Prove this by an application of the Schwarz Lemma.

Lemma 42. *Let $g(z, w)$ be as in Definition 38. Then there is a non-constant analytic function $f : \mathbb{C}^+ \to \mathbb{C}^+$ such that for all $z \in \mathbb{C}^+$, $g(z, f(z)) = f(z)$. The analytic function f is uniquely determined by the fixed point equation.*

Proof: As before we set

$$\Omega = \left\{ w + F_1^{\langle -1 \rangle}(F_2(w)) - F_2(w) \mid w \in \Gamma_{\alpha_2,\beta} \right\} \subseteq \mathbb{C}^+$$

and let, for $z \in \mathbb{C}^+$, $g_z : \mathbb{C}^+ \to \mathbb{C}^+$ be $g_z(u) = g(z, u)$.

The idea of the proof is to define the fixed point of the function g_z by iterations. For $z \in \Omega$, we already know that we have a fixed point; hence, the version of Denjoy-Wolff mentioned above gives the convergence of the iterates in this case. The extension of the statement to all $z \in \mathbb{C}^+$ is then provided by an application of Vitali's Theorem. A minor inconvenience comes from the fact that we have to transcribe our situation from the upper half-plane to the disc, in order to apply the theorems mentioned above. This is achieved by composing our functions with

$$\varphi(z) = i\frac{1+z}{1-z} \qquad \text{and} \qquad \psi(z) = \frac{z-i}{z+i};$$

φ maps $\mathbb{D}$ onto $\mathbb{C}^+$ and ψ maps $\mathbb{C}^+$ onto $\mathbb{D}$; they are inverses of each other.

Let us first consider $z \in \Omega$. Let $\tilde{g}_z : \mathbb{D} \to \mathbb{D}$ be given by $\tilde{g}_z = \psi \circ g_z \circ \varphi$. Since $z \in \Omega$ there exists an $w \in \Gamma_{\alpha_2,\beta}$ with $z = w + F_1^{\langle -1 \rangle}(F_2(w)) - F_2(w)$. Let $\tilde{w} = \psi(w)$. Then

$$\tilde{g}_z(\tilde{w}) = \psi(g_z(\varphi(\tilde{w}))) = \psi(g_z(w)) = \psi(w) = \tilde{w}.$$

So the map $\tilde{g}_z$ has a fixed point in $\mathbb{D}$. In order to apply the Denjoy-Wolff theorem, we have to exclude that $\tilde{g}_z$ is an automorphism. But since we have for all $w \in \mathbb{C}^+$

$$\mathrm{Im}(g_z(w)) = \mathrm{Im}(z) + \mathrm{Im}(H_1(z + H_2(w))) \geq \mathrm{Im}(z),$$

it is clear that g_z cannot be an automorphism of the upper half-plane and hence $\tilde{g}_z$ cannot be an automorphism of the disc. Hence, by Denjoy-Wolff, $\tilde{g}_z^{\circ n}(\tilde{u}) \to \tilde{w}$ for all $\tilde{u} \in \mathbb{D}$. Converting back to $\mathbb{C}^+$ we see that $g_z^{\circ n}(u) \to w$ for all $u \in \mathbb{C}^+$.

Now we define our iterates on all of $\mathbb{C}^+$, where we choose for concreteness the initial point as $u_0 = i$. We define a sequence $\{f_n\}_n$ of analytic functions from $\mathbb{C}^+$ to $\mathbb{C}^+$ by $f_n(z) = g_z^{\circ n}(i)$. We claim that for all $z \in \mathbb{C}^+$, $\lim_n f_n(z)$ exists. We have shown that already for $z \in \Omega$. There $z = w + F_1^{\langle -1 \rangle}(F_2(w)) - F_2(w)$ with $w \in \Gamma_{\alpha_2,\beta}$, and $g_z^{\circ n}(i) \to w$. Thus for all $z \in \Omega$ the sequence $\{f_n(z)\}_n$ converges to the corresponding w. Now let $\tilde{\Omega} = \psi(\Omega)$ and $\tilde{f}_n = \psi \circ f_n \circ \varphi$. then $\tilde{f}_n : \mathbb{D} \to \mathbb{D}$ and for $\tilde{z} \in \tilde{\Omega}$, $\lim_n \tilde{f}_n(\tilde{z})$ exists. Hence, by Vitali's Theorem, $\lim_n \tilde{f}_n(\tilde{z})$ exists for all $\tilde{z} \in \mathbb{D}$. Note that by the maximum modulus principle this limit cannot take on values on the boundary of $\mathbb{D}$ unless it is constant. Since it is clearly not constant on $\tilde{\Omega}$, the limit takes on only values in $\mathbb{D}$. Hence $\lim_n f_n(z)$ exists for all $z \in \mathbb{C}^+$ as an element in $\mathbb{C}^+$. So we define $f : \mathbb{C}^+ \to \mathbb{C}^+$ by $f(z) = \lim_n f_n(z)$; by Vitali's Theorem the convergence is uniform on compact subsets of $\mathbb{C}^+$ and f is analytic. Recall that $f_n(z) = g_z^{\circ n}(i)$, so

$$g_z(f(z)) = \lim_n g_z(f_n(z)) = \lim_n g_z^{\circ(n+1)}(i) = f(z),$$

so we have $g(z, f(z)) = g_z(f(z)) = f(z)$.

By Denjoy-Wolff, the function f is uniquely determined by the fixed point equation on the open set Ω; by analytic continuation it is then unique everywhere.

$\square$

Theorem 43. *There are analytic functions $\omega_1, \omega_2 : \mathbb{C}^+ \to \mathbb{C}^+$ such that for all $z \in \mathbb{C}^+$*

(*i*) $F_1(\omega_1(z)) = F_2(\omega_2(z))$, *and*
(*ii*) $\omega_1(z) + \omega_2(z) = z + F_1(\omega_1(z))$.

The analytic functions ω_1 and ω_2 are uniquely determined by these two equations.

Proof: Let $z \in \mathbb{C}^+$ and $g_z(w) = g(z, w)$. By Lemma 42, g_z has a unique fixed point $f(z)$. So define the function ω_2 by $\omega_2(z) = f(z)$ for $z \in \mathbb{C}^+$, and the function ω_1 by $\omega_1(z) = z + F_2(\omega_2(z)) - \omega_2(z)$. Then ω_1 and ω_2 are analytic on $\mathbb{C}^+$ and

$$\omega_1(z) + \omega_2(z) = z + F_2(\omega_2(z)).$$

By Lemma 40, we have that for $z \in \Omega$, $z = \omega_2(z) + F_1^{\langle -1 \rangle}(F_2(\omega_2(z))) - F_2(\omega_2(z))$ and by construction $z = \omega_2(z) + \omega_1(z) - F_2(\omega_2(z))$. Hence for $z \in \Omega$, $\omega_1(z) = F_1^{\langle -1 \rangle}(F_2(\omega_2(z)))$. Thus for all $z \in \Omega$, and hence by analytic continuation for all $z \in \mathbb{C}^+$, we have $F_1(\omega_1(z)) = F_2(\omega_2(z))$ as required.

For the uniqueness one has to observe that the equations (*i*) and (*ii*) yield

$$\omega_1(z) = z + F_2(\omega_2(z)) - \omega_2(z) = z + H_2(\omega_2(z))$$

and

$$\omega_2(z) = z + F_1(\omega_1(z)) - \omega_1(z) = z + H_1(\omega_1(z)),$$

and thus

$$\omega_2(z) = z + H_1(z + H_2(\omega_2(z))) = g(z, \omega_2(z)).$$

By Lemma 42, we know that an analytic solution of this fixed point equation is unique. Exchanging H_1 and H_2 gives in the same way the uniqueness of ω_1. □

To define the free additive convolution of ν_1 and ν_2, we shall let $F(z) = F_1(\omega_1(z)) = F_2(\omega_2(z))$ and then show that $1/F$ is the Cauchy transform of a probability measure, which will be $\nu_1 \boxplus \nu_2$. The main difficulty is to show that $F(z)/z \to 1$ as $\sphericalangle z \to \infty$. For this we need the following lemma.

Lemma 44. $\displaystyle \lim_{y\to\infty} \frac{\omega_1(iy)}{iy} = \lim_{y\to\infty} \frac{\omega_2(iy)}{iy} = 1.$

Proof: Let us begin by showing that $\lim_{y\to\infty} \omega_2(iy) = \infty$ (in the sense of Definition 16).

We must show that given $\alpha, \beta > 0$ there is $y_0 > 0$ such that $\omega_2(iy) \subseteq \Gamma_{\alpha,\beta}$ whenever $y > y_0$. Note that by the previous Theorem we have $\omega_2(z) = z + H_1(\omega_1(z)) \in z + \mathbb{C}^+$. So we have that $\mathrm{Im}(\omega_2(z)) > \mathrm{Im}(z)$. Since ω_2 maps $\mathbb{C}^+$ to $\mathbb{C}^+$ we have by the Nevanlinna representation of ω_2 (see Exercise 13) that $b_2 =$

$\lim_{y\to\infty}\omega_2(iy)/(iy) \geq 0$. This means that $\mathrm{Im}(\omega_2(iy))/y \to b_2$ and our inequality $\mathrm{Im}(\omega_2(z)) > \mathrm{Im}(z)$ implies that $b_2 \geq 1$. We also have that $\mathrm{Re}(\omega_2(iy))/y \to 0$. So there is $y_0 > 0$ so that for $y > y_0 \geq \beta$ we have

$$\left(\frac{\mathrm{Re}(\omega_2(iy))}{y}\right)^2 + \left(\frac{\mathrm{Im}(\omega_2(iy))}{y}\right)^2 < (\alpha^2+1)\left(\frac{\mathrm{Im}(\omega_2(iy))}{y}\right)^2.$$

For such a y we have

$$\frac{|\omega_2(iy)|^2}{y^2} < (1+\alpha^2)\left(\frac{\mathrm{Im}(\omega_2(iy))}{y}\right)^2.$$

Thus $\omega_2(iy) \in \Gamma_\alpha$ (see Notation 15). Since $\mathrm{Im}(\omega_2(iy)) > y > y_0$, we have that $\omega_2(iy) \in \Gamma_{\alpha,\beta}$. Thus $\lim_{y\to\infty}\omega_2(iy) = \infty$.

Recall that $\omega_1(z) = z + H_2(\omega_2(z)) \in z + \mathbb{C}^+$, so by repeating our arguments above, we have that $b_1 = \lim_{y\to\infty}\omega_1(iy)/(iy) \geq 1$ and $\lim_{y\to\infty}\omega_1(iy) = \infty$.

Since $\lim_{\sphericalangle z\to\infty} F_1(z)/z = 1$ (see Exercise 12), we now have $\lim\limits_{y\to\infty}\dfrac{F_1(\omega_1(iy))}{\omega_2(iy)} = 1$. Moreover the equation $\omega_1(z) + \omega_2(z) = z + F_1(\omega_1(z))$ means that

$$\begin{aligned}
b_1 + b_2 &= \lim_{y\to\infty}\frac{\omega_1(iy)+\omega_2(iy)}{iy} \\
&= \lim_{y\to\infty}\frac{iy + F_1(\omega_1(iy))}{iy} \\
&= 1 + \lim_{y\to\infty}\frac{F_1(\omega_1(iy))}{\omega_1(iy)}\frac{\omega_1(iy)}{iy} \\
&= 1 + b_1.
\end{aligned}$$

Thus $b_2 = 1$. By the same argument, we have $b_1 = 1$. □

Theorem 45. *Let $F = F_2 \circ \omega_2$. Then F is the reciprocal of the Cauchy transform of a probability measure.*

Proof: We have that F maps $\mathbb{C}^+$ to $\mathbb{C}^+$ so by Theorem 10 we must show that $\lim_{y\to\infty}F(iy)/(iy) = 1$. By Lemma 44

$$\lim_{y\to\infty}\frac{F(iy)}{iy} = \lim_{y\to\infty}\frac{F_2(\omega_2(iy))}{iy} = \lim_{y\to\infty}\frac{F_2(\omega_2(iy))}{\omega_2(iy)}\frac{\omega_2(iy)}{iy} = 1.$$

□

Theorem 46. *Let ν_1 and ν_2 be two probability measures on $\mathbb{R}$ then there is ν, a probability measure on $\mathbb{R}$ with R-transform R, such that $R = R_1 + R_2$.*

Proof: Let $F = F_2 \circ \omega_2 = F_1 \circ \omega_1$ be as in Theorem 45 and ν its corresponding probability measure. By Theorem 43 (*ii*) we have

$$\omega_1(F^{\langle -1\rangle}(z^{-1})) + \omega_2(F^{\langle -1\rangle}(z^{-1})) - F_1(\omega_1(F^{\langle -1\rangle}(z^{-1}))) = F^{\langle -1\rangle}(z^{-1}).$$

Also $\omega_1(F^{\langle -1\rangle}(z^{-1})) = F_1^{\langle -1\rangle}(z^{-1})$ and $\omega_2(F^{\langle -1\rangle}(z^{-1})) = F_2^{\langle -1\rangle}(z^{-1})$ so our equation becomes

$$F_1^{\langle -1\rangle}(z^{-1}) + F_2^{\langle -1\rangle}(z^{-1}) - z^{-1} = F^{\langle -1\rangle}(z^{-1}).$$

Hence $R(z) = R_1(z) + R_2(z)$. □

Definition 47. Let $\nu_1 \boxplus \nu_2$ be the probability measure whose Cauchy transform is the reciprocal of F, i.e. for which we have $R = R_1 + R_2$. We call $\nu_1 \boxplus \nu_2$ the *free additive convolution* of ν_1 and ν_2.

Remark 48. In the case of bounded operators x and y which are free, we saw in Section 3.5 that the distribution of their sum gives the free additive convolution of their distributions. Later we shall see how using the theory of unbounded operators affiliated with a von Neumann algebra we can have the same conclusion for probability measures with non-compact support (see Remark 8.16).

Remark 49. 1) There is also a similar analytic theory of *free multiplicative convolution* $\boxtimes$ for the product of free variables; see, for example, [21, 30, 54].

2) There exists a huge body of results around infinitely divisible and stable laws in the free sense; see, for example, [8–10, 22, 29–32, 53, 70, 97, 199].

Chapter 4
Asymptotic Freeness for Gaussian, Wigner, and Unitary Random Matrices

After having developed the basic theory of freeness, we are now ready to have a more systematic look into the relation between freeness and random matrices. In Chapter 1, we showed the asymptotic freeness between independent Gaussian random matrices. This is only the tip of an iceberg. There are many more classes of random matrices which show asymptotic freeness. In particular, we will present such results for Wigner matrices, Haar unitary random matrices and treat also the relation between such ensembles and deterministic matrices. Furthermore, we will strengthen the considered form of freeness from the averaged version (which we considered in Chapter 1) to an almost sure one.

We should point out that our presentation of the notion of freeness is quite orthogonal to its historical development. Voiculescu introduced this concept in an operator algebraic context (we will say more about this in Chapter 6); at the beginning of free probability, when Voiculescu discovered the R-transform and proved the free central limit theorem around 1983, there was no relation at all with random matrices. This connection was only revealed later in 1991 by Voiculescu [180]; he was motivated by the fact that the limit distribution which he found in the free central limit theorem had appeared before in Wigner's semi-circle law in the random matrix context. The observation that operator algebras and random matrices are deeply related had a tremendous impact and was the beginning of new era in the subject of free probability.

4.1 Asymptotic freeness: averaged convergence versus almost sure convergence

The most important random matrix is the GUE random matrix ensemble A_N. Let us recall what this means. Each entry of A_N is a complex-valued random variable a_{ij}, and $a_{ji} = \overline{a_{ij}}$ for $i \neq j$, while $a_{ii} = \overline{a_{ii}}$ thus implying that a_{ii} is in fact a real-valued random variable. A_N is said to be GUE-distributed if each a_{ij} with $i < j$ is of the form

J.A. Mingo, R. Speicher, *Free Probability and Random Matrices*,
Fields Institute Monographs 35, DOI 10.1007/978-1-4939-6942-5_4

$$a_{ij} = x_{ij} + \sqrt{-1}y_{ij}, \tag{4.1}$$

where $x_{ij}, y_{ij}, 1 \leq i < j \leq N$ are independent real Gaussian random variables, each with mean 0 and variance $1/(2N)$. This also determines the entries below the diagonal. Moreover, the GUE requirement means that the diagonal entries a_{ii} are real-valued independent Gaussian random variables which are also independent from the x_{ij}'s and the y_{ij}'s and have mean 0 and variance $1/N$.

Let tr be the normalized trace on the full $N \times N$ matrix algebra over $\mathbb{C}$. Then $\mathrm{tr}(A_N)$ is a random variable. In Chapter 1 we proved Wigner's semi-circle law; namely, that

$$\lim_{N\to\infty} E[\mathrm{tr}(A_N^m)] = \begin{cases} \frac{1}{n+1}\binom{2n}{n}, & m = 2n \\ 0, & m \text{ odd} \end{cases}.$$

In the language we have developed in Chapter 2 (see Definition 2.1), this means that $A_N \xrightarrow{\text{distr}} s$, as $N \to \infty$, where the convergence is in distribution with respect to $E \circ \mathrm{tr}$ and s is a semi-circular element in some non-commutative probability space.

We also saw Voiculescu's remarkable generalization of Wigner's semi-circle law: if $A_N^{(1)}, \ldots, A_N^{(p)}$ are p independent $N \times N$ GUE random matrices (meaning that if we collect the real and imaginary parts of the above diagonal entries together with the diagonal entries, we get a family of independent real Gaussians with mean 0 and variances as explained above), then

$$A_N^{(1)}, \ldots, A_N^{(p)} \xrightarrow{\text{distr}} s_1, \ldots, s_p \text{ as } N \to \infty, \tag{4.2}$$

where $s_1, \ldots, s_p$ is a family of freely independent semi-circular elements. This amounts to proving that for all $m \in \mathbb{N}$ and all $1 \leq i_1, \ldots, i_m \leq p$, we have

$$\lim_{N\to\infty} E[\mathrm{tr}(A_N^{(i_1)} \cdots A_N^{(i_m)}] = \varphi(s_{i_1} \cdots s_{i_m}).$$

Recall that since $s_1, \ldots, s_p$ are free, their mixed cumulants will vanish, and only the second cumulants of the form $\kappa_2(s_i, s_i)$ will be non-zero. With the chosen normalization of the variance for our random matrices, those second cumulants will be 1. Thus,

$$\varphi(s_{i_1} \cdots s_{i_m}) = \sum_{\pi \in NC_2(m)} \kappa_\pi[s_{i_1}, \ldots, s_{i_m}]$$

is given by the number of non-crossing pairings of the $s_{i_1}, \ldots, s_{i_m}$ which connect only s_i's with the same index. Hence (4.2) follows from Lemma 1.9.

The statements above about the limit distribution of Gaussian random matrices are in distribution with respect to the averaged trace $E[\mathrm{tr}(\cdot)]$. However, they also hold in the stronger sense of almost sure convergence. Before formalizing this,

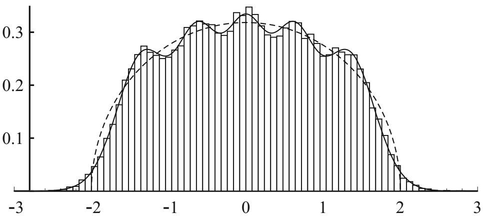


Fig. 4.1 Averaged distribution for 10,000 realizations with $N = 5$. The dashed line is the semi-circle law, and the solid line is the limit as the number of realizations tends to infinity

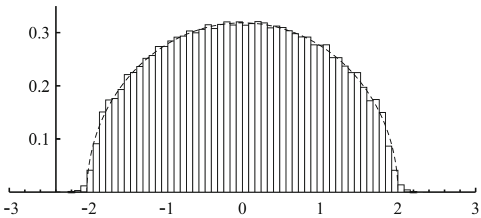


Fig. 4.2 Averaged distribution for 10,000 realizations with $N = 20$. The dashed line is the semi-circle law

let us first look at some numerical simulations in order to get an idea of the difference between *convergence of averaged eigenvalue distribution* and *almost sure convergence of eigenvalue distribution*.

Consider first our usual setting with respect to $E[\mathrm{tr}(\cdot)]$. To simulate this, we have to average for fixed N the eigenvalue distributions of the sampled $N \times N$ matrices. For the Gaussian ensemble, there are infinitely many of those, so we approximate this averaging by choosing a large number of realizations of our random matrices. In the following pictures, we created 10,000 $N \times N$ matrices (by generating the entries independently and according to a normal distribution), calculated for each of those 10,000 matrices the N eigenvalues, and plotted the histogram for the $10{,}000 \times N$ eigenvalues. We show those histograms for $N = 5$ (see Fig. 4.1) and $N = 20$ (see Fig. 4.2). Wigner's theorem in the averaged version tells us that as $N \to \infty$ these averaged histograms have to converge to the semi-circle. The numerical simulations show this very clearly. Note that already for quite small N, for example, $N = 20$, we have a very good agreement with the semi-circular distribution.

Let us now consider the stronger almost sure version of this. In that case, we produce for each N only one $N \times N$ matrix (generated according to the probability measure for our ensemble) and plot the corresponding histogram of the N eigenvalues. The almost sure version of Wigner's theorem says that generically, i.e. for almost all choices of such sequences of $N \times N$ matrices, the corresponding sequence of histograms converges to the semi-circle. This statement is supported by the following pictures of four such samples, for $N = 10$, $N = 100$, $N = 1000$, $N = 4000$ (see Figs. 4.3 and 4.4). Clearly, for small N, the histogram depends on the specific realization of our random matrix, but the larger N gets, the smaller the variations between different realizations get.

Also for the asymptotic freeness of independent Gaussian random matrices, we have an almost sure version. Consider two independent Gaussian random matrices

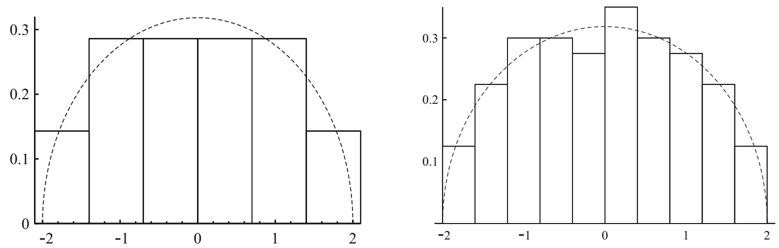


Fig. 4.3 One realization of a $N = 10$ and a $N = 100$ Gaussian random matrix

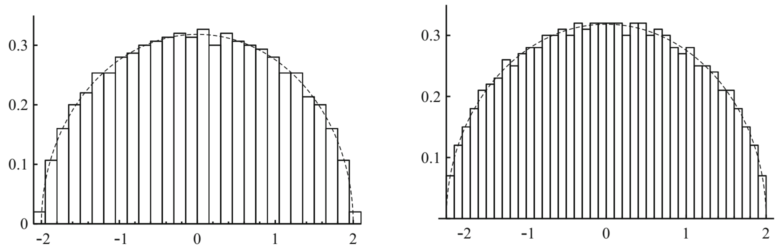


Fig. 4.4 One realization of a $N = 1000$ and a $N = 4000$ Gaussian random matrix

A_N and B_N. We have seen that $A_N, B_N \xrightarrow{\text{distr}} s_1, s_2$, where s_1, s_2 are free semicircular elements.

This means, for example, that

$$\lim_{N\to\infty} E[\mathrm{tr}(A_N A_N B_N B_N A_N B_N B_N A_N)] = \varphi(s_1 s_1 s_2 s_2 s_1 s_2 s_2 s_1).$$

We have $\varphi(s_1 s_1 s_2 s_2 s_1 s_2 s_2 s_1) = 2$, since there are two non-crossing pairings which respect the indices:

The numerical simulation in the first part of the following figure shows the averaged (over 1000 realizations) value of $\mathrm{tr}(A_N A_N B_N B_N A_N B_N B_N A_N)$, plotted against N, for N between 2 and 30. Again, one sees (Fig. 4.5 left) a very good agreement with the asymptotic value of 2 for quite small N.

For the almost sure version of this, we realize for each N just one matrix A_N and (independently) one matrix B_N and calculate for this pair the number $\mathrm{tr}(A_N A_N B_N B_N A_N B_N B_N A_N)$. We expect that generically, as $N \to \infty$, this should also converge to 2. The second part of the above figure shows a simulation for this (Fig. 4.5 right).

Let us now formalize our two notions of asymptotic freeness. For notational convenience, we restrict here to two sequences of matrices. The extension to more random matrices or to sets of random matrices should be clear.

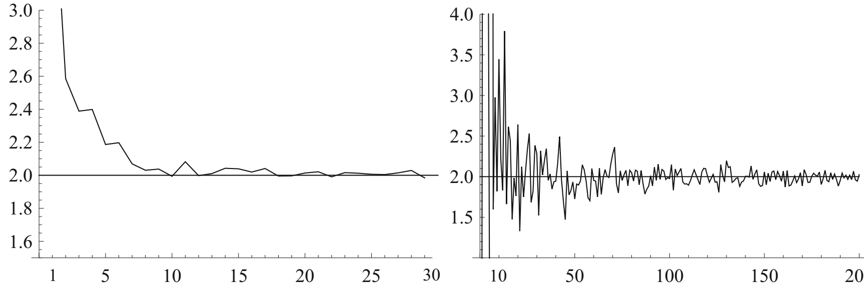


Fig. 4.5 On the left, we have the averaged trace (averaged over 1000 realizations) of the normalized trace of $X_N = A_N A_N B_N B_N A_N B_N B_N A_N$ for N from 1 to 30. On the right, the normalized trace of X_N for N from 1 to 200 (one realization for each N)

Definition 1. Consider two sequences $(A_N)_{N\in\mathbb{N}}$ and $(B_N)_{N\in\mathbb{N}}$ of random $N \times N$ matrices such that for each $N \in \mathbb{N}$, A_N and B_N are defined on the same probability space (Ω_N, P_N). Denote by E_N the expectation with respect to P_N.

1) We say A_N and B_N are *asymptotically free* if $A_N, B_N \in (\mathcal{A}_N, E_N[\mathrm{tr}(\cdot)])$ (where $\mathcal{A}_N$ is the algebra generated by the random matrices A_N and B_N) converge in distribution to some elements a, b (living in some non-commutative probability space $(\mathcal{A}, \varphi)$) such that a, b are free.
2) Consider now the product space $\Omega = \prod_{N\in\mathbb{N}} \Omega_N$ and let $P = \prod_{N\in\mathbb{N}} P_N$ be the product measure of the P_N on Ω. Then we say that A_N and B_N are *almost surely asymptotically free*, if there exists a, b (in some non-commutative probability space $(\mathcal{A}, \varphi)$) which are free and such that we have for almost all $\omega \in \Omega$ that $A_N(\omega), B_N(\omega) \in (M_N(\mathbb{C}), \mathrm{tr}(\cdot))$ converge in distribution to a, b.

Remark 2. What does this mean concretely? Assume we are given our two sequences A_N and B_N and we want to investigate their convergence to some a and b, where a and b are free. Then, for any choice of $m \in \mathbb{N}$ and $p_1, q_1, \dots, p_m, q_m \geq 0$, we have to consider the trace of the corresponding monomial,

$$f_N := \mathrm{tr}(A_N^{q_1} B_N^{p_1} \cdots A_N^{q_m} B_N^{p_m}),$$

and show that this converges to the corresponding expression

$$\alpha := \varphi(a^{q_1} b^{p_1} \cdots a^{q_m} b^{p_m}).$$

For asymptotic freeness, we have to show the convergence of $E_N[f_N]$ to α, whereas in the almost sure case, we have to strengthen this to the almost sure convergence of $\{f_N\}_N$. In order to do so, one usually shows that the variance of the random variables f_N goes to zero fast enough. Namely, assume that $E_N[f_N]$ converges to α; then the fact that $f_N(\omega)$ does not converge to α is equivalent to the fact that the difference between $f_N(\omega)$ and $\alpha_N := E_N[f_N]$ does not converge to zero. But this is the same as the statement that for some $\varepsilon > 0$ we have $|f_N(\omega) - \alpha_N| \geq \varepsilon$ for infinitely many N. Thus, the almost sure convergence of $\{f_N\}_N$ is equivalent to the fact that for any $\varepsilon > 0$

$$P(\{\omega \mid |f_N(\omega) - \alpha_N| \geq \varepsilon \text{ infinitely often}\}) = 0.$$

As this is the probability of the lim sup of events, we can use the first Borel-Cantelli lemma which guarantees that this probability is zero if we have

$$\sum_{N \in \mathbb{N}} P(\{\omega \mid |f_N(\omega) - \alpha_N| \geq \varepsilon\}) < \infty.$$

Note that, since the f_N are independent with respect to P, this is by the second Borel-Cantelli lemma actually equivalent to the almost sure convergence of f_N. On the other hand, Chebyshev's inequality gives us the bound (since $\alpha_N = E[f_N]$)

$$P_N(\{\omega \mid |f_N(\omega) - \alpha_N| \geq \varepsilon\}) \leq \frac{1}{\varepsilon^2} \mathrm{var}[f_N].$$

So if we can show that $\sum_{N \in \mathbb{N}} \mathrm{var}[f_N] < \infty$, then we are done. Usually, one is able to bound the order of these variances by a constant times $1/N^2$, which is good enough.

We will come back to the question of estimating the variances in Remark 5.14. In Theorem 5.13, we will show the variances are of order $1/N^2$, as claimed above. (Actually we will do more there and provide a non-crossing interpretation of the coefficient of this leading order term.) So in the following, we will usually only address the asymptotic freeness of the random matrices under consideration in the averaged sense and postpone questions about the almost sure convergence to Chapter 5. However, in all cases considered, the averaged convergence can be strengthened to almost sure convergence, and we will state our theorems directly in this stronger form.

Remark 3. There is actually another notion of convergence which might be more intuitive than almost sure convergence, namely, *convergence in probability*. Namely, our random matrices A_N and B_N converge in probability to a and b (and hence, if a and b are free, are *asymptotically free in probability*), if we have for each $\epsilon > 0$ that

$$\lim_{N \to \infty} P_N(\{\omega \mid |f_N(\omega) - \alpha_N| \geq \varepsilon\}) = 0.$$

As before, we can use Chebyshev's inequality to insure this convergence if we can show that $\lim_N \mathrm{var}[f_N] = 0$.

It is clear that convergence in probability is weaker than almost sure convergence. Since our variance estimates are usually strong enough to insure almost sure convergence, we will usually state our theorems in terms of almost sure convergence. Almost sure versions of the various theorems were also considered in [96, 160, 173].

4.2 Asymptotic freeness of Gaussian random matrices and deterministic matrices

Consider a sequence $(A_N)_{N\in\mathbb{N}}$ of $N \times N$ GUE random matrices A_N; then we know that $A_N \xrightarrow{\text{distr}} s$. Consider now also a sequence $(D_N)_{N\in\mathbb{N}}$ of deterministic (i.e. non-random) matrices, $D_N \in M_N(\mathbb{C})$. Assume that

$$\lim_{N\to\infty} \operatorname{tr}(D_N^m) \tag{4.3}$$

exists for all $m \geq 1$. Then we have $D_N \xrightarrow{\text{distr}} d$, as $N \to \infty$, where d lives in some non-commutative probability space and where the moments of d are given by the limit moments (4.3) of the D_N. We want to investigate the question whether there is anything definite to say about the relation between s and d.

In order to answer this question, we need to find out whether the limiting mixed moments

$$\lim_{N\to\infty} E[\operatorname{tr}(D_N^{q(1)} A_N D_N^{q(2)} \cdots D_N^{q(m)} A_N)], \tag{4.4}$$

for all $m \geq 1$ (where $q(k)$ can be 0 for some k) exist. In the calculation, let us suppress the dependence on N to reduce the number of indices and write

$$D_N^{q(k)} = (d_{ij}^{(k)})_{i,j=1}^N \qquad \text{and} \qquad A_N = (a_{ij})_{i,j=1}^N. \tag{4.5}$$

The Wick formula allows us to calculate mixed moments in the entries of A :

$$E[a_{i_1 j_1} a_{i_2 j_2} \cdots a_{i_m j_m}] = \sum_{\pi\in\mathcal{P}_2(m)} \prod_{(r,s)\in\pi} E[a_{i_r j_r} a_{i_s j_s}], \tag{4.6}$$

where

$$E[a_{ij} a_{kl}] = \delta_{il}\delta_{jk}\frac{1}{N}. \tag{4.7}$$

Thus, we have

$$\begin{aligned}
E[\operatorname{tr}(D_N^{q(1)} A_N D_N^{q(2)} \cdots D_N^{q(m)} A_N)] &= \frac{1}{N} \sum_{i,j:[m]\to[N]} E[d_{j_1 i_1}^{(1)} a_{i_1 j_2} d_{j_2 i_2}^{(2)} a_{i_2 j_3} \cdots d_{j_m i_m}^{(m)} a_{i_m j_1}] \\
&= \frac{1}{N} \sum_{i,j:[m]\to[N]} E[a_{i_1 j_2} a_{i_2 j_3} \cdots a_{i_m j_1}] d_{j_1 i_1}^{(1)} \cdots d_{j_m i_m}^{(m)} \\
&= \frac{1}{N^{1+m/2}} \sum_{\pi\in\mathcal{P}_2(m)} \sum_{i,j:[m]\to[N]} \prod_{r=1}^{m} \delta_{i_r j_{\gamma\pi(r)}} d_{j_1 i_1}^{(1)} \cdots d_{j_m i_m}^{(m)} \\
&= \frac{1}{N^{1+m/2}} \sum_{\pi\in\mathcal{P}_2(m)} \sum_{j:[m]\to[N]} d_{j_1 j_{\gamma\pi(1)}}^{(1)} \cdots d_{j_m j_{\gamma\pi(m)}}^{(m)}.
\end{aligned}$$

In the calculation above, we regard a pairing $\pi \in \mathcal{P}_2(m)$ as a product of disjoint transpositions in the permutation group S_m (i.e. an involution without fixed point). Also $\gamma \in S_m$ denotes the long cycle $\gamma = (1, 2, \ldots, m)$, and $\#(\sigma)$ is the number of cycles in the factorization of $\sigma \in S_m$ as a product of disjoint cycles.

In order to get a simple formula for the expectation, we need a simple expression for

$$\sum_{j:[m]\to[N]} d^{(1)}_{j_1 j_{\gamma\pi(1)}} \cdots d^{(m)}_{j_m j_{\gamma\pi(m)}}.$$

As always, tr is the normalized trace, and we extend it multiplicatively as a functional on permutations. For example, if $\sigma = (1, 6, 3)(4)(2, 5) \in S_6$, then

$$\mathrm{tr}_\sigma[D_1, D_2, D_3, D_4, D_5, D_6] = \mathrm{tr}(D_1 D_6 D_3)\mathrm{tr}(D_4)\mathrm{tr}(D_2 D_5).$$

In terms of matrix elements, we have the following which we leave as an easy exercise.

Exercise 1. Let $A_1, \ldots, A_n$ be $N \times N$ matrices and let $\sigma \in S_n$ be a permutation. Let the entries of A_k be $(a^{(k)}_{ij})^N_{i,j=1}$. Show that

$$\mathrm{tr}_\sigma(A_1, \ldots, A_n) = N^{-\#(\sigma)} \sum_{i_1,\ldots,i_n=1}^{N} a^{(1)}_{i_1 i_{\sigma(1)}} a^{(2)}_{i_2 i_{\sigma(2)}} \cdots a^{(n)}_{i_n i_{\sigma(n)}}.$$

Thus, we may write

$$E[\mathrm{tr}(D_N^{q(1)} A_N D_N^{q(2)} \cdots D_N^{q(m)} A_N)] = \sum_{\pi\in\mathcal{P}_2(m)} N^{\#(\gamma\pi)-1-m/2} \mathrm{tr}_{\gamma\pi}[D_N^{q(1)}, \ldots, D_N^{q(m)}]. \tag{4.8}$$

Now, as pointed out in Corollary 1.6, one has for $\pi \in \mathcal{P}_2(m)$ that

$$\lim_{N\to\infty} N^{\#(\gamma\pi)-1-m/2} = \begin{cases} 1, & \text{if } \pi \in NC_2(m) \\ 0, & \text{otherwise} \end{cases},$$

so that we finally get

$$\lim_{N\to\infty} E[\mathrm{tr}(D_N^{q(1)} A_N D_N^{q(2)} \cdots D_N^{q(m)} A_N)] = \sum_{\pi\in NC_2(m)} \varphi_{\gamma\pi}[d^{q(1)}, \ldots, d^{q(m)}]. \tag{4.9}$$

We see that the mixed moments of Gaussian random matrices and deterministic matrices have a definite limit. And moreover, we can recognize this limit as something familiar. Namely, compare (4.9) to the formula (2.22) for a corresponding mixed moment in free variables d and s, in the case where s is semi-circular:

$$\varphi[d^{q(1)} s d^{q(2)} s \cdots d^{q(m)} s] = \sum_{\pi \in NC_2(m)} \varphi_{K^{-1}(\pi)}[d^{q(1)}, \ldots, d^{q(m)}]. \tag{4.10}$$

Both formulas, (4.9) and (4.10), are the same provided $K^{-1}(\pi) = \gamma\pi$ where K is the Kreweras complement. But this is indeed true for all $\pi \in NC_2(m)$; see [137, Ex. 18.25]. Consider, for example, $\pi = \{(1,10),(2,3),(4,7),(5,6),(8,9)\} \in NC_2(10)$. Regard this as the involution $\pi = (1,10)(2,3)(4,7)(5,6)(8,9) \in S_{10}$. Then we have $\gamma\pi = (1)(2,4,8,10)(3)(5,7)(6)(9)$, which corresponds exactly to $K^{-1}(\pi)$.

Thus, we have proved that Gaussian random matrices and deterministic matrices become asymptotically free with respect to the averaged trace. The calculations can of course also be extended to the case of several GUE and deterministic matrices. By estimating the covariance of the appropriate traces (see Remark 5.14), one can strengthen this to almost sure asymptotic freeness. So we have the following theorem of Voiculescu [180, 188].

Theorem 4. *Let $A_N^{(1)}, \ldots, A_N^{(p)}$ be p independent $N \times N$ GUE random matrices and let $D_N^{(1)}, \ldots, D_N^{(q)}$ be q deterministic $N \times N$ matrices such that*

$$D_N^{(1)}, \ldots, D_N^{(q)} \xrightarrow{distr} d_1, \ldots, d_q \quad \text{as } N \to \infty.$$

Then

$$A_N^{(1)}, \ldots, A_N^{(p)}, D_N^{(1)}, \ldots, D_N^{(q)} \xrightarrow{distr} s_1, \ldots, s_p, d_1, \ldots, d_q \qquad \text{as } N \to \infty,$$

where each s_i is semi-circular and $s_1, \ldots, s_p, \{d_1, \ldots, d_q\}$ are free. The convergence above also holds almost surely, so in particular, we have that $A_N^{(1)}, \ldots, A_N^{(p)}$, $\{D_N^{(1)}, \ldots, D_N^{(q)}\}$ are almost surely asymptotically free.

The theorem above can be generalized to the situation where the D_N's are also random matrix ensembles. If we assume that the D_N and the A_N are independent and that the D_N have an almost sure limit distribution, then we get almost sure asymptotic freeness by the deterministic version above just by conditioning onto the D_N's. Hence, we have the following random version for the almost sure setting.

Theorem 5. *Let $A_N^{(1)}, \ldots, A_N^{(p)}$ be p independent $N \times N$ GUE random matrices and let $D_N^{(1)}, \ldots, D_N^{(q)}$ be q random $N \times N$ matrices such that almost surely*

$$D_N^{(1)}(\omega), \ldots, D_N^{(q)}(\omega) \xrightarrow{distr} d_1, \ldots, d_q \quad \text{as } N \to \infty.$$

Furthermore, assume that $A_N^{(1)}, \dots, A_N^{(p)}, \{D_N^{(1)}, \dots, D_N^{(q)}\}$ *are independent. Then we have almost surely as* $N \to \infty$

$$A_N^{(1)}(\omega), \dots, A_N^{(p)}(\omega), D_N^{(1)}(\omega), \dots, D_N^{(q)}(\omega) \xrightarrow{distr} s_1, \dots, s_p, d_1, \dots, d_q$$

where each s_i *is semi-circular and* $s_1, \dots, s_p, \{d_1, \dots, d_q\}$ *are free. So in particular, we have that* $A_N^{(1)}, \dots, A_N^{(p)}, \{D_N^{(1)}, \dots, D_N^{(q)}\}$ *are almost surely asymptotically free.*

For the averaged version, on the other hand, the assumption of an averaged limit distribution for random D_N is not enough to guarantee asymptotic freeness in the averaged sense, as the following example shows.

Example 6. Consider a Gaussian random matrix A_N, and let, for each N, D_N be a random matrix which is independent from A_N and just takes on two values, $P(D_N = I_N) = 1/2$ and $P(D_N = -I_N) = 1/2$, where I_N is the identity matrix. Then for each N, D_N has the averaged eigenvalue distribution $\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$ and thus the same distribution in the limit, but A_N and D_N are clearly not asymptotically free. The problem here is that the fluctuations of D_N are too large; there is no almost sure convergence in that case to $\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$. Of course, we have that I_N is asymptotically free from A_N and that $-I_N$ is asymptotically free from A_N, but this does not imply the asymptotic freeness of D_N from A_N.

Let us also remark that in our algebraic framework, it is not obvious how to deal directly with the assumption of almost sure convergence to the limit distribution. We will actually replace this in the next chapter by the more accessible condition that the variance of the normalized traces is of order $1/N^2$. Note that this is a stronger condition in general than almost sure convergence of the eigenvalue distribution, but this stronger assumption in our theorems will be compensated by the fact that we can then also show this stronger behaviour in the conclusion.

4.3 Asymptotic freeness of Haar distributed unitary random matrices and deterministic matrices

Let $\mathcal{U}(N)$ denote the group of unitary $N \times N$ matrices, i.e. $N \times N$ complex matrices which satisfy $U^*U = UU^* = I_N$. Since $\mathcal{U}(N)$ is a compact group, one can take dU to be Haar measure on $\mathcal{U}(N)$ normalized so that $\int_{\mathcal{U}(N)} dU = 1$, which gives a probability measure on $\mathcal{U}(N)$. A *Haar distributed unitary random matrix* is a matrix U_N chosen at random from $\mathcal{U}(N)$ with respect to Haar measure. There is a useful theoretical and practical way to construct Haar unitaries: take an $N \times N$ (non-self-adjoint!) random matrix whose entries are independent standard complex Gaussians and apply the Gram-Schmidt orthogonalization procedure; the resulting matrix is then a Haar unitary.

Exercise 2. Let $\{Z_{ij}\}_{i,j=1}^N$ be N^2 independent standard complex Gaussian random variables with mean 0 and complex variance 1, i.e. $\mathrm{E}(Z_{ij}\overline{Z_{ij}}) = 1$. Show that if $U = (u_{ij})_{ij}$ is a unitary matrix and $Y_{ij} = \sum_{k=1}^N u_{ik}Z_{kj}$, then $\{Y_{ij}\}_{i,j=1}^N$ are N^2 independent standard complex Gaussian random variables with mean 0 and complex variance 1.

Exercise 3. Let $\Phi : GL_N(\mathbb{C}) \to \mathcal{U}(N)$ be the map which takes an invertible complex matrix A and applies the Gram-Schmidt procedure to the columns of A to obtain a unitary matrix. Show that for any $U \in \mathcal{U}(N)$, we have $\Phi(UA) = U\Phi(A)$.

Exercise 4. Let $\{Z_{ij}\}_{ij}$ be as in Exercise 2 and let Z be the $N \times N$ matrix with entries Z_{ij}. Since $Z \in GL_N(\mathbb{C})$, almost surely, we may let $U = \Phi(Z)$. Show that U is Haar distributed.

What is the $*$-distribution of a Haar unitary random matrix with respect to the state $\varphi = E \circ \mathrm{tr}$? Since $U_N^* U_N = I_N = U_N U_N^*$, the $*$-distribution is determined by the values $\varphi(U_N^m)$ for $m \in \mathbb{Z}$. Note that for any complex number $\lambda \in \mathbb{C}$ with $|\lambda| = 1$, λU_N is again a Haar unitary random matrix. Thus, $\varphi(\lambda^m U_N^m) = \varphi(U_N^m)$ for all $m \in \mathbb{Z}$. This implies that we must have $\varphi(U_N^m) = 0$ for $m \neq 0$. For $m = 0$, we have of course $\varphi(U_N^0) = \varphi(I_N) = 1$.

Definition 7. Let $(\mathcal{A}, \varphi)$ be a $*$-probability space. An element $u \in \mathcal{A}$ is called a *Haar unitary* if:

- u is unitary, i.e. $u^*u = 1_{\mathcal{A}} = uu^*$;
- $\varphi(u^m) = \delta_{0,m}$ for $m \in \mathbb{Z}$.

Thus, a Haar unitary random matrix $U_N \in \mathcal{U}(N)$ is a Haar unitary for each $N \geq 1$ (with respect to $\varphi = E \circ \mathrm{tr}$).

We want to see that asymptotic freeness occurs between Haar unitary random matrices and deterministic matrices, as was the case with GUE random matrices. The crucial element in the Gaussian setting was the Wick formula, which of course does not apply when dealing with Haar unitary random matrices, whose entries are neither independent nor Gaussian. However, we do have a replacement for the Wick formula in this context, which is known as the *Weingarten convolution formula*; see [57, 59].

The Weingarten convolution formula asserts the existence of a sequence of functions $(\mathrm{Wg}_N)_{N=1}^\infty$ with each Wg_N a central function in the group algebra $\mathbb{C}[S_n]$ of the symmetric group S_n, for each $N \geq n$. The function Wg_N has the property that for the entries u_{ij} of a Haar distributed unitary random matrix $U = (u_{ij}) \in \mathcal{U}(N)$ and all index tuples $i, j, i', j' : [n] \to [N]$

$$E[u_{i_1 j_1} \cdots u_{i_n j_n} \overline{u_{i'_1 j'_1}} \cdots \overline{u_{i'_n j'_n}}] = \sum_{\sigma,\tau \in S_n} \prod_{r=1}^n \delta_{i_r i'_{\sigma(r)}} \delta_{j_r j'_{\tau(r)}} \mathrm{Wg}_N(\tau\sigma^{-1}). \tag{4.11}$$

Exercise 5. Let us recall a special factorization of a permutation $\sigma \in S_n$ into a product of transpositions. Let $\sigma_1 = \sigma$ and let $n_1 \leq n$ be the largest integer such that $\sigma_1(n_1) \neq n_1$ and $k_1 = \sigma_1(n_1)$. Let $\sigma_2 = (n_1, k_1)\sigma_1$, the product of the transposition (n_1, k_1) and σ_1. Then $\sigma_2(n_1) = n_1$. Let n_2 be the largest integer such that $\sigma_2(n_2) \neq n_2$ and $k_2 = \sigma_2(n_2)$. In this way, we find $n \geq n_1 > n_2 > \cdots > n_l$ and $k_1, \ldots, k_l$ such that $k_i < n_i$ and such that $(n_l, k_l) \cdots (n_1, k_1)\sigma = e$, the identity of S_n. Then $\sigma = (n_1, k_1) \cdots (n_l, k_l)$, and this representation is unique, subject to the conditions on n_i and k_i. Recall that $\#(\sigma)$ denotes the number of cycles in the cycle decomposition of σ and $|\sigma|$ is the minimal number of factors among all factorizations into a product of transpositions of σ.

Moreover $l = |\sigma| = n - \#(\sigma)$ because $|\sigma_{i-1}| = |\sigma_i| - 1$. Recall the *Jucys-Murphy* elements in $\mathbb{C}[S_n]$; let

$$J_1 = 0, \quad J_2 = (1,2), \quad \ldots \quad J_k = (1,k) + (2,k) + \cdots + (k-1,k).$$

Show that J_k and J_l commute for all k and l.

Exercise 6. Let N be an integer. Using the factorization in Exercise 5, show that

$$(N + J_1) \cdots (N + J_n) = \sum_{\sigma \in S_n} N^{\#(\sigma)} \sigma.$$

Exercise 7. Let $G \in \mathbb{C}[S_n]$ be the function $G(\sigma) = N^{\#(\sigma)}$. Thus, as operators we have $G = (N + J_1) \cdots (N + J_n)$. Show that $\|J_k\| \leq k - 1$ and for $N \geq n$, G is invertible in $\mathbb{C}[S_n]$. Let Wg_N be the inverse of G.
By writing

$$N^n \mathrm{Wg}_N = (1 + N^{-1} J_1)^{-1} \cdots (1 + N^{-1} J_n)^{-1}$$

show that

$$N^n \mathrm{Wg}_N(\sigma) = \mathrm{O}(N^{-|\sigma|}).$$

Thus, one knows the asymptotic decay

$$\mathrm{Wg}_N(\pi) \sim \frac{1}{N^{2n - \#(\pi)}} \text{ as } N \to \infty \tag{4.12}$$

for any $\pi \in S_n$. The convolution formula and the asymptotic estimate allow us to prove the following result of Voiculescu [180, 188].

Theorem 8. *Let $U_N^{(1)}, \ldots, U_N^{(p)}$ be p independent $N \times N$ Haar unitary random matrices, and let $D_N^{(1)}, \ldots, D_N^{(q)}$ be q deterministic $N \times N$ matrices such that*

$$D_N^{(1)}, \ldots, D_N^{(q)} \xrightarrow{distr} d_1, \ldots, d_q \qquad as \qquad N \to \infty.$$

Then, for $N \to \infty$,

$$U_N^{(1)}, U_N^{(1)*}, \dots, U_N^{(p)}, U_N^{(p)*}, D_N^{(1)}, \dots, D_N^{(q)} \xrightarrow{distr} u_1, u_1^*, \dots, u_p, u_p^*, d_1, \dots, d_q,$$

where each u_i is a Haar unitary and $\{u_1, u_1^*\}, \dots, \{u_p, u_p^*\}, \{d_1, \dots, d_q\}$ are free. The above convergence holds also almost surely. In particular, $\{U_N^{(1)}, U_N^{(1)*}\}, \dots, \{U_N^{(p)}, U_N^{(p)*}\}, \{D_N^{(1)}, \dots, D_N^{(q)}\}$ are almost surely asymptotically free.

The proof proceeds in a fashion similar to the Gaussian setting and will not be given here. We refer to [137, Lecture 23].

Note that in general if u is a Haar unitary such that $\{u, u^*\}$ is free from elements $\{a, b\}$, then a and ubu^* are free. In order to prove this, consider

$$\varphi\big(p_1(a)q_1(ubu^*)\cdots p_r(a)q_r(ubu^*)\big)$$

where p_i, q_i are polynomials such that for all $i = 1, \dots, r$

$$\varphi(p_i(a)) = 0 = \varphi(q_i(ubu^*)).$$

Note that by the unitary condition, we have $q_i(ubu^*) = uq_i(b)u^*$. Thus, by the freeness between $\{u, u^*\}$ and b,

$$0 = \varphi(q_i(ubu^*)) = \varphi(uq_i(b)u^*) = \varphi(uu^*)\varphi(q_i(b)) = \varphi(q_i(b)).$$

But then

$$\varphi\big(p_1(a)q_1(ubu^*)\cdots p_r(a)q_r(ubu^*)\big) = \varphi\big(p_1(a)uq_1(b)u^* p_2(a)\cdots p_r(a)uq_r(b)u^*\big)$$

is zero, since $\{u, u^*\}$ is free from $\{a, b\}$ and φ vanishes on all the factors in the latter product.

Thus, our Theorem 8 yields also the following as a corollary.

Theorem 9. *Let A_N and B_N be two sequences of deterministic $N \times N$ matrices with $A_N \xrightarrow{distr} a$ and $B_N \xrightarrow{distr} b$. Let U_N be a sequence of $N \times N$ Haar unitary random matrices. Then $A_N, U_N B_N U_N^* \xrightarrow{distr} a, b$, where a and b are free. This convergence holds also almost surely. So in particular, we have that A_N and $U_N B_N U_N^*$ are almost surely asymptotically free.*

The reader might notice that this theorem is, strictly speaking, not a consequence of Theorem 8, because in order to use the latter we would need the assumption that also mixed moments in A_N and B_N converge to some limit, which we do not assume in Theorem 9. However, the proof of Theorem 8, for the special case where we only need to consider moments in which U_N and U_N^* come alternatingly, reveals that we never encounter a mixed moment in A_N and B_N. The structure of the Weingarten formula ensures that they will never interact. A detailed proof of Theorem 9 can be found in [137, Lecture 23].

Conjugation by a Haar unitary random matrix corresponds to a *random rotation*. Thus, the above theorem says that randomly rotated deterministic matrices become asymptotically free in the limit of large matrix dimension. Another way of saying this is that random matrix ensembles which are *unitarily invariant* (i.e. such that the joint distribution of their entries is not changed by conjugation with any unitary matrix) are asymptotically free from deterministic matrices.

Note that the eigenvalue distribution of B_N is not changed if we consider $U_N B_N U_N^*$ instead. Only the relation between A_N and B_N is brought into a generic form by applying a random rotation between the eigenspaces of A_N and of B_N.

Again one can generalize Theorems 8 and 9 by replacing the deterministic matrices by random matrices, which are independent from the Haar unitary matrices and which have an almost sure limit distribution. As outlined at the end of the last section, we will replace in Chapter 5 the assumption of almost sure convergence by the vanishing of fluctuations $\mathrm{var}[\mathrm{tr}(\cdot), \mathrm{tr}(\cdot)]$ like $1/N^2$. See also our discussions in Chapter 5 around Remark 5.26 and Theorem 5.29.

Note also that Gaussian random matrices are invariant under conjugation by unitary matrices, i.e. if B_N is GUE, then also $U_N B_N U_N^*$ is GUE. Furthermore, the fluctuations of GUE random matrices vanish of the right order, and hence we have almost sure convergence to the semi-circle distribution. Thus, Theorem 9 (in the version where B_N is allowed to be a random matrix ensemble with almost sure limit distribution) contains the asymptotic freeness of Gaussian random matrices and deterministic random matrices (Theorem 4) as a special case.

4.4 Asymptotic freeness between Wigner and deterministic random matrices

Wigner matrices are generalizations of Gaussian random matrices: the entries are, apart from symmetry conditions, independent and identically distributed, but with arbitrary, not necessarily Gaussian, distribution. Whereas Gaussian random matrices are unitarily invariant, this is not true any more for general Wigner matrices; thus, we cannot use the results about Haar unitary random matrices to derive asymptotic freeness results for Wigner matrices. Nevertheless, there are many results in the literature which show that Wigner matrices behave with respect to eigenvalue questions in the same way as Gaussian random matrices. For example, their eigenvalue distribution converges always to a semi-circle. In order to provide a common framework and possible extensions for such investigations, it is important to settle the question of asymptotic freeness for Wigner matrices. We will show that in this respect Wigner matrices also behave like Gaussian random matrices. It turns out that the estimates for the subleading terms are, compared to the Gaussian case, more involved. However, there is actually a nice combinatorial structure behind these estimates, which depends on a general estimate for sums given in terms of graphs. This quite combinatorial approach goes back to work of Yin and Krishnaiah who considered the product of two random matrices, one of them being a covariance matrix (i.e. a Wishart matrix; see Section 4.5). Their moment calculations are special cases of the general asymptotic freeness calculations which we have to address in this section.

Another proof for the asymptotic freeness of Wigner matrices which does not rely on the precise graph sum estimates for the subleading terms can be found in the book of Anderson, Guionnet, Zeitouni [7]. The special case where the deterministic matrices are of a block-diagonal form was already treated by Dykema in [66].

We will extend our calculations from Section 4.2 to Wigner matrices. So let $(A_N)_{N\geq 1}$ now be a sequence of Wigner matrices and $(D_N^{(i)})_{N\geq 1}$ sequences of deterministic matrices whose joint limit distribution exists. We have to look at alternating moments in Wigner matrices and deterministic matrices. Again, we consider just one Wigner matrix, but it is clear that the same arguments work also for a family of independent Wigner matrices, by just decorating the A_N with an additional index. In order to simplify the notation, it is also advantageous to consider the case where the entries of our Wigner matrices are real random variables. So now let us first give a precise definition what we mean by a Wigner matrix.

Notation 10. *Let μ be a probability distribution on $\mathbb{R}$. Let a_{ij} for $i, j \in \mathbb{N}$ with $i \leq j$ be independent identically distributed real random variables with distribution μ. We also put $a_{ij} := a_{ji}$ for $i > j$. Then the corresponding $N \times N$* Wigner random matrix ensemble *is given by the self-adjoint random matrix*

$$A_N = \frac{1}{\sqrt{N}} \left(a_{ij}\right)_{i,j=1}^{N}.$$

Let A_N be now such a Wigner matrix; clearly, in our algebraic frame, we have to assume that all moments of μ exist; furthermore, we have to assume that the mean of μ is zero, and we normalize the variance of μ to be 1.

Remark 11. We want to comment on our assumption that μ has mean zero. In analytic proofs involving Wigner matrices, one usually does not need this assumption. For example, Wigner's semi-circle law holds for Wigner matrices, even if the entries have non-vanishing mean. The general case can, by using properties of weak convergence, be reduced to the case of vanishing mean. However, in our algebraic frame, we cannot achieve this reduction. The reason for this discrepancy is that our notion of convergence in distribution is actually stronger than weak convergence in situations where mass might escape to infinity. For example, consider a deterministic diagonal matrix D_N, with $a_{11} = N$, and all other entries zero. Then $\mu_{D_N} = (1 - 1/N)\delta_0 + 1/N\delta_N$; thus, μ_{D_N} converges weakly to δ_0, for $N \to \infty$. However, the second and higher moments of D_N with respect to tr do not converge; thus, D_N does not converge in distribution.

Another simplifying assumption we have made is that the distribution of the diagonal entries is the same as that of the off-diagonal entries. With a little more work, the method given here can be made to work without this assumption.

We examine now an averaged alternating moment in our deterministic matrices $D_N^{(k)} = (d_{ij}^{(k)})$ and the Wigner matrix $A_N = \frac{1}{\sqrt{N}}(a_{ij})$. We have

$$\begin{aligned} E\left[\operatorname{tr}\bigl(D_N^{(1)} A_N \cdots D_N^{(m)} A_N\bigr)\right] &= \frac{1}{N^{m/2+1}} \sum_{i_1,\dots,i_{2m}=1}^{N} E\left[d_{i_1 i_2}^{(1)} a_{i_2 i_3} \cdots d_{i_{2m-1} i_{2m}}^{(m)} a_{i_{2m} i_1}\right] \\ &= \frac{1}{N^{m/2+1}} \sum_{i_1,\dots,i_{2m}=1}^{N} E\left[a_{i_2 i_3} \cdots a_{i_{2m} i_1}\right] d_{i_1 i_2}^{(1)} \cdots d_{i_{2m-1} i_{2m}}^{(m)} \\ &= \frac{1}{N^{m/2+1}} \sum_{i_1,\dots,i_{2m}=1}^{N} \sum_{\sigma \in \mathcal{P}(m)} k_\sigma(a_{i_2 i_3}, \dots, a_{i_{2m} i_1}) d_{i_1 i_2}^{(1)} \cdots d_{i_{2m-1} i_{2m}}^{(m)}. \end{aligned}$$

In the last step, we have replaced the Wick formula for Gaussian random variables by the general expansion of moments in terms of classical cumulants. Now we use the independence of the entries of A_N. A cumulant in the a_{ij} is only different from zero if all its arguments are the same; of course, we have to remember that $a_{ij} = a_{ji}$. (Not having to bother about the complex conjugate here is the advantage of looking at real Wigner matrices.) Thus, in order that $k_\sigma[a_{i_2 i_3}, \dots, a_{i_{2m} i_1}]$ is different from zero, we must have: if k and l are in the same block of σ, then we must have $\{i_{2k}, i_{2k+1}\} = \{i_{2l}, i_{2l+1}\}$. Note that now we do not prescribe whether i_{2k} has to agree with i_{2l} or with i_{2l+1}. In order to deal with partitions of the indices $i_1, \dots, i_{2m}$ instead of partitions of the pairs $(i_2, i_3), (i_4, i_5) \dots, (i_{2m}, i_1)$, we say that a partition $\pi \in \mathcal{P}(2m)$ is a *lift* of a partition $\sigma \in \mathcal{P}(m)$ if we have for all $k, l = 1, \dots, m$ with $k \neq l$ that

$$k \sim_\sigma l \Leftrightarrow \Bigl\{[2k \sim_\pi 2l \text{ and } 2k+1 \sim_\pi 2l+1] \text{ or } [2k \sim_\pi 2l+1 \text{ and } 2k+1 \sim_\pi 2l]\Bigr\}.$$

Here we are using the notation $k \sim_\sigma l$ to mean that k and l are in the same block of σ. Then the condition that $k_\sigma(a_{i_2 i_3}, \dots, a_{i_{2m} i_1})$ is different from zero can also be paraphrased as $\ker i \geq \pi$, for some lift π of σ. Note that the value of $k_\sigma(a_{i_2 i_3}, \dots, a_{i_{2m} i_1})$ depends only on $\ker(i)$ because we have assumed that the diagonal and off-diagonal elements have the same distribution. Let us denote this common value by $k_{\ker(i)}$. Thus, we can rewrite the equation above as

$$E\left[\operatorname{tr}\bigl(D_N^{(1)} A_N \cdots D_N^{(m)} A_N\bigr)\right] = \frac{1}{N^{m/2+1}} \sum_{\sigma \in \mathcal{P}(m)} \sum_{\substack{i:[2m]\to[N] \\ \ker i \geq \pi \text{ for some lift } \pi \text{ of } \sigma}} k_{\ker(i)}\, d_{i_1 i_2}^{(1)} \cdots d_{i_{2m-1} i_{2m}}^{(m)}. \tag{4.13}$$

Note that in general, there is not a unique lift of a given σ. For example, for the one block partition $\sigma = \{(1,2,3)\} \in \mathcal{P}(3)$, we have the following lifts in $\mathcal{P}(6)$:

$$\{(1,3,5),(2,4,6)\}, \qquad \{(1,3,4),(2,5,6)\}, \qquad \{(1,2,4),(3,5,6)\},$$
$$\{(1,2,5),(3,4,6)\}, \qquad \{(1,2,3,4,5,6)\}.$$

If σ consists of several blocks, then one can make the corresponding choice for each block of σ. If σ is a pairing, there is a special lift π of σ which we call the *standard lift* of σ; if (r,s) is a block of σ, then π will have the blocks $(2r+1, 2s)$ and $(2r, 2s+1)$.

If we want to rewrite the sum over i in (4.13) in terms of sums of the form

$$\sum_{\substack{i:[2m]\to[N] \\ \ker i \geq \pi}} d^{(1)}_{i_1 i_2} \cdots d^{(m)}_{i_{2m-1} i_{2m}} \tag{4.14}$$

for fixed lifts π, then we have to notice that in general a multi-index i will show up with different π's; indeed, the lifts of a given σ are partially ordered by inclusion and form a poset; thus, we can rewrite the sum over i with $\ker i \geq \pi$ for some lift π of σ in terms of sums over fixed lifts, with some well-defined coefficients (given by the Möbius function of this poset – see Exercise 8). However, the precise form of these coefficients is not needed since we will show that at most one of the corresponding sums has the right asymptotic order (namely, $N^{m/2+1}$), so all the other terms will play no role asymptotically. So our main goal will now be to examine the sum (4.14) and show that for all $\pi \in \mathcal{P}(2m)$ which are lifts of σ, a term of the form (4.14) grows in N with order at most $m/2+1$, and furthermore, this maximal order is achieved only in the case in which σ is a non-crossing pairing and π is the standard lift of σ. After identifying these terms, we must relate them to Equation (4.9); this is achieved in Exercise 9.

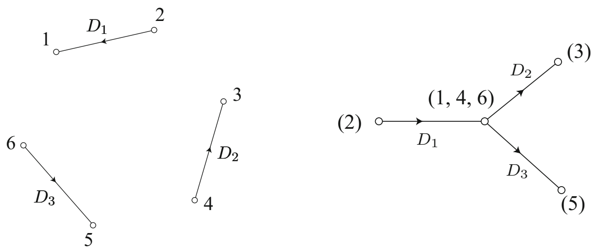
Exercise 8. Let σ be a partition of $[m]$ and $M = \{\pi \in \mathcal{P}(2m) \mid \pi \text{ is a lift of } \sigma\}$. For a subset L of M, let $\pi_L = \sup_{\pi\in L} \pi$; here sup denotes the join in the lattice of all partitions. Use the principle of inclusion-exclusion to show that

$$\sum_{\substack{i:[2m]\to[N] \\ \ker i \geq \pi \text{ for some } \pi\in M}} d^{(1)}_{i_1 i_2} \cdots d^{(m)}_{i_{2m-1} i_{2m}} = \sum_{L\subset M} (-1)^{|L|-1} \sum_{\substack{i:[2m]\to[N] \\ \ker i \geq \pi_L}} d^{(1)}_{i_1 i_2} \cdots d^{(m)}_{i_{2m-1} i_{2m}}.$$

Exercise 9. Let σ be a pairing of $[m]$ and π be the standard lift of σ. Then

$$\sum_{\ker(i)\geq\pi} d^{(1)}_{i_1 i_2} \cdots d^{(m)}_{i_{2m-1} i_{2m}} = \mathrm{Tr}_{\gamma_m \sigma}(D^{(1)}_N, \ldots, D^{(m)}_N).$$

Fig. 4.6 On the left, we have Γ_6. We let π be the partition of [6] with blocks $\{(1,4,6), (2), (3), (5)\}$. The graph on the right is G_π. We have $S_\pi(N) = \sum_{i,j,k,l} d_{ij}^{(1)} d_{jk}^{(2)} d_{jl}^{(3)}$ and $\mathfrak{r}(G_\pi) = 3/2$



Let us first note that, because of our assumption that the entries of the Wigner matrices have vanishing mean, first order cumulants are zero and thus only those σ which have no singletons will contribute to (4.13). This implies the same property for the lifts, and in (4.14), we can restrict ourselves to considering π without singletons.

It turns out that it is convenient to associate to π a graph G_π. Let us start with the directed graph Γ_{2m} with $2m$ vertices labelled $1, 2, \ldots, 2m$ and directed edges $(1,2), (3,4), \ldots, (2m-1, 2m)$; $(2i-1, 2i)$ starts at $2i$ and goes to $2i-1$. Given a $\pi \in \mathcal{P}(2m)$, we obtain a directed graph G_π by identifying the vertices which belong to the same block of π. We will not identify the edges (actually, the direction of two edges between identified vertices might even not be the same) so that G_π will in general have multiple edges, as well as loops. The sum (4.14) can then be rewritten in terms of the graph $G = G_\pi$ as

$$S_G(N) := \sum_{i:V(G)\to[N]} \prod_{e\in E(G)} d^{(e)}_{i_{t(e)}, i_{s(e)}}, \tag{4.15}$$

where we sum over all functions $i : V(G) \to [N]$, and for each such function we take the product of $d^{(e)}_{i_{t(e)}, i_{s(e)}}$ as e runs over all the edges of the graph and $s(e)$ and $t(e)$ denote, respectively, the source and terminus of the edge e. Note that we keep all edges under the identification according to π; thus, the m matrices $D^{(1)}, \ldots, D^{(m)}$ in (4.14) show up in (4.15) as the various D_e for the m edges of G_π. See Fig. 4.6.

What we have to understand about such graph sums is their asymptotic behaviour as $N \to \infty$. This problem has a nice answer for arbitrary graphs, namely, one can estimate such graph sums (4.15) in terms of the norms of the matrices corresponding to the edges and properties of the graph G. The relevant feature of the graph is the structure of its two-edge connected components.

Definition 12. A *cutting edge* of a connected graph is an edge whose removal would disconnect the graph. A connected graph is *two-edge connected* if it does not contain a cutting edge, i.e. if it cannot be disconnected by the removal of an edge. A *two-edge connected component* of a graph is a two-edge connected subgraph which is not properly contained is a larger two-edge connected subgraph.

A *forest* is a graph without cycles. A *tree* is a connected component of a forest, i.e. a connected graph without cycles. A tree is *trivial* if it consists of only one vertex. A *leaf* of a non-trivial tree is a vertex which meets only one edge. The sole vertex of a trivial tree will also be called a *trivial leaf*.

It is clear that if one shrinks each two-edge connected component of a graph to a vertex and removes the loops, then one does not have any more cycles; thus, one is left with a forest.

Notation 13. *For a graph G, we denote by $\mathfrak{F}(G)$ its* forest of two-edge connected components*; the vertices of $\mathfrak{F}(G)$ consist of the two-edge connected components of G, and two distinct vertices of $\mathfrak{F}(G)$ are connected by an edge if there is a cutting edge between vertices from the two corresponding two-edge connected components in G.*

We can now state the main theorem on estimates for graph sums. The special case for two-edge connected graphs goes back to the work of Yin and Krishnaiah [206]; see also the book of Bai and Silverstein [15]. The general case, which is stronger than the corresponding statement in [15, 206], is proved in [129].

Theorem 14. *Let G be a directed graph, possibly with multiple edges and loops. Let for each edge e of G be given an $N \times N$ matrix $D_e = (d_{ij}^{(e)})_{i,j=1}^N$. Then the associated graph sum (4.15) satisfies*

$$|S_G(N)| \leq N^{\mathfrak{r}(G)} \cdot \prod_{e \in E(G)} \|D_e\|, \tag{4.16}$$

where $\mathfrak{r}(G)$ is determined as follows from the structure of the graph G. Let $\mathfrak{F}(G)$ be the forest of two-edge connected components of G. Then

$$\mathfrak{r}(G) = \sum_{\mathfrak{l} \text{ leaf of } \mathfrak{F}(G)} \mathfrak{r}(\mathfrak{l}),$$

where

$$\mathfrak{r}(\mathfrak{l}) := \begin{cases} 1, & \text{if } \mathfrak{l} \text{ is a trivial leaf} \\ \frac{1}{2}, & \text{if } \mathfrak{l} \text{ is a leaf of a non-trivial tree} \end{cases}.$$

Note that each tree of the forest $\mathfrak{F}(G)$ makes at least a contribution of 1 in $\mathfrak{r}(G)$, because a non-trivial tree has at least two leaves. One can also make the description above more uniform by having a factor $1/2$ for each leaf, but then counting a trivial leaf as two actual leaves. Note also that the direction of the edges plays no role for the estimate above. The direction of an edge is only important in order to define the contribution of an edge to the graph sum. One direction corresponds to the matrix D_e, and the other direction corresponds to the transpose D_e^t. Since the norm of a matrix is the same as the norm of its transpose, the estimate is the same for all graph sums which correspond to the same undirected graph.

Let us now apply Theorem 14 to G_π. We have to show that $\mathfrak{r}(G_\pi) \leq m/2+1$ for our graphs G_π, $\pi \in \mathcal{P}(2m)$. Of course, for general $\pi \in \mathcal{P}(2m)$, this does not need to be true. For example, if $\pi = \{(1,2),(3,4),\ldots,(2m-1,2m)\}$, then G_π consists of m isolated points and thus $\mathfrak{r}(G_\pi) = m$. Clearly, we have to take into account that we can restrict in (4.13) to lifts of a σ without singletons.

Definition 15. Let $G = (V, E)$ be a graph and $w_1, w_2 \in V$. Let us consider the graph G' obtained by *merging* the two vertices w_1 and w_2 into a single vertex w. This means that the vertices V' of G' are $(V \setminus \{w_1, w_2\}) \cup \{w\}$. Also each edge of G becomes an edge of G', except that if the edge started (or ended) at w_1 or w_2, then the corresponding edge of G' starts (or ends) at w.

Lemma 16. *Suppose π_1 and π_2 are partitions of $[2m]$ and $\pi_1 \leq \pi_2$. Then $\mathfrak{r}(G_{\pi_2}) \leq \mathfrak{r}(G_{\pi_1})$.*

Proof: We only have to consider the case where π_2 is obtained from π_1 by joining two blocks w_1 and w_2 of π_1 and then use induction.

We have to consider three cases. Let C_1 and C_2 be the two-edge connected components of G_{π_1} containing w_1 and w_2, respectively. Recall that $\mathfrak{r}(G_{\pi_1})$ is the sum of the contributions of each connected component and the contribution of a connected component is either 1 or one half the number of leaves in the corresponding tree of $\mathfrak{F}(G_{\pi_1})$, whichever is larger.

Case 1. Suppose the connected component of G_{π_1} containing w_1 is two-edge connected, i.e. C_1 becomes the only leaf of a trivial tree in $\mathfrak{F}(G_{\pi_1})$. Then the contribution of this component to $\mathfrak{r}(G_{\pi_1})$ is 1. If w_2 is in C_1, then merging w_1 and w_2 has no effect on $\mathfrak{r}(G_{\pi_1})$ and thus $\mathfrak{r}(G_{\pi_1}) = \mathfrak{r}(G_{\pi_2})$. If w_2 is not in C_1, then C_1 gets joined to some other connected component of G_{π_1}, which will leave the contribution of this other component unchanged. In this latter case, we shall have $\mathfrak{r}(G_{\pi_2}) = \mathfrak{r}(G_{\pi_1}) - 1$.

For the rest of the proof, we shall assume that neither w_1 nor w_2 lies in a connected component of G_{π_1} which has only one two-edge connected component.

Case 2. Suppose w_1 and w_2 lie in different connected components of G_{π_1}. When w_1 and w_2 are merged, the corresponding two-edge connected components are joined. If either of these corresponded to a leaf in $\mathfrak{F}(G_{\pi_1})$, then the number of leaves would be reduced by 1 or 2 (depending on whether both two-edge components were leaves in $\mathfrak{F}(G_{\pi_1})$). Hence, $\mathfrak{r}(G_{\pi_2})$ is either $\mathfrak{r}(G_{\pi_1}) - 1/2$ or $\mathfrak{r}(G_{\pi_1}) - 1$.

Case 3. Suppose that both w_1 and w_2 are in the same connected component of G_{π_1}. Then the two-edge connected components C_1 and C_2 become vertices of a tree T in $\mathfrak{F}(G_{\pi_1})$ (see Fig. 4.7). When we merge w_1 and w_2, we form a two-edge connected component C of G_{π_2}, which consists of all the two-edge connected components corresponding to the vertices of T along the unique path from C_1 to C_2. On the level of T, this corresponds to collapsing all the edges between C_1 and C_2 into a single vertex. This may reduce the number of leaves by 0, 1, or 2. If there were only two leaves, we might end up with a single vertex, but the contribution to $\mathfrak{r}(G_{\pi_1})$ would still not increase. Thus, $\mathfrak{r}(G_{\pi_1})$ can only decrease.

□

Definition 17. Let G be a directed graph and let v be a vertex of G. Suppose that v has one incoming edge e_1 and one outgoing edge e_2. Let G' be the graph obtained by removing e_1, e_2, and v and replacing these with an edge e from $s(e_1)$ to $t(e_2)$. We say that G' is the graph obtained from G by *removing the vertex* v. See Fig. 4.8.

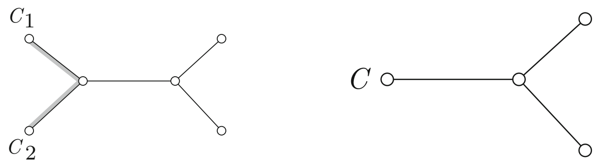


Fig. 4.7 Suppose w_1 and w_2 are in the same connected component of G_{π_1}, but in different, say C_1 and C_2, two-edge connected components of G_{π_1}, we collapse the edge (shown here shaded) joining C_1 to C_2 in $\mathfrak{F}(G_{\pi_1})$ (See Case 3 in the proof of Lemma 16)

Fig. 4.8 If we remove the vertex v from a graph, we replace the edges e_1 and e_2 by the edge e (See Definition 17)

We say that the degree of a vertex is the number of edges to which it is incident, using the convention that a loop contributes 2. The total degree of a subgraph is the sum of the degrees of all its vertices.

Using the usual order on partitions of $[2m]$, we say that a partition π is a minimal lift of σ if it is not larger than some other lift of σ.

Lemma 18. *Let σ be a partition of $[m]$ without singletons and $\pi \in \mathcal{P}(2m)$ be a minimal lift of σ. Suppose that G_π contains a two-edge connected component of total degree strictly less than 3 and which becomes a leaf in $\mathfrak{F}(G_\pi)$. Then*

(i) $(k-1, k)$ *is a block of σ; and*
(ii) $(2k-2, 2k+1)$ *and* $(2k-1, 2k)$ *are blocks of π.*

Let σ' be the partition obtained by deleting the block $(k-1, k)$ from σ and π' the partition obtained by deleting $(2k-2, 2k+1)$ and $(2k-1, 2k)$ from π. Then π' is a minimal lift of σ', and the graph $G_{\pi'}$ is obtained from G_π by:

(a) *deleting the connected component $(2k-1, 2k)$ and;*
(b) *deleting the vertex obtained from $(2k-2, 2k+1)$;*
(c) $\mathfrak{r}(G_\pi) = \mathfrak{r}(G_{\pi'}) + 1$.

Proof: Since σ has no singletons, each block of σ contains at least two elements, and thus each block of the lift π contains at least two points. Thus, every vertex of G_π has degree at least 2. So a two-edge connected component with total degree less than 3 must consist of a single vertex. Moreover, if this vertex has distinct incoming and outgoing edges, then this two-edge connected component cannot become a leaf in $\mathfrak{F}(G_\pi)$. Thus, G_π has a two-edge connected component C which consists of a vertex with a loop. Moreover, C will also be a connected component. Since an edge always goes from $2k-1$ to $2k$, π must have a block consisting of the two elements $2k-1$ and $2k$. Since π is a lift of σ, σ must have the block $(k-1, k)$. Since π is a minimal lift of σ, π has the two blocks $(2k-2, 2k+1)$, $(2k-1, 2k)$. This proves (i) and (ii).

Now π' is a minimal lift of σ' because π was minimal on all the other blocks of σ. Also the block $(2k-2, 2k+1)$ corresponds to a vertex of G_π with one incoming edge and one outgoing edge. Thus, by removing this block from π, we remove a vertex from G_π, as described in Definition 17. Hence, $G_{\pi'}$ is obtained from G_π by removing the connected component C and the vertex $(2k-2, 2k+1)$.

Finally, the contribution of C to $\mathfrak{r}(G_\pi)$ is 1. If the connected component, C', of G_π containing the vertex $(2k-2, 2k+1)$ has only one other vertex, which would have to be $(2k-3, 2k+2)$, the contribution of this component to $\mathfrak{r}(G_\pi)$ will be 1, and $G_{\pi'}$ will have as a connected component this vertex $(2k-3, 2k+2)$ and a loop whose contribution to $\mathfrak{r}(G_{\pi'})$ will still be 1. On the other hand, if C' has more than one other vertex, then the number of leaves will not be diminished when the vertex $(2k-1, 2k+1)$ is removed, and thus also in this case, the contribution of C' to $\mathfrak{r}(G_\pi)$ is unchanged. Hence, in both cases $\mathfrak{r}(G_\pi) = \mathfrak{r}(G_{\pi'}) + 1$. □

Lemma 19. *Consider $\sigma \in \mathcal{P}(m)$ without singletons and let $\pi \in \mathcal{P}(2m)$ be a lift of σ. Then we have for the corresponding graph G_π that*

$$\mathfrak{r}(G_\pi) \leq \frac{m}{2} + 1, \tag{4.17}$$

and we have equality if and only if σ is a non-crossing pairing and π the corresponding standard lift

$$k \sim_\sigma l \Leftrightarrow \{2k \sim_\pi 2l + 1 \text{ and } 2k + 1 \sim_\pi 2l\}.$$

Proof: By Lemma 16, we may suppose that π is a minimal lift of σ. Let the connected components of G_π be $C_1, \ldots, C_p$. Let the number of edges in C_i be m_i, and the number of leaves in the tree of $\mathfrak{F}(G_\pi)$ corresponding to C_i be l_i. The contribution of C_i to $\mathfrak{r}(G_\pi)$ is $\mathfrak{r}_i = \max\{1, l_i/2\}$.

Suppose σ has no blocks of the form $(k-1, k)$. Then by Lemma 18 each two-edge connected component of G_π which becomes a leaf in $\mathfrak{F}(G_\pi)$ must have total degree at least 3. Thus, $m_i \geq 2$ for each i. Moreover, the contribution of each leaf to the total degree must be at least 3. Thus, $3l_i \leq 2m_i$. If $l_i \geq 2$, then $\mathfrak{r}_i = l_i/2 \leq m_i/3$. If $l_i = 1$, then, as $m_i \geq 2$, we have $\mathfrak{r}_i = 1 \leq m_i/2$. So in either case, $\mathfrak{r}_i \leq m_i/2$. Summing over all components, we have $\mathfrak{r}(G_\pi) \leq m/2$.

If σ does contain a block of the form $(k-1, k)$ and π blocks $(2k-2, 2k+1)$, $(2k-1, 2k)$, then we may repeatedly remove these blocks from σ and π until we reach σ' and π' such that either (a) σ' contains no blocks which are a pair of adjacent elements or (b) $\sigma' = \{(1, 2)\}$ (after renumbering) and π' is a minimal lift of σ'. In either case by Lemma 18, $\mathfrak{r}(G_\pi) = \mathfrak{r}(G_{\pi'}) + q$ where q is the number of times we have removed a pair of adjacent elements of σ.

In case (a), we have by the earlier part of the proof that $\mathfrak{r}(G_{\pi'}) \leq m'/2$. Thus, $\mathfrak{r}(G_\pi) = \mathfrak{r}(G_{\pi'}) + q \leq m'/2 + q = m/2$.

In case (b) we have that $\sigma' = \{(1, 2)\}$ and either $\pi = \{(1, 2), (3, 4)\}$ (π is standard) or $\pi = \{(1, 3), (2, 4)\}$ (π is not standard). In the first case, see Fig. 4.9,

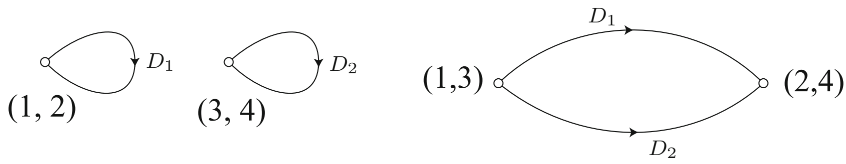


Fig. 4.9 If $\sigma = \{(1,2)\}$ there are two possible minimal lifts: $\pi_1 = \{(1,2),(3,4)\}$ and $\pi_2 = \{(1,3),(2,4)\}$. We show G_{π_1} on the left and G_{π_2} on the right. The graph sum for π_1 is $\mathrm{Tr}(D_1)\mathrm{Tr}(D_2)$ and the graph sum for π_2 is $\mathrm{Tr}(D_1 D_2^t)$ (See the conclusion of the proof of Lemma 19)

$G_{\pi'}$ has two vertices, each with a loop and so $\mathfrak{r}(G_{\pi'}) = 2 = m'/2 + 1$, and hence $\mathfrak{r}(G_\pi) = q + m'/2 + 1 = m/2 + 1$. In the second case, $G_{\pi'}$ is two-edge connected and so $\mathfrak{r}(G_{\pi'}) = 1 = m'/2$, and hence $\mathfrak{r}(G_\pi) = q + m'/2 = m/2$. So we can only have $\mathfrak{r}(G_\pi) = m/2 + 1$ when σ is a non-crossing pairing and π is standard; in all other cases, we have $\mathfrak{r}(G_\pi) \le m/2$. □

Equipped with this lemma, the investigation of the asymptotic freeness of Wigner matrices and deterministic matrices is now quite straightforward. Lemma 19 shows that the sum (4.14) has at most the order $N^{m/2+1}$ and that the maximal order is achieved exactly for σ which are non-crossing pairings and for π which are the corresponding standard lifts. But for those we get in (4.13) the same contribution as for Gaussian random matrices. The other terms in (4.13) will vanish, as long as we have uniform bounds on the norms of the deterministic matrices. Thus, the result for Wigner matrices is the same as for Gaussian matrices, provided we assume a uniform bound on the norm of the deterministic matrices.

Moreover, the forgoing arguments can be extended to several independent Wigner matrices. Thus, we have proved the following theorem.

Theorem 20. *Let $\mu_1, \dots, \mu_p$ be probability measures on $\mathbb{R}$, for which all moments exist and for which the means vanish. Let $A_N^{(1)}, \dots, A_N^{(p)}$ be p independent $N \times N$ Wigner random matrices with entry distributions $\mu_1, \dots, \mu_p$, respectively, and let $D_N^{(1)}, \dots, D_N^{(q)}$ be q deterministic $N \times N$ matrices such that for $N \to \infty$*

$$D_N^{(1)}, \dots, D_N^{(q)} \xrightarrow{distr} d_1, \dots, d_q$$

and such that

$$\sup_{\substack{N \in \mathbb{N} \\ r=1,\dots,q}} \|D_N^{(r)}\| < \infty.$$

Then, as $N \to \infty$,

$$A_N^{(1)}, \dots, A_N^{(p)}, D_N^{(1)}, \dots, D_N^{(q)} \xrightarrow{distr} s_1, \dots, s_p, d_1, \dots, d_q,$$

where each s_i is semi-circular and $s_1, \ldots, s_p, \{d_1, \ldots, d_q\}$ are free. In particular, we have that $A_N^{(1)}, \ldots, A_N^{(p)}, \{D_N^{(1)}, \ldots, D_N^{(q)}\}$ are asymptotically free.

By estimating the variance of the traces, one can show that one also has almost sure convergence in the above theorem; also, one can extend those statements to random matrices $D_N^{(k)}$ which are independent from the Wigner matrices, provided one assumes the almost sure version of a limit distribution and of the norm boundedness condition. We leave the details to the reader.

Exercise 10. Show that under the same assumptions as in Theorem 20, one can bound the variance of the trace of a word in Wigner and deterministic matrices as

$$\operatorname{var}\left[\operatorname{tr}\left(D_N^{(1)} A_N \cdots D_N^{(m)} A_N\right)\right] \leq \frac{C}{N^2},$$

where C is a constant, depending on the word.

Show that this implies that Wigner matrices and deterministic matrices are almost surely asymptotically free under the assumptions of Theorem 20.

Exercise 11. State (and possibly prove) the version of Theorem 20, where the $D_N^{(1)}, \ldots, D_N^{(q)}$ are allowed to be random matrices.

4.5 Examples of random matrix calculations

In the following, we want to look at some examples which show how the machinery of free probability can be used to calculate asymptotic eigenvalue distributions of random matrices.

4.5.1 Wishart matrices and the Marchenko-Pastur distribution

Besides the Gaussian random matrices, the most important random matrix ensemble are the *Wishart random matrices* [203]. They are of the form $A = \frac{1}{N} XX^*$, where X is an $N \times M$ random matrix with independent Gaussian entries. There are two forms: a complex case when the entries x_{ij} are standard complex Gaussian random variables with mean 0 and $\mathrm{E}(|x_{ij}|^2) = 1$ and a real case where the entries are real-valued Gaussian random variables with mean 0 and variance 1. Again, one has an almost sure convergence to a limiting eigenvalue distribution (which is the same in both cases), if one sends N and M to infinity in such a way that the ratio M/N is kept fixed. Figure 4.10 below shows the eigenvalue histograms with $M = 2N$, for $N = 100$ and $N = 2000$. For $N = 100$, we have averaged over 3000 realizations.

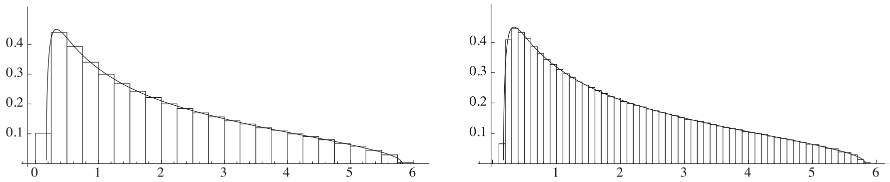


Fig. 4.10 On the left, we have the eigenvalue distribution of a Wishart random matrix with $N = 100$ and $M = 200$ averaged over 3000 instances, and on the right we have one instance with $N = 2000$ and $M = 4000$. The solid line is the graph of the density of the limiting distribution

By similar calculations as for the Gaussian random matrices, one can show that in the limit $N, M \to \infty$ such that the ratio $M/N \to c$, for some $0 < c < \infty$, the asymptotic averaged eigenvalue distribution is given by

$$\lim_{\substack{N,M\to\infty \\ M/N\to c}} \mathrm{E}[\mathrm{tr}(A^k)] = \sum_{\pi\in NC(k)} c^{\#(\pi)}. \tag{4.18}$$

Exercise 12. Show that for $A = \frac{1}{N}XX^*$, a Wishart matrix as above, we have

$$\mathrm{E}(\mathrm{Tr}(A^k)) = \frac{1}{N^k} \sum_{i_1,\ldots,i_k=1}^{N} \sum_{i_{-1},\ldots,i_{-k}=1}^{M} \mathrm{E}(x_{i_1 i_{-1}} \overline{x_{i_2 i_{-1}}} \cdots x_{i_k i_{-k}} \overline{x_{i_1 i_{-k}}}).$$

Then use Exercise 1.7 to show that, in the case of standard complex Gaussian entries for X, we have the "genus expansion"

$$\mathrm{E}(\mathrm{tr}(A^k)) = \sum_{\sigma\in S_k} N^{\#(\sigma)+\#(\gamma_k\sigma^{-1})-(k+1)} \left(\frac{M}{N}\right)^{\#(\sigma)}.$$

Then use Proposition 1.5 to prove (4.18).

This means that all free cumulants of the limiting distribution are equal to c. This qualifies the limiting distribution to be called a *free Poisson distribution of rate* c. Since this limiting distribution of Wishart matrices was first calculated by Marchenko and Pastur [123], it is in the random matrix literature usually called the *Marchenko-Pastur distribution*. See Definition 2.11, Exercises 2.10, 2.11, and Remark 3.11 and the subsequent exercises.

Exercise 13. We have chosen the normalization for Wishart matrices that simplifies the free cumulants. The standard normalization is $\frac{1}{M}XX^*$. If we let $A' = \frac{1}{M}XX^*$, then $A = \frac{M}{N}A'$ so in the limit we have scaled the distribution by c. Using Exercise 2.12, show that the limiting eigenvalue distribution of A' is ρ_y where $y = 1/c$ (using the notation of Remark 2.12).

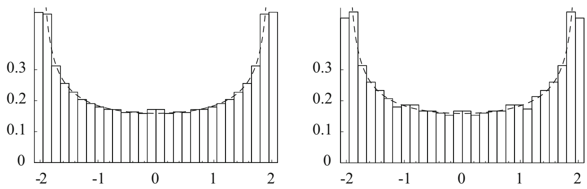


Fig. 4.11 The eigenvalue distribution of $A + UAU^*$. In the left graph, we have 1000 realizations with $N = 100$, and in the right, one realization with $N = 1000$

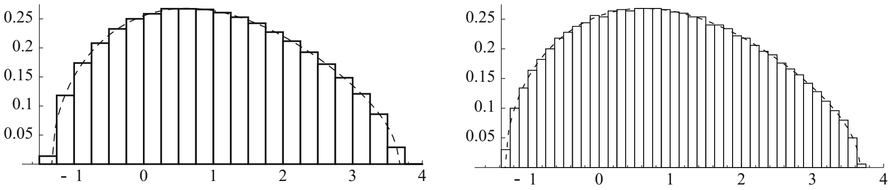


Fig. 4.12 On the left, we display the averaged eigenvalue distribution for 3000 realizations of the sum of a GUE and a complex Wishart random matrix with $M = 200$ and $N = 100$. On the right, we display the eigenvalue distribution of a single realization of the sum of a GUE and a complex Wishart random matrix with $M = 8000$ and $N = 4000$

4.5.2 Sum of random matrices

Let us now consider the sum of random matrices. If the two matrices are asymptotically free, then we can apply the R-transform machinery for calculating the asymptotic distribution of their sum. Namely, for each of the two matrices, we calculate the Cauchy transform of their asymptotic eigenvalue distribution and from this their R-transform. Then the sum of the R-transforms gives us the R-transform of the sum of the matrices, and from there we can go back to the Cauchy transform and, via Stieltjes inversion theorem, to the density of the sum.

Example 21. As an example, consider $A + UAU^*$, where U is a Haar unitary random matrix and A is a diagonal matrix with $N/2$ eigenvalues -1 and $N/2$ eigenvalues 1. (See Fig. 4.11.)

Thus, by Theorem 9, the asymptotic eigenvalue distribution of the sum is the same as the distribution of the sum of two free Bernoulli distributions. The latter can be easily calculated as the arc-sine distribution. See [137, Example 12.8].

Example 22. Consider now independent GUE and Wishart matrices. They are asymptotically free; thus, the asymptotic eigenvalue distribution of their sum is given by the free convolution of a semi-circle and a Marchenko-Pastur distribution.

Figure 4.12 shows the agreement (for $c = 2$) between numerical simulations and the predicted distribution using the R-transform. The first is averaged over 3000 realizations with $N = 100$, and the second is one realization for $N = 4000$.

4.5.3 Product of random matrices

One can also rewrite the combinatorial description (2.23) of the product of free variables into an analytic form. The following theorem gives this version in terms of Voiculescu's *S-transform* [178]. For more details and a proof of that theorem, we refer to [137, Lecture 18].

Theorem 23. *Put $M_a(z) := \sum_{m=0}^{\infty} \varphi(a^m) z^m$ and define the S-transform of a by*

$$S_a(z) := \frac{1+z}{z} M_a^{\langle -1 \rangle}(z),$$

where $M^{\langle -1 \rangle}$ denotes the inverse under composition of M. Then if a and b are free, we have $S_{ab}(z) = S_a(z) \cdot S_b(z)$.

Again, this allows one to do analytic calculations for the asymptotic eigenvalue distribution of a product of asymptotically free random matrices. One should note in this context that the product of two self-adjoint matrices is in general not self-adjoint; thus, it is not clear why all its eigenvalues should be real. (If they are not real, then the S-transform does not contain enough information to recover the eigenvalues.) However, if one makes the restriction that at least one of the matrices has positive spectrum, then, because the eigenvalues of AB are the same as those of the self-adjoint matrix $B^{1/2}AB^{1/2}$, one can be sure that the eigenvalues of AB are real as well, and one can use the S-transform to recover them. One should also note that a priori the S-transform of a is only defined if $\varphi(a) \neq 0$. However, by allowing formal power series in $\sqrt{z}$, one can also extend the definition of the S-transform to the case where $\varphi(a) = 0$, $\varphi(a^2) > 0$. For more on this, and the corresponding version of Theorem 23 in that case, see [146].

Example 24. Consider two independent Wishart matrices. They are asymptotically free; this follows either by the fact that a Wishart matrix is unitarily invariant or, alternatively, by an easy generalization of the genus expansion from (4.18) to the case of several independent Wishart matrices. So the asymptotic eigenvalue distribution of their product is given by the distribution of the product of two free Marchenko-Pastur distributions.

As an example consider two independent Wishart matrices for $c = 5$. Figure 4.13 compares simulations with the analytic formula derived from the S-transform. The first is one realization for $N = 100$ and $M = 500$, the second is one realization for $N = 2000$ and $M = 10000$.

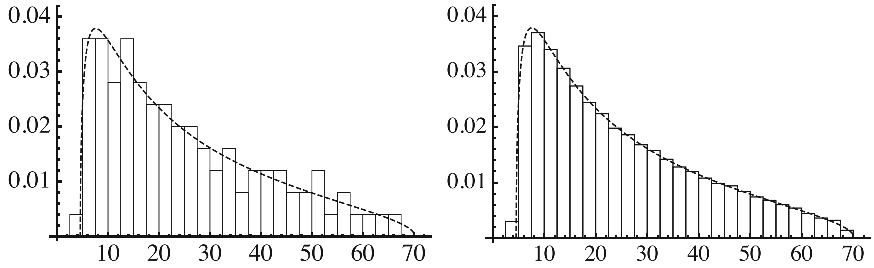


Fig. 4.13 The eigenvalue distribution of the product of two independent complex Wishart matrices. On the left we have one realization with $N = 100$ and $M = 500$. On the right we have one realization with $N = 2000$ and $M = 10000$. See Example 24

Chapter 5
Fluctuations and Second Order Freeness

Given an $N \times N$ random matrix ensemble, we often want to know, in addition to its limiting eigenvalue distribution, how the eigenvalues fluctuate around the limit. This is important in random matrix theory because in many ensembles, the eigenvalues exhibit repulsion, and this feature is often important in applications (see, e.g. [112]). If we take a diagonal random matrix ensemble with independent entries, then the eigenvalues are just the diagonal entries of the matrix and by independence do not exhibit any repulsion. If we take a self-adjoint ensemble with independent entries, i.e. the Wigner ensemble, the eigenvalues are not independent and appear to spread evenly, i.e. there are few bald spots and there is much less clumping; see Fig. 5.1. For some simple ensembles, one can obtain exact formulas measuring this repulsion, i.e. the two-point correlation functions; unfortunately these exact expressions are usually rather complicated. However, just as in the case of the eigenvalue distributions themselves, the large N limit of these distributions is much simpler and can be analysed.

We saw earlier that freeness allows us to find the limiting distributions of $X_N + Y_N$ or $X_N Y_N$ provided we know the limiting distributions of X_N and Y_N individually and X_N and Y_N are asymptotically free. The theory of second order freeness, which was developed in [60, 128, 131], provides an analogous machinery for calculating the fluctuations of sums and products from those of the constituent matrices, provided one has asymptotic second order freeness.

We want to emphasize that on the level of fluctuations, the theory is less robust than on the level of expectations. In particular, whereas on the first order level it does not make any difference for most results whether we consider real or complex random matrices, this is not true any more for second order. What we are going to present here is the theory of second order freeness for *complex* random matrices (modelled according to the GUE). There exists also a real second order freeness theory (modelled according to the GOE, i.e. Gaussian orthogonal ensemble); the general structure of the real theory is the same as in the complex case, but details

J.A. Mingo, R. Speicher, *Free Probability and Random Matrices*,
Fields Institute Monographs 35, DOI 10.1007/978-1-4939-6942-5_5

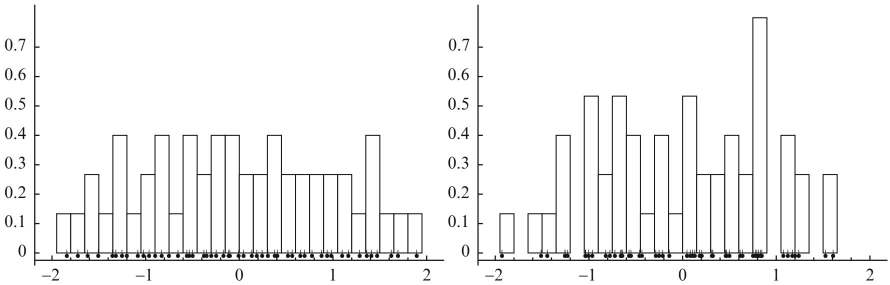


Fig. 5.1 On the left is a histogram of the eigenvalues of an instance of a 50×50 GUE random matrix. The tick marks at the bottom show the actual eigenvalues. On the right we have independently sampled a semi-circular distribution 50 times. We can see that the spacing is more "uniform" in the eigenvalue plot (on the left). The fluctuation moments are a way of measuring this quantitatively

are different. In particular, in the real case, there will be additional contributions in the combinatorial formulas, which correspond to non-orientable surfaces. We will not say more on the real case, but refer to [127, 149].

5.1 Fluctuations of GUE random matrices

To start let us return to our basic example, the GUE. Let X_N be an $N \times N$ self-adjoint Gaussian random matrix, that is, if we write $X_N = (f_{ij})_{i,j=1}^N$ with $f_{ij} = x_{ij} + \sqrt{-1}\, y_{ij}$, then $\{x_{ij}\}_{i \le j} \cup \{y_{ij}\}_{i<j}$ is an independent set of Gaussian random variables with

$$\mathrm{E}(f_{ij}) = 0, \qquad \mathrm{E}(x_{ii}^2) = 1/N, \qquad \text{and} \qquad \mathrm{E}(x_{ij}^2) = \mathrm{E}(y_{ij}^2) = 1/(2N) \quad (\text{for } i \neq j).$$

The eigenvalue distribution of X_N converges almost surely to Wigner's semi-circular law $(2\pi)^{-1}\sqrt{4-t^2}\,dt$, and in particular if f is a polynomial and $\mathrm{tr} = N^{-1}\mathrm{Tr}$ is the normalized trace, then $\{\mathrm{tr}(f(X_N))\}_N$ converges almost surely as $N \to \infty$ to $(2\pi)^{-1}\int_{-2}^{2} f(t)\sqrt{4-t^2}\,dt$. Thus, if f is a polynomial centred with respect to the semi-circle law, i.e.

$$\frac{1}{2\pi}\int_{-2}^{2} f(t)\sqrt{4-t^2}\,dt = 0, \tag{5.1}$$

then $\{\mathrm{tr}(f(X_N))\}_N$ converges almost surely to 0; however, if we rescale by multiplying by N, $\{\mathrm{Tr}(f(X_N))\}_N$ becomes a convergent sequence of random variables, and the limiting covariances for various f's give the *fluctuations* of X_N. Assuming a growth condition on the first two derivatives of f, Johansson [104] was able to show the result below for more general functions f, but we shall just state it for polynomials.

Theorem 1. *Let f be a polynomial such that the centredness condition (5.1) is satisfied and let $\{X_N\}_N$ be the* GUE. *Then $\mathrm{Tr}(f(X_N))$ converges to a Gaussian random variable. Moreover, if $\{C_n\}_n$ are the Chebyshev polynomials of the first kind (rescaled to $[-2,2]$), then $\{\mathrm{Tr}(C_n(X_N))\}_{n=1}^{\infty}$ converge to independent Gaussian random variables with $\lim_N \mathrm{Tr}(C_n(X_N))$ having mean 0 and variance n.*

The Chebyshev polynomials of the first kind are defined by the relation $T_n(\cos\theta) = \cos n\theta$. They are the orthogonal polynomials on $[-1,1]$ which are orthogonal with respect to the arc-sine law $\pi^{-1}(1-x^2)^{-1/2}$. Rescaling to the interval $[-2,2]$ means using the measure $\pi^{-1}(4-x^2)^{-1/2}dx$ and setting $C_n(x) = 2\,T_n(x/2)$. We thus have

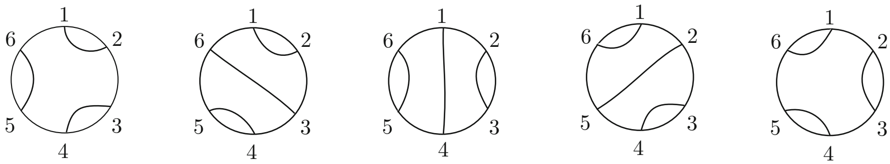
$$
\begin{array}{l|l}
C_0(x) = 2 & C_3(x) = x^3 - 3x \\
C_1(x) = x & C_4(x) = x^4 - 4x^2 + 2 \\
C_2(x) = x^2 - 2 & C_5(x) = x^5 - 5x^3 + 5x
\end{array}
$$

$$\text{and for } n \geq 1,\ C_{n+1}(x) = xC_n(x) - C_{n-1}(x).$$

The reader will be asked to prove some of the above-mentioned properties of C_n (as well as corresponding properties of the second kind analogue U_n) in Exercise 12. We will provide a proof of Theorem 1 at the end of this chapter; see Section 5.6.1.

Recall that in the case of first order freeness, the moments of the GUE had a combinatorial interpretation in terms of planar diagrams. These diagrams led to the notion of free cumulants and the R-transform, which unlocked the whole theory.

For the GUE the moments $\{\alpha_k\}_k$ of the limiting eigenvalue distribution are 0 for k odd and the Catalan numbers for k even. For example, when $k = 6$, $\alpha_6 = 5$, the third Catalan number, and the corresponding diagrams are the five non-crossing pairings on $[6]$.



To understand the fluctuations, we shall introduce another type of planar diagram, this time on an annulus. We shall confine our discussion to ensembles that have what we shall call a second order limiting distribution.

Definition 2. Let $\{X_N\}_N$ be a sequence of random matrices. We say that $\{X_N\}_N$ has a *second order limiting distribution* if there are sequences $\{\alpha_k\}_k$ and $\{\alpha_{p,q}\}_{p,q}$ such that

- for all k, $\alpha_k = \lim_N \mathrm{E}(\mathrm{tr}(X_N^k))$ and
- for all $p \geq 1$ and $q \geq 1$,

$$\alpha_{p,q} = \lim_N \mathrm{cov}\left(\mathrm{Tr}(X_N^p), \mathrm{Tr}(X_N^q)\right)$$

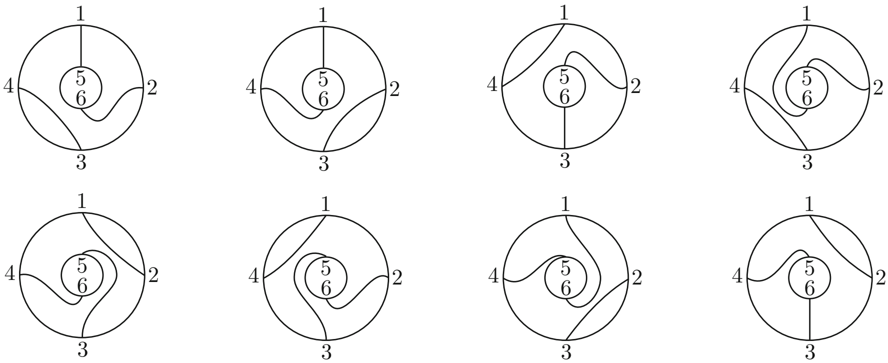
- for all $r > 2$ and all integers $p_1, \dots, p_r \geq 1$

$$\lim_N k_r \left(\mathrm{Tr}(X_N^{p_1}), \mathrm{Tr}(X_N^{p_2}), \dots, \mathrm{Tr}(X_N^{p_r})\right) = 0.$$

Here, k_r are the classical cumulants; note that the α_k are the limits of k_1 (which is the expectation) and $\alpha_{p,q}$ are the limits of k_2 (which is the covariance).

Remark 3. Note that the first condition says that X_N has a limiting eigenvalue distribution in the averaged sense. By the second condition, the variances of normalized traces go asymptotically like $1/N^2$. Thus, by Remark 4.2, the existence of a second order limiting distribution implies actually almost sure convergence to the limit distribution.

We shall next show that the GUE has a second order limiting distribution. The numbers $\{\alpha_{p,q}\}_{p,q}$ that are obtained have an important combinatorial significance as the number of non-crossing annular pairings. Informally, a pairing of the (p,q)-annulus is *non-crossing* or *planar* if when we arrange the numbers 1, 2, 3, ..., p in clockwise order on the outer circle and the numbers $p+1, \dots, p+q$ in counterclockwise order on the inner circle there is a way to draw the pairings so that the lines do not cross and there is at least one string that connects the two circles. For example, $\alpha_{4,2} = 8$, and the eight drawings are shown below.



In Definition 2.7, we defined a partition π of $[n]$ to be non-crossing if a certain configuration, called a crossing, did not appear. A *crossing* was defined to be four points $a < b < c < d \in [n]$ such that a and c are in one block of π and b and d are in another block of π. In [126] a permutation of $[p+q]$ was defined to be a non-crossing annular permutation if no one of five proscribed configurations appeared. It was then shown that under a connectedness condition, this definition was equivalent to the algebraic condition $\#(\pi) + \#(\pi^{-1}\gamma) = p + q$, where $\gamma = (1, 2, 3, \dots, p)(p+1, \dots, p+q)$. In [128, §2.2] another definition was given.

Here we wish to present a natural topological definition (Definition 5) and show that it is equivalent to the algebraic condition in [126]. The key idea is to relate a non-crossing annular permutation to a non-crossing partition and then use an algebraic condition found by Biane [33]. To state the theorem of Biane (Theorem 4), it is necessary to regard a partition as a permutation by putting the elements of its blocks in increasing order. It is also convenient not to distinguish notationally between a partition and the corresponding permutation.

As before, we denote by $\#(\pi)$ the number of blocks or cycles of π. We let (i, j) denote the transposition that switches i and j.

The following theorem tells us when a permutation came from a non-crossing partition. See [137, Prop. 23.23] for a proof. The proof uses induction and two simple facts about permutations.

- If $\pi \in S_n$ and $i, j \in [n]$, then

$$\#(\pi\,(i,j)) = \#(\pi) + 1 \text{ if } i \text{ and } j \text{ are in the same cycle of } \pi$$
$$\#(\pi\,(i,j)) = \#(\pi) - 1 \text{if } i \text{ and } j \text{ are in different cycles of } \pi.$$

- If $|\pi|$ is the minimum number of factors among all factorizations of π into a product of transpositions, then

$$|\pi| + \#(\pi) = n. \tag{5.2}$$

Theorem 4. *Let γ_n denote the permutation in S_n which has the one cycle* $(1, 2, 3, \ldots, n)$. *For all $\pi \in S_n$, we have*

$$\#(\pi) + \#(\pi^{-1}\gamma_n) \leq n + 1; \tag{5.3}$$

and π, considered as a partition, is non-crossing if and only if

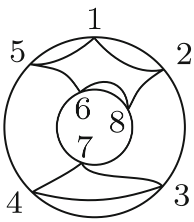
$$\#(\pi) + \#(\pi^{-1}\gamma_n) = n + 1. \tag{5.4}$$

Definition 5. The (p, q)-annulus is the annulus with the integers 1 to p arranged clockwise on the outside circle and $p+1$ to $p+q$ arranged counterclockwise on the inner circle. A permutation π in S_{p+q} is a *non-crossing permutation* on the (p, q)-annulus (or just a *non-crossing annular permutation*) if we can draw the cycles of π between the circles of the annulus so that:

(i) the cycles do not cross,
(ii) each cycle encloses a region between the circles homeomorphic to the disc with boundary oriented clockwise, and
(iii) at least one cycle connects the two circles.

We denote by $S_{NC}(p, q)$ the set of non-crossing permutations on the (p, q)-annulus. The subset consisting of non-crossing pairings on the (p, q)-annulus is denoted by $NC_2(p, q)$.

Example 6. Let $p = 5$ and $q = 3$ and $\pi_1 = (1, 2, 8, 6, 5)(3, 4, 7)$ and $\pi_2 = (1, 2, 8, 6, 5)(3, 7, 4)$. Then π_1 is a non-crossing permutation of the $(5, 3)$-annulus; we can find a drawing which satisfies (i) and (ii) of Definition 5:



But for π_2 we can find a drawing satisfying one of (i) or (ii) but not both. Notice also that if we try to draw π_1 on a disc, we will have a crossing, so π_1 is non-crossing on the annulus but not on the disc. See also Fig. 5.2.

Notice that when we have a partition π of $[n]$ and we want to know if π is non-crossing in the disc sense, property (ii) of Definition 5 is automatic because we always put the elements of the blocks of π in increasing order.

Remark 7. Note that in general we have to distinguish between non-crossing annular permutations and the corresponding partitions. On the disc, the non-crossing condition ensures that for each $\pi \in NC(n)$ there is exactly one corresponding non-crossing permutation (by putting the elements in a block of π in increasing order to read it as a cycle of a permutation). On the annulus, however, this one-to-one correspondence breaks down. Namely, if $\pi \in S_{NC}(p, q)$ has only one through-cycle (a *through-cycle* is a cycle which contains elements from both circles), then the block structure of this cycle is not enough to recover its cycle structure. For example, in $S_{NC}(2, 2)$, we have the following four non-crossing annular permutations:

$$(1, 2, 3, 4), \quad (1, 2, 4, 3), \quad (1, 3, 4, 2), \quad (1, 4, 3, 2).$$

As partitions all four are the same, having one block $\{1, 2, 3, 4\}$; but as permutations they are all different. It is indeed the permutations, and not the partitions, which are relevant for the description of the fluctuations. One should, however, also note that this difference disappears if one has more than one through-cycle. Also for pairings there is no difference between non-crossing annular permutations and partitions. This justifies the notation $NC_2(p, q)$ in this case.

Exercise 1. (i) Let π_1 and π_2 be two non-crossing annular permutations in $S_{NC}(p, q)$, which are the same as partitions. Show that if they have more than one through-cycle, then $\pi_1 = \pi_2$.

(ii) Show that the number of non-crossing annular permutations which are the same as partitions is, in the case of one through-cycle, given by mn, where m and n are the number of elements of the through-cycle on the first and the second circle, respectively.

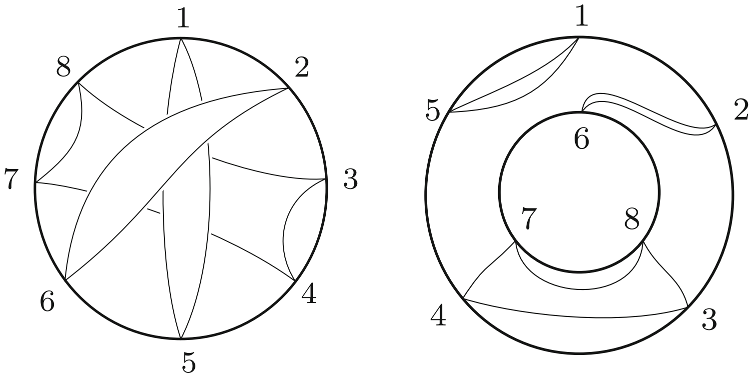
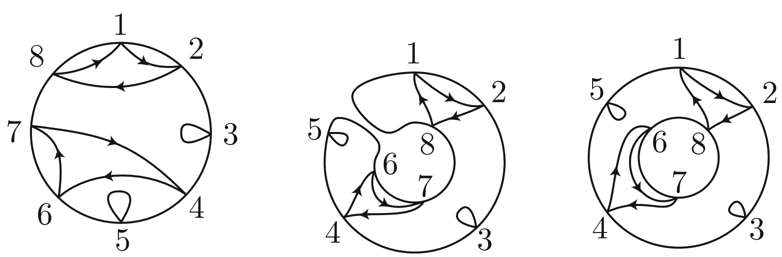


Fig. 5.2 Consider the permutation $\pi = (1,5)(2,6)(3,4,7,8)$. As a disc permutation, it cannot be drawn in a non-crossing way. However, on the $(5,3)$-annulus, it has a non-crossing presentation. Note that we have $\pi^{-1}\gamma_{5,3} = (1,6,4)(2,8)(3)(5)(7)$. So $\#(\pi) + \#(\pi^{-1}\gamma_{5,3}) = 8$

Theorem 8. *Let $\gamma = (1,2,3,\ldots,p)(p+1,\ldots,p+q)$ and $\pi \in S_{p+q}$ be a permutation that has at least one cycle that connects the two cycles of γ. Then π is a non-crossing permutation of the (p,q)-annulus if and only if $\#(\pi) + \#(\pi^{-1}\gamma) = p+q$.*

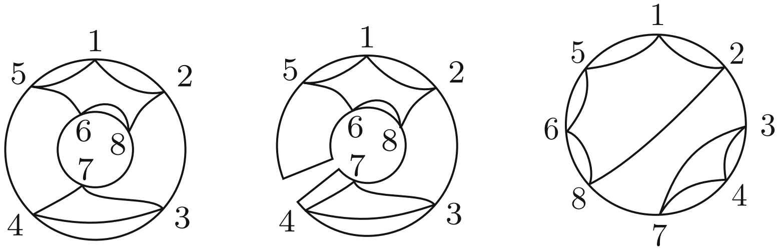
Proof: We must show that the topological property of Definition 5 is equivalent to the algebraic property $\#(\pi) + \#(\pi^{-1}\gamma) = p+q$. A similar equivalence was given in Theorem 4; and we shall use this equivalence to prove Theorem 8.

To begin let us observe that if π is a non-crossing partition of $[p+q]$, we can deform the planar drawing for π on the disc into a drawing on the annulus satisfying the two first conditions of Definition 5 as follows. We deform the disc so that it appears as an annulus with a channel with one side between p and $p+1$ and the other between $p+q$ and 1. We then close the channel and obtain a non-crossing permutation of the (p,q)-annulus.



We have thus shown that every non-crossing partition of $[p+q]$ that satisfies the connectedness condition gives a non-crossing annular permutation of the (p,q)-annulus. We now wish to reverse the procedure.

So let us start with π a non-crossing permutation of the (p,q)-annulus. We chose i such that i and $\pi(i)$ are on different circles, in fact we can assume that $1 \le i \le p$ and $p+1 \le \pi(i) \le p+q$. Such an i always exists because π always has at least one cycle that connects the two circles. We then cut the annulus by making a channel from i to $\gamma^{-1}\pi(i)$. In the illustration below, $i = 4$.



Hence, our π is non-crossing in the disc; however, the order of the points on the disc produced by cutting the annulus is not the standard order – it is the order given by

$$\begin{aligned}\tilde{\gamma} &= \gamma\,(i, \gamma^{-1}\pi(i)) \\ &= (1, \ldots, i, \pi(i), \gamma(\pi(i)), \ldots, p+q, p+1, \ldots, \gamma^{-1}(\pi(i)), \gamma(i), \ldots, p).\end{aligned}$$

Thus, we must show that for i and $\pi(i)$ on different circles, the following are equivalent:

(a) π is non-crossing in the disc with respect to $\tilde{\gamma} = \gamma\,(i, \gamma^{-1}\pi(i))$, and
(b) $\#(\pi) + \#(\pi^{-1}\gamma) = p+q$.

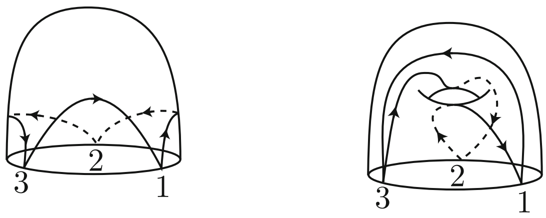
If i and $\pi(i)$ are in different cycles of γ, then i and $\pi^{-1}\gamma(i)$ are in the same cycle of $\pi^{-1}\gamma$. Hence, $\#(\pi^{-1}\gamma\,(i, \pi^{-1}\gamma(i))) = \#(\pi^{-1}\gamma) + 1$. Thus, $\#(\pi) + \#(\pi^{-1}\tilde{\gamma}) = \#(\pi) + \#(\pi^{-1}\gamma) + 1$. Since $\tilde{\gamma}$ has only one cycle, we know, by Theorem 4, that π is non-crossing with respect to $\tilde{\gamma}$ if and only if $\#(\pi) + \#(\pi^{-1}\tilde{\gamma}) = p+q+1$. Thus, π is non-crossing with respect to $\tilde{\gamma}$ if and only if $\#(\pi) + \#(\pi^{-1}\gamma) = p+q$. This shows the equivalence of (a) and (b). □

This result is part of a more general theory of maps on surfaces found by Jacques [102] and Cori [61]. Suppose we have two permutations π and γ in S_n and that π and γ generate a subgroup of S_n that acts transitively on $[n]$. Suppose also that γ has k cycles and we draw k discs on a surface of genus g and arrange the points in the cycles of γ around the circles so that when viewed from the outside, the numbers appear in the same order as in the cycles of γ. We then draw the cycles of π on the surface such that:

- the cycles do not cross, and
- each cycle of π is the oriented boundary of a region on the sphere, oriented with an outward pointing normal, homeomorphic to a disc.

The *genus* of π relative to γ is the smallest g such that the cycles of π can be drawn on a surface of genus g. When $g = 0$, i.e. we can draw π on a sphere, we say that π is *γ-planar*.

In the example below, we let $n = 3$, $\gamma = (1, 2, 3)$ and in the first example $\pi_1 = (1, 2, 3)$ and in the second $\pi_2 = (1, 3, 2)$.



Since π_1 and π_2 have only one cycle, there is no problem with the blocks crossing; it is only to get the correct orientation that we must add a handle for π_2.

Theorem 9. *Suppose $\pi, \gamma \in S_n$ generate a subgroup which acts transitively on $[n]$ and g is the genus of π relative to γ. Then*

$$\#(\pi) + \#(\pi^{-1}\gamma) + \#(\gamma) = n + 2(1 - g). \tag{5.5}$$

Sketch The idea of the proof is to use Euler's formula for the surface of genus g on which we have drawn the cycles of π, as in the definition. Each cycle of γ is a disc numbered according to γ, and we shrink each of these to a point to make the vertices of our simplex. Thus, $V = \#(\gamma)$. The resulting surface will have one face for each cycle of π and one for each cycle of $\pi^{-1}\gamma$. Thus, $F = \#(\pi) + \#(\pi^{-1}\gamma)$. Finally the edges will be the boundaries between the cycles of π and the cycles of $\pi^{-1}\gamma$, and there will be n of these. Thus, $2(1-g) = F - E + V = \#(\pi) + \#(\pi^{-1}\gamma) - n + \#(\gamma)$. □

Remark 10. The requirement that the subgroup generated by π and γ acts transitively is needed to get a connected surface. In the disconnected case, we can replace $2(1 - g)$ by the Euler characteristic of the union of the surfaces.

Now let us return to our discussion of the second order limiting distribution of the GUE.

Theorem 11. *Let $\{X_N\}_N$ be the* GUE. *Then $\{X_N\}_N$ has a second order limiting distribution with fluctuation moments $\{\alpha_{p,q}\}_{p,q}$ where $\alpha_{p,q}$ is the number of non-crossing pairings on a (p, q)-annulus.*

Proof: We have already seen in Theorem 1.7 that

$$\alpha_k = \lim_N \mathrm{E}(\mathrm{tr}(X_N^k))$$

exists for all k and is given by the number of non-crossing pairings of $[k]$. Let us next fix $r \geq 2$ and positive integers $p_1, p_2, \ldots, p_r$, and we shall find a formula for $k_r(\mathrm{Tr}(X_N^{p_1}), \mathrm{Tr}(X_N^{p_2}), \ldots, \mathrm{Tr}(X_N^{p_r}))$.

We shall let $p = p_1 + p_2 + \cdots + p_r$ and γ be the permutation in S_p with the r cycles

$$\gamma = (1, 2, 3, \ldots, p_1)(p_1+1, \ldots, p_1+p_2) \cdots (p_1+\cdots+p_{r-1}+1, \ldots, p_1+\cdots+p_r).$$

Now, with $X_N = (f_{ij})_{i,j=1}^N$,

$$\begin{aligned}\mathrm{E}(\mathrm{Tr}(X_N^{p_1}) \cdots \mathrm{Tr}(X_N^{p_r})) &= \sum \mathrm{E}(f_{i_1,i_2} f_{i_2,i_3} \cdots f_{i_{p_1},i_1} \cdot f_{i_{p_1+1},i_{p_1+2}} \cdots f_{i_{p_1+p_2},i_{p_1+1}} \times \cdots \\ &\qquad \cdots \times f_{i_{p_1+\cdots+p_{r-1}+1},i_{p_1+\cdots+p_{r-1}+2}} \cdots f_{i_{p_1+\cdots+p_r},i_{p_1+\cdots+p_{r-1}+1}}) \\ &= \sum \mathrm{E}(f_{i_1,i_{\gamma(1)}} \cdots f_{i_p,i_{\gamma(p)}})\end{aligned}$$

because the indices of the f's follow the cycles of γ.

Recall that Wick's formula (1.8) tells us how to calculate the expectation of a product of Gaussian random variables. In particular, the expectation will be 0 unless the number of factors is even. Thus, we must have p even and

$$\mathrm{E}(f_{i_1,i_{\gamma(1)}} \cdots f_{i_p,i_{\gamma(p)}}) = \sum_{\pi \in \mathcal{P}_2(p)} \mathrm{E}_\pi(f_{i_1,i_{\gamma(1)}}, \ldots, f_{i_p,i_{\gamma(p)}}).$$

Given a pairing π and a pair (s,t) of π, $\mathrm{E}(f_{i_s,i_{\gamma(s)}} f_{i_t,i_{\gamma(t)}})$ will be 0 unless $i_s = i_{\gamma(t)}$ and $i_t = i_{\gamma(s)}$. Following our usual convention of regarding partitions as permutations and a p-tuple $(i_1, \ldots, i_p)$ as a function $i : [p] \to [N]$, this last condition can be written as $i(s) = i(\gamma(\pi(s)))$ and $i(t) = i(\gamma(\pi(t)))$. Thus, for $\mathrm{E}_\pi(f_{i_1,i_{\gamma(1)}}, \ldots, f_{i_p,i_{\gamma(p)}})$ to be non-zero, we require $i = i \circ \gamma \circ \pi$ or the function i to be constant on the cycles of $\gamma\pi$. When $\mathrm{E}_\pi(f_{i_1,i_{\gamma(1)}}, \ldots, f_{i_p,i_{\gamma(p)}}) \neq 0$, it equals $N^{-p/2}$ (by our normalization of the variance, $\mathrm{E}(|f_{ij}|^2) = 1/N$). An important quantity will then be the number of functions $i : [p] \to [N]$ that are constant on the cycles of $\gamma\pi$; since we can choose the value of the function arbitrarily on each cycle, this number is $N^{\#(\gamma\pi)}$. Hence,

$$\begin{aligned}\mathrm{E}(\mathrm{Tr}(X_N^{p_1}) \cdots \mathrm{Tr}(X_N^{p_r})) &= \sum_{i_1,\ldots,i_p=1}^{N} \sum_{\pi \in \mathcal{P}_2(p)} \mathrm{E}_\pi(f_{i_1,i_{\gamma(1)}}, \ldots, f_{i_p,i_{\gamma(p)}}) \\ &= \sum_{\pi \in \mathcal{P}_2(p)} \sum_{i_1,\ldots,i_p=1}^{N} \mathrm{E}_\pi(f_{i_1,i_{\gamma(1)}}, \ldots, f_{i_p,i_{\gamma(p)}}) \\ &= \sum_{\pi \in \mathcal{P}_2(p)} N^{-p/2} \cdot \#\big(\{i : [p] \to [N] \mid i = i \circ \gamma \circ \pi\}\big) \\ &= \sum_{\pi \in \mathcal{P}_2(p)} N^{\#(\gamma\pi) - p/2}.\end{aligned}$$

The next step is to find which pairings π contribute to the cumulant k_r. Recall that if $Y_1, \dots, Y_r$ are random variables, then

$$k_r(Y_1, \dots, Y_r) = \sum_{\sigma \in \mathcal{P}(r)} \mathrm{E}_\sigma(Y_1, Y_2, \dots, Y_r)\, \mu(\sigma, 1_r)$$

where μ is the Möbius function of the partially ordered set $\mathcal{P}(r)$; see Exercise 1.14. If σ is a partition of $[r]$, there is an associated partition $\tilde{\sigma}$ of $[p]$ where each block of $\tilde{\sigma}$ is a union of cycles of γ; in fact if s and t are in the same block of σ, then the r^{th} and s^{th} cycles of γ

$$(p_1 + \cdots + p_{s-1} + 1, \dots, p_1 + \cdots + p_s) \text{ and } (p_1 + \cdots + p_{t-1} + 1, \dots, p_1 + \cdots + p_t)$$

are in the same block of $\tilde{\sigma}$. Using the same calculation as was used above, we have for $\sigma \in \mathcal{P}(r)$

$$\mathrm{E}_\sigma(\mathrm{Tr}(X_N^{p_1}), \dots, \mathrm{Tr}(X_N^{p_r})) = \sum_{\substack{\pi \in \mathcal{P}_2(p) \\ \pi \leq \tilde{\sigma}}} N^{\#(\gamma\pi) - p/2}.$$

Now given $\pi \in \mathcal{P}(p)$, we let $\hat{\pi}$ be the partition of $[r]$ such that s and t are in the same block of $\hat{\pi}$ if there is a block of π that contains both elements of s^{th} and t^{th} cycles of π. Thus,

$$\begin{aligned} k_r(\mathrm{Tr}(X_N^{p_1}), \dots, \mathrm{Tr}(X_N^{p_r})) &= \sum_{\sigma \in \mathcal{P}(r)} \mu(\sigma, 1_r) \sum_{\substack{\pi \in \mathcal{P}_2(p) \\ \pi \leq \tilde{\sigma}}} N^{\#(\gamma\pi) - p/2} \\ &= \sum_{\pi \in \mathcal{P}_2(p)} N^{\#(\gamma\pi) - p/2} \sum_{\substack{\sigma \in \mathcal{P}(r) \\ \sigma \geq \hat{\pi}}} \mu(\sigma, 1_r). \end{aligned}$$

A fundamental fact of the Möbius function is that for an interval $[\sigma_1, \sigma_2]$ in $\mathcal{P}(r)$ we have $\sum_{\sigma_1 \leq \sigma \leq \sigma_2} \mu(\sigma, \sigma_2) = 0$ unless $\sigma_1 = \sigma_2$ in which case the sum is 1. Thus, $\sum_{\sigma \geq \hat{\pi}} \mu(\sigma, 1_r) = 0$ unless $\hat{\pi} = 1_r$ in which case the sum is 1. Hence,

$$k_r(\mathrm{Tr}(X_N^{p_1}), \dots, \mathrm{Tr}(X_N^{p_r})) = \sum_{\substack{\pi \in \mathcal{P}_2(p) \\ \hat{\pi} = 1_r}} N^{\#(\gamma\pi) - p/2}.$$

When $\hat{\pi} = 1_r$, the subgroup generated by γ and π acts transitively on $[p]$, and thus Euler's formula (5.5) can be applied. Thus, for the π which appear in the sum, we have

$$\begin{aligned}\#(\gamma\pi) &= \#(\pi^{-1}\gamma)\\ &= p + 2(1-g) - \#\pi - \#\gamma\\ &= p + 2(1-g) - p/2 - r\\ &= p/2 + 2(1-g) - r,\end{aligned}$$

and thus $\#(\gamma\pi) - p/2 = 2 - r - 2g$. So the leading order of k_r, corresponding to the γ-planar π, is given by N^{2-r}. Taking the limit $N \to \infty$ gives the assertion. It shows that k_r goes to zero for $r > 2$, and for $r = 2$ the limit is given by the number of γ-planar π, i.e. by $\#(NC_2(p,q))$. □

5.2 Fluctuations of several matrices

Up to now, we have looked on the limiting second order distribution of one GUE random matrix. One can generalize those calculations quite easily to the case of several independent GUE random matrices.

Exercise 2. Suppose $X_1^{(N)}, \dots, X_s^{(N)}$ are s independent $N \times N$ GUE random matrices. Then we have, for all $p, q \geq 1$ and for all $1 \leq r_1, \dots, r_{p+q} \leq s$ that

$$\lim_N k_2\big(\mathrm{Tr}(X_{r_1}^{(N)} \cdots X_{r_p}^{(N)}), \mathrm{Tr}(X_{r_{p+1}}^{(N)} \cdots X_{r_{p+q}}^{(N)})\big) = \#\big(NC_2^{(r)}(p,q)\big),$$

where $NC_2^{(r)}(p,q)$ denotes the non-crossing annular pairings which respect the colour, i.e. those $\pi \in NC_2(p,q)$ such that $(k,l) \in \pi$ only if $r_k = r_l$. Furthermore, all higher order cumulants of unnormalized traces go to zero.

Maybe more interesting is the situation where we also include deterministic matrices. Similarly to the first order case, we expect to see some second order freeness structure appearing there. Of course, the calculation of the asymptotic fluctuations of mixed moments in GUE and deterministic matrices will involve the (first order) limiting distribution of the deterministic matrices. Let us first recall what we mean by this.

Definition 12. Suppose that we have, for each $N \in \mathbb{N}$, deterministic $N \times N$ matrices $D_1^{(N)}, \dots, D_s^{(N)} \in M_N(\mathbb{C})$ and a non-commutative probability space $(\mathcal{A}, \varphi)$ with elements $d_1, \dots, d_s \in \mathcal{A}$ such that we have for each polynomial $p \in \mathbb{C}\langle x_1, \dots, x_s\rangle$ in s non-commuting variables

$$\lim_N \mathrm{tr}\big(p(D_1^{(N)}, \dots, D_s^{(N)})\big) = \varphi\big(p(d_1, \dots, d_s)\big).$$

Then we say that $(D_1^{(N)}, \dots, D_s^{(N)})_N$ has a *limiting distribution* given by $(d_1, \dots, d_s) \in (\mathcal{A}, \varphi)$.

Theorem 13. *Suppose $X_1^{(N)}, \dots, X_s^{(N)}$ are s independent $N \times N$ GUE random matrices. Fix $p, q \geq 1$ and let $\{D_1^{(N)}, \dots, D_{p+q}^{(N)}\} \subseteq M_N(\mathbb{C})$ be deterministic $N \times N$ matrices with limiting distribution given by $d_1, \dots, d_{p+q} \in (\mathcal{A}, \varphi)$. Then we have for all $1 \leq r_1, \dots, r_{p+q} \leq s$ that*

$$\lim_N k_2 \left(\mathrm{Tr}(D_1^{(N)} X_{r_1}^{(N)} \cdots D_p^{(N)} X_{r_p}^{(N)}), \mathrm{Tr}(D_{p+1}^{(N)} X_{r_{p+1}}^{(N)} \cdots D_{p+q}^{(N)} X_{r_{p+q}}^{(N)}) \right) = \sum_{\pi \in NC_2^{(r)}(p,q)} \varphi_{\gamma_{p,q}\pi}(d_1, \dots, d_{p+q}),$$

where the sum runs over all $\pi \in NC_2(p,q)$ such that $(k,l) \in \pi$ only if $r_k = r_l$ and where

$$\gamma_{p,q} = (1, \dots, p)(p+1, \dots, p+q) \in S_{p+q}. \tag{5.6}$$

Proof: Let us first calculate the expectation of the product of the two traces. For better legibility, we suppress in the following the upper index N. We write as usual $X_r^{(N)} = (f_{ij}^{(r)})$ and $D_p^{(N)} = (d_{ij}^{(p)})$. We will denote by $\mathcal{P}_2^{(r)}(p+q)$ the pairings of $[p+q]$ which respect the colour $r = (r_1, \dots, r_{p+q})$ and by $\mathcal{P}_{2,c}^{(r)}(p+q)$ the pairings in $\mathcal{P}_2^{(r)}(p+q)$ where at least one pair connects a point in $[p]$ to a point in $[p+1, p+q] = \{p+1, p+2, \dots, p+q\}$.

$$\begin{aligned}
&\mathrm{E}\big(\mathrm{Tr}(D_1 X_{r_1} \cdots D_p X_{r_p}) \mathrm{Tr}(D_{p+1} X_{r_{p+1}} \cdots D_{p+q} X_{r_{p+q}})\big) \\
&= \sum_{\substack{i_1, \dots, i_{p+q} \\ j_1, \dots, j_{p+q}}} \mathrm{E}\left(d_{i_1 j_1}^{(1)} f_{j_1 i_2}^{(r_1)} d_{i_2 j_2}^{(2)} \cdots d_{i_p j_p}^{(p)} f_{j_p i_1}^{(r_p)} \cdot d_{i_{p+1} j_{p+1}}^{(p+1)} \cdots d_{i_{p+q} j_{p+q}}^{(p+q)} f_{j_{p+q} i_{p+1}}^{(r_{p+q})} \right) \\
&= \sum_{\substack{i_1, \dots, i_{p+q} \\ j_1, \dots, j_{p+q}}} \mathrm{E}\left(f_{j_1 i_2}^{(r_1)} \cdots f_{j_{p+q} i_{p+1}}^{(r_{p+q})} \right) d_{i_1 j_1}^{(1)} \cdots d_{i_{p+q} j_{p+q}}^{(p+q)} \\
&= \sum_{\substack{i_1, \dots, i_{p+q} \\ j_1, \dots, j_{p+q}}} \sum_{\pi \in \mathcal{P}_2^{(r)}(p+q)} N^{-(p+q)/2} \delta_{j, i \circ \gamma_{p,q} \circ \pi} \cdot d_{i_1 j_1}^{(1)} \cdots d_{i_{p+q} j_{p+q}}^{(p+q)} \\
&= N^{-(p+q)/2} \sum_{\pi \in \mathcal{P}_2^{(r)}(p+q)} \sum_{\substack{i_1, \dots, i_{p+q} \\ j_1, \dots, j_{p+q} \\ j = i \circ \gamma_{p,q} \circ \pi}} d_{i_1 j_1}^{(1)} \cdots d_{i_{p+q} j_{p+q}}^{(p+q)} \\
&= N^{-(p+q)/2} \sum_{\pi \in \mathcal{P}_2^{(r)}(p+q)} \sum_{i_1, \dots, i_{p+q}} d_{i_1, i_{\gamma_{p,q} \circ \pi(1)}}^{(1)} \cdots d_{i_{p+q}, i_{\gamma_{p,q} \circ \pi(p+q)}}^{(p+q)} \\
&= N^{-(p+q)/2} \sum_{\pi \in \mathcal{P}_2^{(r)}(p+q)} \mathrm{Tr}_{\gamma_{p,q}\pi}(D_1, \dots, D_{p+q}).
\end{aligned}$$

Thus, by subtracting the disconnected pairings, we get for the covariance

$$
\begin{aligned}
k_2(\mathrm{Tr}(D_1X_{r_1}\cdots D_pX_{r_p}),&\mathrm{Tr}(D_{p+1}X_{r_{p+1}}\cdots D_{p+q}X_{r_{p+q}}))\\
&= \sum_{\pi\in\mathcal{P}_{2,c}^{(r)}(p+q)} N^{-(p+q)/2}\mathrm{Tr}_{\gamma_{p,q}\pi}(D_1,\ldots,D_{p+q})\\
&= \sum_{\pi\in\mathcal{P}_{2,c}^{(r)}(p+q)} N^{\#(\gamma_{p,q}\pi)-(p+q)/2}\mathrm{tr}_{\gamma_{p,q}\pi}(D_1,\ldots,D_{p+q}).
\end{aligned}
$$

For $\pi\in\mathcal{P}_{2,c}(p+q)$, we have $\#(\pi)+\#(\gamma_{p,q}\pi)+\#(\gamma_{p,q})=p+q+2(1-g)$, and hence $\#(\gamma_{p,q}\pi)-\frac{p+q}{2}=-2g$. The genus g is always ≥ 0 and equal to 0 only when π is non-crossing. Thus,

$$
\begin{aligned}
k_2(\mathrm{Tr}(D_1X_{r_1}\cdots D_pX_{r_p}),&\,\mathrm{Tr}(D_{p+1}X_{r_{p+1}}\cdots D_{p+q}X_{r_{p+q}}))\\
&= \sum_{\pi\in NC_2^{(r)}(p,q)} \mathrm{tr}_{\gamma_{p,q}\pi}(D_1,\ldots,D_{p+q})+O(N^{-1}),
\end{aligned}
$$

and the assertion follows by taking the limit $N\to\infty$. □

Remark 14. Note that Theorem 13 shows that the variance of the corresponding normalized traces is $O(N^{-2})$. Indeed the theorem shows that the variance of the unnormalized traces converges, so by normalizing the trace, we get that the variance of the normalized traces decreases like N^{-2}. This proves then the almost sure convergence claimed in Theorem 4.4.

We would like to replace the deterministic matrices $D_1^{(N)},\ldots,D_{p+q}^{(N)}$ in Theorem 13 by random matrices and see if we can still conclude that the variances of the normalized mixed traces decrease like N^{-2}. As was observed at the end of Section 4.2, we have to assume more than just the existence of a limiting distribution of the $D^{(N)}$'s. In the following definition, we isolate this additional property.

Definition 15. We shall say the random matrix ensemble $\{D_1^{(N)},\ldots,D_p^{(N)}\}_N$ has *bounded higher cumulants* if we have for all $r\geq 2$ and for any unnormalized traces $Y_1,\ldots,Y_r$ of monomials in $D_1^{(N)},\ldots,D_p^{(N)}$ that

$$
\sup_N |k_r(Y_1,\ldots,Y_r)|<\infty.
$$

Note that this is a property of the algebra generated by the D's. We won't prove it here, but for many examples, we have $k_r(Y_1,\ldots,Y_r)=O(N^{2-r})$ with the Y_i's as above. These examples include the GUE, Wishart, and Haar distributed unitary random matrices.

Theorem 16. *Suppose $X_1^{(N)}, \dots, X_s^{(N)}$ are s independent $N \times N$ GUE random matrices. Fix $p, q \geq 1$ and let $\{D_1^{(N)}, \dots, D_{p+q}^{(N)}\} \subseteq M_N(\mathbb{C})$ be random $N \times N$ matrices with a limiting distribution and with bounded higher cumulants. Then we have for all $1 \leq r_1, \dots, r_{p+q} \leq s$ that*

$$k_2\left(\mathrm{tr}(D_1^{(N)} X_{r_1}^{(N)} \cdots D_p^{(N)} X_{r_p}^{(N)}), \mathrm{tr}(D_{p+1}^{(N)} X_{r_{p+1}}^{(N)} \cdots D_{p+q}^{(N)} X_{r_{p+q}}^{(N)})\right) = O(N^{-2}).$$

Proof: We rewrite the proof of Theorem 13 with the change that the D's are now random to get

$$\begin{aligned} &\mathrm{E}\left(\mathrm{Tr}(D_1 X_{r_1} \cdots D_p X_{r_p}) \mathrm{Tr}(D_{p+1} X_{r_{p+1}} \cdots D_{p+q} X_{r_{p+q}})\right) \\ &\qquad = N^{-(p+q)/2} \sum_{\pi \in \mathcal{P}_2^{(r)}(p+q)} \mathrm{E}\big(\mathrm{Tr}_{\gamma_{p,q}\pi}(D_1, \dots, D_{p+q})\big), \end{aligned}$$

and

$$\begin{aligned} &\mathrm{E}\big(\mathrm{Tr}(D_1 X_{r_1} \cdots D_p X_{r_p})\big) \cdot \mathrm{E}\big(\mathrm{Tr}(D_{p+1} X_{r_{p+1}} \cdots D_{p+q} X_{r_{p+q}})\big) \\ &\quad = N^{-(p+q)/2} \sum_{\substack{\pi_1 \in \mathcal{P}_2^{(r)}(p) \\ \pi_2 \in \mathcal{P}_2^{(r)}(q)}} \mathrm{E}\big(\mathrm{Tr}_{\gamma_p \pi_1}(D_1, \dots, D_p)\big) \cdot \mathrm{E}\big(\mathrm{Tr}_{\gamma_q \pi_2}(D_{p+1}, \dots, D_{p+q})\big). \end{aligned}$$

Here, γ_p denotes as usual the one cycle permutation $\gamma_p = (1, 2, \dots, p) \in S_p$ and similar for γ_q. We let $\mathcal{P}_{2,d}^{(r)}(p+q)$ be the pairings in $\mathcal{P}_2^{(r)}(p+q)$ which do not connect $[p]$ to $[p+1, p+q]$. Then we can write $\mathcal{P}_2^{(r)}(p+q) = \mathcal{P}_{2,c}^{(r)}(p+q) \cup \mathcal{P}_{2,d}^{(r)}(p+q)$, as a disjoint union. Moreover, we can identify $\mathcal{P}_{2,d}^{(r)}(p+q)$ with $\mathcal{P}_2^{(r)}(p) \times \mathcal{P}_2^{(r)}(q)$.

Thus, by subtracting the disconnected pairings, we get for the covariance

$$\begin{aligned} &k_2\left(\mathrm{Tr}(D_1 X_{r_1} \cdots D_p X_{r_p}), \mathrm{Tr}(D_{p+1} X_{r_{p+1}} \cdots D_{p+q} X_{r_{p+q}})\right) \\ &\quad = \sum_{\pi \in \mathcal{P}_{2,c}^{(r)}(p+q)} N^{-(p+q)/2} \mathrm{E}\big(\mathrm{Tr}_{\gamma_{p,q}\pi}(D_1, \dots, D_{p+q})\big) \\ &\qquad + \sum_{\substack{\pi_1 \in \mathcal{P}_2^{(r)}(p) \\ \pi_2 \in \mathcal{P}_2^{(r)}(q)}} N^{-(p+q)/2} \mathrm{E}\big(\mathrm{Tr}_{\gamma_{p,q}\pi_1\pi_2}(D_1, \dots, D_{p+q})\big) \\ &\qquad - \sum_{\substack{\pi_1 \in \mathcal{P}_2^{(r)}(p) \\ \pi_2 \in \mathcal{P}_2^{(r)}(q)}} N^{-(p+q)/2} \mathrm{E}\big(\mathrm{Tr}_{\gamma_p \pi_1}(D_1, \dots, D_p)\big) \cdot \mathrm{E}\big(\mathrm{Tr}_{\gamma_q \pi_2}(D_{p+1}, \dots, D_{p+q})\big) \end{aligned}$$

$$= \sum_{\pi \in \mathcal{P}_{2,c}^{(r)}(p+q)} N^{-(p+q)/2} \mathrm{E}\big(\mathrm{Tr}_{\gamma_{p,q}\pi}(D_1, \dots, D_{p+q})\big)$$

$$+ \sum_{\substack{\pi_1 \in \mathcal{P}_2^{(r)}(p) \\ \pi_2 \in \mathcal{P}_2^{(r)}(q)}} N^{-(p+q)/2} k_2 \left(\mathrm{Tr}_{\gamma_p \pi_1}(D_1, \dots, D_p), \mathrm{Tr}_{\gamma_q \pi_2}(D_{p+1}, \dots, D_{p+q})\right).$$

We shall show that both of these terms are $O(1)$, and thus after normalizing the traces, $k_2 = O(N^{-2})$. For the first term, this is the same argument as in the proof of Theorem 5.13. So let $\pi_1 \in \mathcal{P}_2^{(r)}(p)$ and $\pi_2 \in \mathcal{P}_2^{(r)}(q)$. We let $s = \#(\gamma_p \pi_1)$ and $t = \#(\gamma_q \pi_2)$. Since $\gamma_p \pi_1$ has s cycles, we may write $\mathrm{Tr}_{\gamma_p \pi_1}(D_1, \dots, D_p) = Y_1 \cdots Y_s$ with each Y_i of the form $\mathrm{Tr}(D_{l_1} \cdots D_{l_k})$. Likewise since $\gamma_q \pi_2$ has t cycles, we may write $\mathrm{Tr}_{\gamma_q \pi_2}(D_{p+1}, \dots, D_{p+q}) = Y_{s+1} \cdots Y_{s+t}$ with the Y's of the same form as before. Now by our assumption on the D's, we know that for $u \geq 2$ we have $k_u(Y_{i_1}, \dots, Y_{i_u}) = O(1)$. Using the product formula for classical cumulants, see Equation (1.16), we have that

$$k_2(Y_1 \cdots Y_s, Y_{s+1} \cdots Y_{s+t}) = \sum_{\tau \in \mathcal{P}(s+t)} k_\tau(Y_1, \dots, Y_{s+t})$$

where τ must connect $[s]$ to $[s+1, s+t]$. Now $k_\tau(Y_1, \dots, Y_{s+t}) = O(N^c)$ where c is the number of singletons in τ. Thus, the order of $N^{-(p+q)/2} k_\tau(Y_1, \dots, Y_{s+t})$ is $N^{c-(p+q)/2}$. So we are reduced to showing that $c \leq (p+q)/2$. Since τ connects $[s]$ to $[s+1, s+t]$, τ must have a block with at least two elements. Thus, the number of singletons is at most $s + t - 2$. But $s = \#(\gamma_p \pi_1) \leq p/2 + 1$ and $t = \#(\gamma_q \pi_2) \leq q/2 + 1$ by Corollary 1.6. Thus, $c \leq (p+q)/2$ as claimed. □

5.3 Second order probability space and second order freeness

Recall that a *non-commutative probability space* $(\mathcal{A}, \varphi)$ consists of an algebra over $\mathbb{C}$ and a linear functional $\varphi : \mathcal{A} \to \mathbb{C}$, with $\varphi(1) = 1$. Such a non-commutative probability space is called *tracial*, if φ is also a trace, i.e. if $\varphi(ab) = \varphi(ba)$ for all $a, b \in \mathcal{A}$.

To provide the general framework for second order freeness, we introduce now the idea of a second order probability space, $(\mathcal{A}, \varphi, \varphi_2)$.

Definition 17. Let $(\mathcal{A}, \varphi)$ be a tracial non-commutative probability space. Suppose that we have in addition a bilinear functional $\varphi_2 : \mathcal{A} \times \mathcal{A} \to \mathbb{C}$ such that:

- φ_2 is symmetric in its two variables, i.e. we have $\varphi_2(a, b) = \varphi_2(b, a)$ for all $a, b \in \mathcal{A}$
- φ_2 is tracial in each variable
- $\varphi_2(1, a) = 0 = \varphi_2(a, 1)$ for all $a \in \mathcal{A}$.

Then we say that $(\mathcal{A}, \varphi, \varphi_2)$ is a *second order non-commutative probability space*.

Usually our second order limit elements will arise as limits of random matrices, where φ encodes the asymptotic behaviour of the expectation of traces, whereas φ_2 does the same for the covariances of traces. As we have seen before, in typical examples (as the GUE), we should consider the expectation of the normalized trace tr, but the covariances of the unnormalized traces Tr.

As we have seen in Theorem 16, one usually also needs some control over the higher order cumulants; requiring bounded higher cumulants for the unnormalized traces of the D's was enough to control the variances of the mixed unnormalized traces. However, as in the case of one matrix (see Definition 2), we will in the following definition require instead of boundedness of the higher cumulants the stronger condition that they converge to zero. This definition from [131] makes some arguments easier and is usually satisfied in all relevant random matrix models. Let us point out that, as remarked in [127], the whole theory could also be developed with the boundedness condition instead.

Definition 18. Suppose we have a sequence of random matrices $\{A_1^{(N)}, \ldots, A_s^{(N)}\}_N$ and random variables $a_1, \ldots, a_s$ in a second order non-commutative probability space. We say that $(A_1^{(N)}, \ldots, A_s^{(N)})_N$ has the *second order limit* $(a_1, \ldots, a_s)$ if we have:

- for all $p \in \mathbb{C}\langle x_1, \ldots, x_s\rangle$

$$\lim_N \mathrm{E}\left(\mathrm{tr}(p(A_1^{(N)}, \ldots, A_s^{(N)}))\right) = \varphi\big(p(a_1, \ldots, a_s)\big);$$

- for all $p_1, p_2 \in \mathbb{C}\langle x_1, \ldots, x_s\rangle$

$$\lim_N \mathrm{cov}\left(\mathrm{Tr}(p_1(A_1^{(N)}, \ldots, A_s^{(N)})), \mathrm{Tr}(p_2(A_1^{(N)}, \ldots, A_s^{(N)}))\right) = \varphi_2\big(p_1(a_1, \ldots, a_s), p_2(a_1, \ldots, a_s)\big);$$

- for all $r \geq 3$ and all $p_1, \ldots, p_r \in \mathbb{C}\langle x_1, \ldots, x_s\rangle$

$$\lim_N k_r\left(\mathrm{Tr}(p_1(A_1^{(N)}, \ldots, A_s^{(N)})), \ldots, \mathrm{Tr}(p_r(A_1^{(N)}, \ldots, A_s^{(N)}))\right) = 0.$$

Remark 19. As in Remark 3, the second condition implies that we have almost sure convergence of the (first order) distribution of the $\{A_1^{(N)}, \ldots, A_s^{(N)}\}_N$. So in particular, if the $a_1, \ldots, a_s$ are free, then the existence of a second order limit includes also the fact that $A_1^{(N)}, \ldots, A_s^{(N)}$ are almost surely asymptotically free.

Example 20. A trivial example of second order limit is given by deterministic matrices. If $\{D_1^{(N)}, \ldots, D_s^{(N)}\}$ are deterministic $N \times N$ matrices with limiting distribution, then $k_r(Y_1, \ldots, Y_r) = 0$ for $r > 1$ and for any polynomials Y_i in the D's. So $(D_1^{(N)}, \ldots, D_s^{(N)})_N$ has a second order limiting distribution; φ is given by the limiting distribution and φ_2 is identically zero.

Example 21. Define $(\mathcal{A}, \varphi, \varphi_2)$ by $\mathcal{A} = \mathbb{C}\langle s\rangle$ and

$$\varphi(s^k) = \#\big(NC_2(k)\big) \qquad \text{and} \qquad \varphi_2(s^p, s^q) = \#\big(NC_2(p,q)\big). \tag{5.7}$$

Then $(\mathcal{A}, \varphi, \varphi_2)$ is a second order probability space and s is, by Theorem 11, the second order limit of a GUE random matrix. In first order, s is, of course, just a semi-circular element in $(\mathcal{A}, \varphi)$. We will address a distribution given by (5.7) as a *second order semi-circle distribution.*

Exercise 3. Prove that the second order limit of a Wishart random matrix with rate c (see Section 4.5.1) is given by $(\mathcal{A}, \varphi, \varphi_2)$ with $\mathcal{A} = \mathbb{C}\langle x\rangle$ and

$$\varphi(x^n) = \sum_{\pi \in NC(n)} c^{\#(\pi)} \qquad \text{and} \qquad \varphi_2(x^m, x^n) = \sum_{\pi \in S_{NC}(m,n)} c^{\#(\pi)}. \tag{5.8}$$

We will address a distribution given by (5.8) as a *second order free Poisson distribution (of rate c).*

Example 22. Define $(\mathcal{A}, \varphi, \varphi_2)$ by $\mathcal{A} = \mathbb{C}\langle u, u^{-1}\rangle$ and, for $k, p, q \in \mathbb{Z}$,

$$\varphi(u^k) = \begin{cases} 0, & k \neq 0 \\ 1, & k = 0 \end{cases} \qquad \text{and} \qquad \varphi_2(u^p, u^q) = |p|\delta_{p,-q}. \tag{5.9}$$

Then $(\mathcal{A}, \varphi, \varphi_2)$ is a second order probability space, and u is the second order limit of Haar distributed unitary random matrices. In first order u is of course just a Haar unitary in $(\mathcal{A}, \varphi)$. We will address a distribution given by (5.9) as a *second order Haar unitary.*

Exercise 4. Prove the statement from the previous example: Show that for Haar distributed $N \times N$ unitary random matrices U we have

$$\lim_N k_2\left(\mathrm{Tr}(U^p), \mathrm{Tr}(U^q)\right) = \begin{cases} |p|, & \text{if } p = -q \\ 0, & \text{otherwise} \end{cases}$$

and that the higher order cumulants of unnormalized traces of polynomials in U and U^* go to zero.

Example 23. Let us now consider the simplest case of several variables, namely, the limit of s independent GUE. According to Exercise 2, their second order limit is given by $(\mathcal{A}, \varphi, \varphi_2)$ where $\mathcal{A} = \mathbb{C}\langle x_1, \dots, x_s\rangle$ and

$$\varphi\big(x_{r(1)} \cdots x_{r(k)}\big) = \#\big(NC_2^{(r)}(k)\big)$$

and

$$\varphi_2\big(x_{r(1)} \cdots x_{r(p)}, x_{r(p+1)} \cdots x_{r(p+q)}\big) = \#\big(NC_2^{(r)}(p,q)\big).$$

In the same way as we used in Chapter 1 the formula for φ as our guide to the definition of the notion of freeness, we will now have a closer look on the corresponding formula for φ_2 and try to extract from this a concept of second order freeness.

As in the first order case, let us consider φ_2 applied to alternating products of centred variables, i.e. we want to understand

$$\varphi_2\big((x_{i_1}^{m_1} - c_{m_1}1)\cdots(x_{i_p}^{m_p} - c_{m_p}1), (x_{j_1}^{n_1} - c_{n_1}1)\cdots(x_{j_q}^{n_q} - c_{n_q}1)\big),$$

where $c_m := \varphi(x_i^m)$ (which is independent of i). The variables are here assumed to be alternating in each argument, i.e. we have

$$i_1 \neq i_2 \neq \cdots \neq i_{p-1} \neq i_p \qquad \text{and} \qquad j_1 \neq j_2 \neq \cdots \neq j_{q-1} \neq j_q.$$

In addition, since the whole theory relies on φ_2 being tracial in each of its arguments (as the limit of variances of traces), we will actually assume that it is alternating in a cyclic way, i.e. that we also have $i_p \neq i_1$ and $j_q \neq j_1$.

Let us put $m := m_1 + \cdots + m_p$ and $n := n_1 + \cdots + n_q$. Furthermore, we call the consecutive numbers corresponding to the factors in our arguments "intervals"; so the intervals on the first circle are

$$(1, \ldots, m_1), (m_1 + 1, \ldots, m_1 + m_2), \ldots, (m_1 + \cdots + m_{p-1} + 1, \ldots, m),$$

and the intervals on the second circle are

$$(m + 1, \ldots, m + n_1), \ldots, (m + n_1 + \cdots + n_{q-1} + 1, \ldots, m + n).$$

By the same arguing as in Chapter 1, one can convince oneself that the subtraction of the means has the effect that instead of counting all $\pi \in NC_2(m, n)$, we count now only those where each interval is connected to at least one other interval. In the first order case, because of the non-crossing property, there were no such π, and the corresponding expression was zero. Now, however, we can connect an interval from one circle to an interval of the other circle, and there are possibilities to do this in a non-crossing way. Renaming $a_k := x_{i_k}^{m_k} - c_{m_k}1$ and $b_l := x_{j_l}^{n_l} - c_{n_l}1$ leads then exactly to the formula which will be our defining property of second order freeness in the next definition.

Definition 24. Let $(\mathcal{A}, \varphi, \varphi_2)$ be a second order non-commutative probability space and $(\mathcal{A}_i)_{i\in I}$ a family of unital subalgebras of $\mathcal{A}$. We say that $(\mathcal{A}_i)_{i\in I}$ are *free of second order* if (i) the subalgebras $(\mathcal{A}_i)_{i\in I}$ are free of first order, i.e. in the sense of §1.11, and (ii) the fluctuation moments of centred and cyclically alternating elements can be computed from their ordinary moments in the following way. Recall that $a_1, \ldots, a_n \in \cup_i \mathcal{A}_i$ are *cyclically alternating* if $a_i \in \mathcal{A}_{j_i}$ and

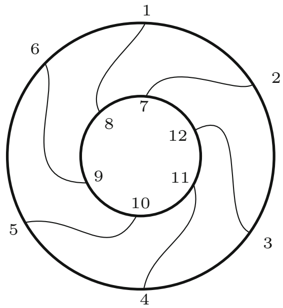


Fig. 5.3 The spoke diagram for $\pi = (1, 8)(2, 7)(3, 12)(4, 11)(5, 10)(6, 9)$. For this permutation, we have $\varphi_\pi(a_1, \ldots, a_{12}) = \varphi(a_1a_8)\varphi(a_2a_7)\varphi(a_3a_{12})\varphi(a_4a_{11})\varphi(a_5a_{10})\varphi(a_6a_9)$

$j_1 \neq j_2 \neq \cdots j_n \neq j_1$. The second condition (ii) is that given two tuples $a_1, \ldots, a_m$ and $b_1, \ldots, b_n$ which are centred and cyclically alternating, then for $(m, n) \neq (1, 1)$

$$\varphi_2(a_1 \cdots a_m, b_1 \cdots b_n) = \delta_{mn} \sum_{k=0}^{n-1} \prod_{i=1}^{n} \varphi(a_i b_{k-i}), \tag{5.10}$$

where the indices of b_i are interpreted modulo n; when $m = n = 1$, we have $\varphi_2(a_1, b_1) = 0$ if a_1 and b_1 come from different $\mathcal{A}_i$'s.

Second order freeness for random variables or for sets is, as usual, defined as second order freeness for the unital subalgebras generated by the variables or the sets, respectively.

Equation (5.10) has the following diagrammatic interpretation. A non-crossing permutation of an (m, n)-annulus is called a *spoke diagram* if all cycles have just two elements (i, j), and the elements are on different circles, i.e. $i \in [m]$ and $j \in [m + 1, m + n]$; see Fig. 5.3. We can only have a spoke diagram if $m = n$; the set of spoke diagrams is denoted by $Sp(n)$. With this notation, Equation (5.10) can also be written as

$$\varphi_2(a_1 \cdots a_m, b_1 \cdots b_n) = \delta_{mn} \sum_{\pi \in Sp(n)} \varphi_\pi(a_1, \ldots, a_m, b_1, \ldots, b_n). \tag{5.11}$$

Exercise 5. Let $\mathcal{A}_1, \mathcal{A}_2 \subset \mathcal{A}$ be free of second order in $(\mathcal{A}, \varphi, \varphi_2)$ and consider $a_1, a_2 \in \mathcal{A}_1$ and $b_1, b_2 \in \mathcal{A}_2$. Show that the definition of second order freeness implies the following formula for the first non-trivial mixed fluctuation moment:

$$\begin{aligned}\varphi_2(a_1b_1, a_2b_2) = {} & \varphi(a_1a_2)\varphi(b_1b_2) - \varphi(a_1a_2)\varphi(b_1)\varphi(b_2) \\ & - \varphi(a_1)\varphi(a_2)\varphi(b_1b_2) + \varphi(a_1)\varphi(a_2)\varphi(b_1)\varphi(b_2) \\ & + \varphi_2(a_1, a_2)\varphi(b_1)\varphi(b_2) + \varphi(a_1)\varphi(a_2)\varphi_2(b_1, b_2).\end{aligned}$$

Let us define now the asymptotic version of second order freeness.

Definition 25. We say $\{A_1^{(N)},\dots,A_s^{(N)}\}_N$ and $\{B_1^{(N)},\dots,B_t^{(N)}\}_N$ are *asymptotically free of second order* if there is a second order non-commutative probability space $(\mathcal{A},\varphi,\varphi_2)$ and elements $a_1,\dots,a_s,b_1,\dots,b_t\in\mathcal{A}$ such that

- ∘ $(A_1^{(N)},\dots,A_s^{(N)},B_1^{(N)},\dots,B_t^{(N)})_N$ has a second order limit $(a_1,\dots,a_s,b_1,\dots,b_t)$
- ∘ $\{a_1,\dots,a_s\}$ and $\{b_1,\dots,b_t\}$ are free of second order.

Remark 26. Note that asymptotic freeness of second order is much stronger than having almost sure asymptotic freeness (of first order). According to Remark 19, we can guarantee the latter by the existence of a second order limit plus freeness of first order in the limit. Having also freeness of second order in the limit makes a much more precise statement on the asymptotic structure of the covariances.

In Example 23, we showed that several independent GUE random matrices are asymptotically free of second order. The same is also true if we include deterministic matrices. This follows from the explicit description in Theorem 13 of the second order limit in this case. We leave the proof of this as an exercise.

Theorem 27. *Let $\{X_1^{(N)},\dots,X_s^{(N)}\}_N$ be s independent* GUE*s, and, in addition, let $\{D_1^{(N)},\dots,D_t^{(N)}\}_N$ be t deterministic matrices with limiting distribution. Then $X_1^{(N)},\dots,X_s^{(N)},\{D_1^{(N)},\dots,D_t^{(N)}\}$ are asymptotically free of second order.*

Exercise 6. Prove Theorem 27 by using the explicit formula for the second order limit distribution given in Theorem 13.

Exercise 7. Show that Theorem 27 remains also true if the deterministic matrices are replaced by random matrices which are independent from the GUE's and which have a second order limit distribution. For this, upgrade first Theorem 16 to a situation where $\{D_1^{(N)},\dots,D_{p+q}^{(N)}\}$ have a second order limit distribution.

As in the first order case, one can also show that Haar unitary random matrices are asymptotically free of second order from deterministic matrices and, more generally, from random matrices which have a second order limit distribution and which are independent from the Haar unitary random matrices; this can then be used to deduce the asymptotic freeness of second order between unitarily invariant ensembles. The calculations in the Haar case rely again on properties of the Weingarten functions and get a bit technical. Here we only state the results; we refer to [131] for the details of the proof.

Definition 28. Let $B_1,\dots,B_t$ be t $N\times N$ random matrices with entries $b_{ij}^{(k)}$ ($k=1,\dots,t;\ i,j=1,\dots,N$). Let $U\in\mathcal{U}_N$ be unitary and $UB_kU^*=(\tilde b_{ij}^{(k)})_{i,j=1}^N$. If the joint distribution of all entries $\{b_{ij}^{(k)}\mid k=1,\dots,t;\ i,j=1,\dots,N\}$ is, for each $U\in\mathcal{U}_N$, the same as the joint distribution of all entries in the conjugated matrices $\{\tilde b_{ij}^{(k)}\mid k=1,\dots,t;\ i,j=1,\dots,N\}$, then we say that the joint distribution of the entries of $B_1,\dots,B_t$ is *unitarily invariant*.

Theorem 29. *Let $\{A_1^{(N)}, \dots, A_s^{(N)}\}_N$ and $\{B_1^{(N)}, \dots, B_t^{(N)}\}_N$ be two ensembles of random matrices such that:*

- *for each N, all entries of $A_1^{(N)}, \dots, A_s^{(N)}$ are independent from all entries of $B_1^{(N)}, \dots, B_t^{(N)}$*
- *for each N, the joint distribution of the entries of $B_1^{(N)}, \dots, B_t^{(N)}$ is unitarily invariant*
- *each of $(A_1^{(N)}, \dots, A_s^{(N)})_N$ and $(B_1^{(N)}, \dots, B_t^{(N)})_N$ has a second order limiting distribution.*

Then $\{A_1^{(N)}, \dots, A_s^{(N)}\}_N$ and $\{B_1^{(N)}, \dots, B_t^{(N)}\}_N$ are asymptotically free of second order.

5.4 Second order cumulants

In the context of usual (first order) freeness, it was advantageous to go over from moments to cumulants – the latter were easier to use to detect freeness, by the characterization of the vanishing of mixed cumulants. In the same spirit, we will now try to express also the fluctuations φ_2 in terms of cumulants. The following theory of second order cumulants was developed in [60]. Let us reconsider our combinatorial description of φ_2 for two of our main examples. In the case of a second order semi-circular element (i.e. for the limit of GUE random matrices, see Example 21), we have

$$\varphi_2(s^m, s^n) = \#\big(NC_2(m,n)\big) = \sum_{\pi \in NC_2(m,n)} 1 = \sum_{\pi \in S_{NC}(m,n)} \kappa_\pi. \tag{5.12}$$

The latter form comes from the fact that the free cumulants κ_n for semi-circulars are 1 for $n = 2$ and zero otherwise, i.e. κ_π is 1 for a non-crossing pairing and zero otherwise. For the second order free Poisson (i.e. for the limit of Wishart random matrices, see Exercise 3), we have

$$\varphi_2(x^m, x^n) = \sum_{\pi \in S_{NC}(m,n)} c^{\#(\pi)} = \sum_{\pi \in S_{NC}(m,n)} \kappa_\pi. \tag{5.13}$$

The latter form comes here from the fact that the free cumulants for a free Poisson are all equal to c. So in both cases, the value of φ_2 is expressed as a sum over the annular versions of non-crossing partitions, and each such permutation π is weighted by a factor κ_π, which is given by the product of first order cumulants, one factor κ_r for each cycle of π of length r. This is essentially the same formula as for φ, the only difference is that we sum over annular permutations instead over circle partitions. However, it turns out that in general the term

$$\sum_{\pi \in S_{NC}(m,n)} \kappa_\pi(a_1, \dots, a_m, a_{m+1}, \dots, a_{m+n})$$

is only one part of $\varphi_2(a_1 \cdots a_m, a_{m+1} \cdots a_{m+n})$; there will also be another contribution which involves genuine “second order cumulants”.

To see that we need in general such an additional contribution, let us rewrite the expression from Exercise 5 for $\varphi_2(a_1b_1, a_2b_2)$, for $\{a_1, a_2\}$ and $\{b_1, b_2\}$ being free of second order, in terms of first order cumulants.

$$\varphi_2(a_1b_1, a_2b_2) = \kappa_2(a_1, a_2)\kappa_2(b_1, b_2) + \kappa_2(a_1, a_2)\kappa_1(b_1)\kappa_1(b_2) \\ + \kappa_1(a_1)\kappa_1(a_2)\kappa_2(b_1, b_2) + \text{something else.}$$

The three displayed terms are the three non-vanishing terms κ_π for $\pi \in S_{NC}(2,2)$ (there are of course more such π, but they do not contribute because of the vanishing of mixed cumulants in free variables). But we have some additional contributions which we write in the form

$$\text{something else} = \kappa_{1,1}(a_1, a_2)\kappa_1(b_1)\kappa_1(b_2) + \kappa_1(a_1)\kappa_1(a_2)\kappa_{1,1}(b_1, b_2)$$

where we have set

$$\kappa_{1,1}(a_1, a_2) := \varphi_2(a_1, a_2) - \kappa_2(a_1, a_2).$$

The general structure of the additional terms is the following. We have second order cumulants $\kappa_{m,n}$ which have as arguments m elements from the first circle and n elements from the second circle. As one already sees in the above simple example, one only has summands which contain at most one such second order cumulant as factor. All the other factors are first order cumulants. So these terms can also be written as κ_σ, but now σ is of the form $\sigma = \pi_1 \times \pi_2 \in NC(m) \times NC(n)$ where one block of σ_1 and one block of σ_2 is marked. The two marked blocks go together as arguments into a second order cumulant; all the other blocks give just first order cumulants. Let us make this more rigorous in the following definition.

Definition 30. The *second order non-crossing annular partitions* $[NC(m) \times NC(n)]$ consist of elements $\sigma = (\pi_1, W_1) \times (\pi_2, W_2)$, where $\pi_1 \times \pi_2 \in NC(m) \times NC(n)$ and where $W_1 \in \pi_1$ and $W_2 \in \pi_2$. The blocks W_1 and W_2 are designated as *marked blocks* of π_1 and π_2, respectively. In examples, we will often mark those blocks as boldface (Fig. 5.4).

Definition 31. Let $(\mathcal{A}, \varphi, \varphi_2)$ be a second order probability space. The *second order cumulants*

$$\kappa_{m,n} : \mathcal{A}^m \times \mathcal{A}^n \to \mathbb{C}$$

are $m + n$-linear functionals on $\mathcal{A}$, where we distinguish the group of the first m arguments from the group of the last n arguments. Those second order cumulants are implicitly defined by the following moment-cumulant formula.

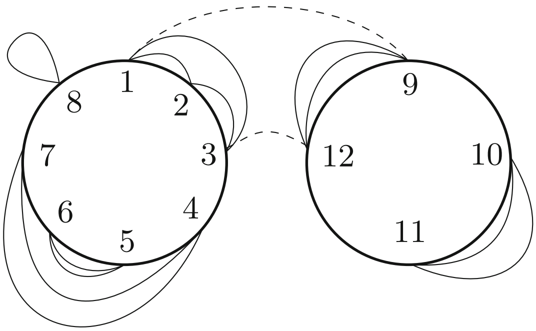


Fig. 5.4 The second order non-crossing annular partition $\sigma = \{(\mathbf{1},\mathbf{2},\mathbf{3}),(4,7),(5,6),(8)\} \times \{(\mathbf{9},\mathbf{12}),(10,11)\}$. Its contribution in the moment-cumulant formula is $\kappa_\sigma(a_1,\ldots,a_{12}) = \kappa_{3,2}(a_1,a_2,a_3,a_9,a_{12})\kappa_2(a_4,a_7)\kappa_2(a_5,a_6)\kappa_1(a_8)\kappa_2(a_{10},a_{11})$

$$\varphi_2(a_1\cdots a_m, a_{m+1}\cdots a_{m+n}) = \sum_{\pi\in S_{NC}(m,n)} \kappa_\pi(a_1,\ldots,a_{m+n}) + \sum_{\sigma\in[NC(m)\times NC(n)]} \kappa_\sigma(a_1,\ldots,a_{m+n}). \tag{5.14}$$

Here we have used the following notation. For a $\pi = \{V_1,\ldots,V_r\} \in S_{NC}(m,n)$, we put

$$\kappa_\pi(a_1,\ldots,a_{m+n}) := \prod_{i=1}^{r} \kappa_{\#(V_i)}\big((a_k)_{k\in V_i}\big),$$

where the κ_n are the already defined first order cumulants in the probability space $(\mathcal{A},\varphi)$. For a $\sigma \in [NC(m)\times NC(n)]$, we define κ_σ as follows. If $\sigma = (\pi_1, W_1)\times(\pi_2, W_2)$ is of the form $\pi_1 = \{W_1, V_1,\ldots,V_r\} \in NC(m)$ and $\pi_2 = \{W_2, \tilde{V}_1,\ldots,\tilde{V}_s\} \in NC(n)$, where W_1 and W_2 are the two marked blocks, then

$$\kappa_\sigma(a_1,\ldots,a_{m+n}) := \prod_{i=1}^{r} \kappa_{\#(V_i)}\big((a_k)_{k\in V_i}\big) \cdot \prod_{j=1}^{s} \kappa_{\#(\tilde{V}_j)}\big((a_l)_{l\in \tilde{V}_j}\big) \cdot \kappa_{\#(W_1),\#(W_2)}\big((a_u)_{u\in W_1}, (a_v)_{v\in W_2}\big).$$

The first sum only involves first order cumulants, and in the second sum, each term is a product of one second order cumulant and some first order cumulants. Thus, since we already know all first order cumulants, the first sum is totally determined in terms of moments of φ. The second sum, on the other side, contains exactly the highest order term $\kappa_{m,n}(a_1,\ldots,a_{m+n})$ and some lower order cumulants. Thus, by recursion, we can again solve the moment-cumulant formulas for the determination of $\kappa_{m,n}(a_1,\ldots,a_{m+n})$.

Example 32. 1) For $m = n = 1$, we have one first order contribution

$$\pi = (1,2) \in S_{NC}(1,1)$$

and one second order contribution

$$\sigma = \{(\mathbf{1})\} \times \{(\mathbf{2})\} \in [NC(1) \times NC(1)],$$

and thus we get

$$\varphi_2(a_1, a_2) = \kappa_\pi(a_1, a_2) + \kappa_\sigma(a_1, a_2) = \kappa_2(a_1, a_2) + \kappa_{1,1}(a_1, a_2).$$

By invoking the definition of κ_2, $\kappa_2(a_1, a_2) = \varphi(a_1 a_2) - \varphi(a_1)\varphi(a_2)$, this can be solved for $\kappa_{1,1}$ in terms of moments with respect to φ and φ_2:

$$\kappa_{1,1}(a_1, a_2) = \varphi_2(a_1, a_2) - \varphi(a_1 a_2) + \varphi(a_1)\varphi(a_2). \tag{5.15}$$

2) For $m = 2$ and $n = 1$, we have four first order contributions in $S_{NC}(2, 1)$,

$$\pi_1 = (1, 2, 3), \quad \pi_2 = (2, 1, 3), \quad \pi_3 = (1, 3)(2), \quad \pi_4 = (1)(2, 3)$$

and three second order contributions in $[NC(2) \times NC(1)]$,

$$\sigma_1 = \{(\mathbf{1}, \mathbf{2})\} \times \{(\mathbf{3})\}, \quad \sigma_2 = \{(\mathbf{1}), (2)\} \times \{(\mathbf{3})\}, \quad \sigma_3 = \{(1), (\mathbf{2})\} \times \{(\mathbf{3})\},$$

resulting in

$$\begin{aligned}\varphi_2(a_1 a_2, a_3) = {} & \kappa_3(a_1, a_2, a_3) + \kappa_3(a_2, a_1, a_3) + \kappa_2(a_1, a_3)\kappa_1(a_2) + \kappa_2(a_2, a_3)\kappa_1(a_1) \\ & + \kappa_{2,1}(a_1, a_2, a_3) + \kappa_{1,1}(a_1, a_3)\kappa_1(a_2) + \kappa_{1,1}(a_2, a_3)\kappa_1(a_1).\end{aligned}$$

By using the known formulas for κ_1, κ_2, κ_3, and the formula for $\kappa_{1,1}$ from above, this can be solved for $\kappa_{2,1}$:

$$\begin{aligned}\kappa_{2,1}(a_1, a_2, a_3) = {} & \varphi_2(a_1 a_2, a_3) - \varphi(a_1)\varphi_2(a_2, a_3) - \varphi(a_2)\varphi_2(a_1, a_3) \\ & - \varphi(a_1 a_2 a_3) - \varphi(a_1 a_3 a_2) + 2\varphi(a_1)\varphi(a_2 a_3) \\ & + 2\varphi(a_1 a_3)\varphi(a_2) + 2\varphi(a_1 a_2)\varphi(a_3) - 4\varphi(a_1)\varphi(a_2)\varphi(a_3).\end{aligned}$$

Example 33. 1) Let s be a second order semi-circular element, i.e. the second order limit of GUE random matrices, with second order distribution as described in Example 21. Then the second order cumulants all vanish in this case, i.e. we have for all $m, n \in \mathbb{N}$

$$\kappa_n(s, \ldots, s) = \delta_{n2} \qquad \text{and} \qquad \kappa_{m,n}(s, \ldots, s) = 0.$$

2) For the second order limit y of Wishart random matrices of parameter c (i.e. for a second order free Poisson element), it follows from Exercise 3 that again

all second order cumulants vanish and the distribution of y can be described as follows: for all $m, n \in \mathbb{N}$, we have

$$\kappa_n(y, \ldots, y) = c, \qquad \text{and} \qquad \kappa_{m,n}(y, \ldots, y) = 0.$$

3) For an example with non-vanishing second order cumulants, let us consider the square $a := s^2$ of the variable s from above. Then, by Equation (5.15), we have

$$\begin{aligned}\kappa_{1,1}(s^2, s^2) = \kappa_{1,1}(a, a) &= \varphi_2(a, a) - \varphi(aa) + \varphi(a)\varphi(a) \\ &= \varphi_2(s^2, s^2) - \varphi(s^2 s^2) + \varphi(s^2)\varphi(s^2) = 2 - 2 + 1 = 1.\end{aligned}$$

Exercise 8. Let $X_N = 1/\sqrt{N}\big(x_{ij}\big)_{i,j=1}^N$ be a Wigner random matrix ensemble, where $x_{ij} = x_{ji}$ for all i, j; all x_{ij} for $i \geq j$ are independent; all diagonal entries x_{ii} are identically distributed according to a distribution ν; and all off-diagonal entries x_{ij}, for $i \neq j$, are identically distributed according to a distribution μ. Show that $\{X_N\}_N$ has a second order limit $x \in (\mathcal{A}, \varphi, \varphi_2)$ which is in terms of cumulants given by: all first order cumulants are zero but $\kappa_2^x = k_2^\mu$; all second order cumulants are zero but $\kappa_{2,2}^x = k_4^\mu$, where k_2^μ and k_4^μ are the second and fourth classical cumulant of μ, respectively.

The usefulness of the notion of second order cumulants comes from the following second order analogue of the characterization of freeness by the vanishing of mixed cumulants.

Theorem 34. *Let $(\mathcal{A}, \varphi, \varphi_2)$ be a second order probability space. Consider unital subalgebras $\mathcal{A}_1, \ldots, \mathcal{A}_s \subset \mathcal{A}$. Then the following statements are equivalent:*

(*i*) *The algebras $\mathcal{A}_1, \ldots, \mathcal{A}_s$ are free of second order.*

(*ii*) *Mixed cumulants, both of first and second order, of the subalgebras vanish:*

- *whenever we choose, for $n \in \mathbb{N}$, $a_j \in \mathcal{A}_{i_j}$ $(j = 1, \ldots, n)$ in such a way that $i_k \neq i_l$ for some $k, l \in [n]$, then the corresponding first order cumulants vanish, $\kappa_n(a_1, \ldots, a_n) = 0$;*
- *and whenever we choose, for $m, n \in \mathbb{N}$, $a_j \in \mathcal{A}_{i_j}$ $(j = 1, \ldots, m + n)$ in such a way that $i_k \neq i_l$ for some $k, l \in [m + n]$, then the corresponding second order cumulants vanish, $\kappa_{m,n}(a_1, \ldots, a_{m+n}) = 0$.*

Sketch Let us give a sketch of the proof. The statement about the first order cumulants is just Theorem 2.14.

That the vanishing of mixed cumulants implies second order freeness follows quite easily from the moment-cumulant formula. In the case of cyclically alternating centred arguments, the only remaining contributions are given by spoke diagrams, and then the moment-cumulant formula (5.14) reduces to the defining formula (5.11) of second order freeness.

For the other direction, note first that second order freeness implies the vanishing of $\kappa_{m,n}(a_1,\dots,a_{m+n}) = 0$ whenever all the a_i are centred and both groups of arguments are cyclically alternating, i.e. $i_1 \neq i_2 \neq \cdots \neq i_m \neq i_1$ and $i_{m+1} \neq i_{m+2} \neq \cdots \neq i_{m+n} \neq i_{m+1}$. Next, because centring does not change the value of second order cumulants, we can drop the assumption of centredness. For also getting rid of the assumption that neighbours must be from different algebras, one has, as in the first order case (see Theorem 3.14), to invoke a formula for second order cumulants which have products as arguments. □

In the following theorem, we state the formula for the $\kappa_{m,n}$ with products as arguments. For the proof, we refer to [132].

Theorem 35. *Suppose $n_1,\dots,n_r,n_{r+1},\dots,n_{r+s}$ are positive integers, $m := n_1 + \cdots + n_r$, $n = n_{r+1} + \cdots + n_{r+s}$. Given a second order probability space $(\mathcal{A},\varphi,\varphi_2)$ and $a_1,a_2,\dots,a_{m+n} \in \mathcal{A}$, let*

$$A_1 = a_1 \cdots a_{n_1}, \quad A_2 = a_{n_1+1} \cdots a_{n_1+n_2}, \quad \dots, \quad A_{r+s} = a_{n_1+\cdots+n_{r+s-1}+1} \cdots a_{m+n}.$$

Then

$$\kappa_{r,s}(A_1,\dots,A_r,A_{r+1},\dots,A_{r+s}) = \sum_{\pi \in S_{NC}(m,n) \text{ with} \dots} \kappa_\pi(a_1,\dots,a_{m+n}) + \sum_{\sigma \in [NC(m)\times NC(n)] \text{ with} \dots} \kappa_\sigma(a_1,\dots,a_{m+n}), \tag{5.16}$$

where the summation is over

(*i*) *those $\sigma = (\pi_1,W_1)\times(\pi_2,W_2) \in [NC(m)\times NC(n)]$ where π_1 connects on one circle the groups corresponding to $A_1,\dots,A_r$ and π_2 connects on the other circle the groups corresponding to $A_{r+1},\dots,A_{r+s}$, where "connecting" is here used in the same sense as in the first order case (see Theorem 2.13). More precisely, this means that*

$$\pi_1 \vee \{(1,\dots,n_1),\dots,(n_1+\cdots+n_{r-1}+1,\dots,m)\} = 1_m$$

and that

$$\pi_2 \vee \{(m+1,\dots,m+n_{r+1}),\dots,(m+n_{r+1}+\cdots+n_{r+s-1}+1,\dots,m+n)\} = 1_n;$$

note that the marked blocks do not play any role for this condition.

(*ii*) *those $\pi \in S_{NC}(m,n)$ which connect the groups corresponding to all A_i on both circles in the following annular way: for such a π, all the groups must be connected, but it is not possible to cut the annulus open by cutting on each of the two circles between two groups.*

Example 36. 1) Let us reconsider the second order cumulant $\kappa_{1,1}(A_1,A_2)$ for $A_1 = A_2 = s^2$ from Example 33, by calculating it via the above theorem. Since all second order cumulants and all but the second first order cumulants of s are zero, in the formula (5.16), there is no contributing σ, and the only two possible π's are $\pi_1 = \{(1,3),(2,4)\}$ and $\pi_2 = \{(1,4),(2,3)\}$. Both connect both groups (a_1, a_2) and (a_3, a_4), but whereas π_1 does this in an annular way, in the case of π_2, the annulus could be cut open outside these groups. So π_1 contributes and π_2 does not. Hence,

$$\kappa_{1,1}(s^2, s^2) = \kappa_{\pi_1}(s,s,s,s) = \kappa_2(s,s)\kappa_2(s,s) = 1,$$

which agrees with the more direct calculation in Example 33.

2) Consider for general random variables a_1, a_2, a_3 the cumulant $\kappa_{1,1}(a_1a_2, a_3)$. The only contributing annular permutation in Equation (5.16) is $\pi = (1,3,2)$ (note that $(1,2,3)$ connects the two groups (a_1, a_2) and a_3, but not in an annular way), whereas all second order annular partitions in $[NC(2) \times NC(1)]$, namely,

$$\sigma_1 = \{(\mathbf{1},\mathbf{2})\} \times \{(\mathbf{3})\}, \quad \sigma_2 = \{(\mathbf{1}),(2)\} \times \{(\mathbf{3})\}, \quad \sigma_3 = \{(1),(\mathbf{2})\} \times \{(\mathbf{3})\},$$

are permitted and thus we get

$$\begin{aligned}\kappa_{1,1}(a_1a_2, a_3) &= \kappa_\pi(a_1,a_2,a_3) + \sum_{i=1}^{3} \kappa_{\sigma_i}(a_1,a_2,a_3)\\ &= \kappa_3(a_1,a_3,a_2) + \kappa_{2,1}(a_1,a_2,a_3)\\ &\quad + \kappa_{1,1}(a_1,a_3)\kappa_1(a_2) + \kappa_{1,1}(a_2,a_3)\kappa_1(a_1).\end{aligned}$$

As in the first order case, one can, with the help of this product formula, also get a version of the characterization of freeness in terms of vanishing of mixed cumulants for random variables instead of subalgebras.

Theorem 37. *Let $(\mathcal{A}, \varphi, \varphi_2)$ be a second order probability space. Consider $a_1, \ldots, a_s \in \mathcal{A}$. Then the following statements are equivalent:*

(*i*) *The variables $a_1, \ldots, a_s$ are free of second order.*
(*ii*) *Mixed cumulants, both of first and second order, of the variables vanish, i.e. $\kappa_n(a_{i_1}, \ldots, a_{i_n}) = 0$ and $\kappa_{m,n}(a_{i_1}, \ldots, a_{i_{m+n}}) = 0$ for all $m, n \in \mathbb{N}$ and all $1 \le i_k \le s$ (for all relevant k) and such that among the variables there are at least two different ones: there exist k, l such that $i_k \ne i_l$.*

Exercise 9. The main point in reducing this theorem to the version for subalgebras consists in using the product formula to show that the vanishing of mixed cumulants in the variables implies also the vanishing of mixed cumulants in elements in the generated subalgebras. As an example of this, show that the vanishing of all mixed first and second order cumulants in a_1 and a_2 implies also the vanishing of the mixed cumulants $\kappa_{2,1}(a_1^3, a_1, a_2^2)$ and $\kappa_{1,2}(a_1^3, a_1, a_2^2)$.

5.5 Functional relation between second order moment and cumulant series

Let us now consider the situation where all our random variables are the same, $a_1 = \cdots = a_{m+n} = a$. Then we write as before for the first order quantities $\alpha_n := \varphi(a^n)$ and $\kappa_n^a := \kappa_n(a, \ldots, a)$ and on second order level $\alpha_{m,n} := \varphi_2(a^m, a^n)$ and $\kappa_{m,n}^a := \kappa_{m,n}(a, \ldots, a)$. The vanishing of mixed cumulants for free variables gives then again that our cumulants linearize the addition of free variables.

Theorem 38. *Let $(\mathcal{A}, \varphi, \varphi_2)$ be a second order probability space, and let $a, b \in \mathcal{A}$ be free of second order. Then we have for all $m, n \in \mathbb{N}$*

$$\kappa_n^{a+b} = \kappa_n^a + \kappa_n^b \qquad \text{and} \qquad \kappa_{m,n}^{a+b} = \kappa_{m,n}^a + \kappa_{m,n}^b. \tag{5.17}$$

As in the first order case, one can translate the combinatorial relation between moments and cumulants into a functional relation between generating power series. In the following theorem, we give this as a relation between the corresponding Cauchy and R-transforms. Again, we refer to [60] for the proof and more details.

Theorem 39. *The moment-cumulant relations*

$$\alpha_n = \sum_{\pi \in NC(n)} \kappa_\pi \qquad \text{and} \qquad \alpha_{m,n} = \sum_{\pi \in S_{NC}(m,n)} \kappa_\pi + \sum_{\sigma \in [NC(m) \times NC(n)]} \kappa_\sigma$$

are equivalent to the functional relations

$$\frac{1}{G(z)} + R(G(z)) = z \tag{5.18}$$

and

$$G(z,w) = G'(z)G'(w)R(G(z), G(w)) + \frac{\partial^2}{\partial z \partial w} \log\left(\frac{F(z) - F(w)}{z - w}\right) \tag{5.19}$$

between the following formal power series: the Cauchy transforms

$$G(z) = \frac{1}{z} \sum_{n \geq 0} \alpha_n z^{-n} \qquad \text{and} \qquad G(z,w) = \frac{1}{zw} \sum_{m,n \geq 1} \alpha_{m,n} z^{-m} w^{-n}$$

and the R-transforms

$$R(z) = \frac{1}{z} \sum_{n \geq 1} \kappa_n z^n \qquad \text{and} \qquad R(z,w) = \frac{1}{zw} \sum_{m,n \geq 1} \kappa_{m,n} z^m w^n;$$

and where $F(z) = 1/G(z)$.

Equation (5.19) *can also be written in the form*

$$G(z,w) = G'(z)G'(w)\left\{R(G(z),G(w)) + \frac{1}{(G(z)-G(w))^2}\right\} - \frac{1}{(z-w)^2}. \tag{5.20}$$

Equation (5.18) is just the well-known functional relation (2.27) from Chapter 2 between first order moments and cumulants. Equation (5.19) determines a sequence of equations relating the first and second order moments with the second order cumulants; if we also express the first order moments in terms of first order cumulants, then this corresponds to the moment-cumulant relation $\alpha_{m,n} = \sum_{\pi \in S_{NC}(m,n)} \kappa_\pi + \sum_{\sigma \in [NC(m)\times NC(n)]} \kappa_\sigma$.

Note that formally the second term on the right-hand side of (5.19) can also be written as

$$\frac{\partial^2}{\partial z \partial w} \log\left(\frac{F(z)-F(w)}{z-w}\right) = \frac{\partial^2}{\partial z \partial w} \log\left(\frac{G(w)-G(z)}{z-w}\right); \tag{5.21}$$

but since $(G(w)-G(z))/(z-w)$ has no constant term, the power series expansion of $\log[(G(w)-G(z))/(z-w)]$ is not well defined.

Below is a table, produced from (5.19), giving the first few equations:

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

Remark 40. Note that the Cauchy transforms can also be written as

$$G(z) = \lim_{N\to\infty} \mathrm{E}\left(\mathrm{tr}(\frac{1}{z-A_N})\right) = \varphi\left(\frac{1}{z-a}\right) \tag{5.22}$$

and

$$G(z,w) = \lim_{N\to\infty} \mathrm{cov}\left(\mathrm{Tr}(\frac{1}{z-A_N}), \mathrm{Tr}(\frac{1}{w-A_n})\right) = \varphi_2\left(\frac{1}{z-a}, \frac{1}{w-a}\right), \tag{5.23}$$

if A_N has a as second order limit distribution.

In the case where all the second order cumulants are zero, i.e. $R(z,w) = 0$, Equation (5.19) expresses the second order Cauchy transform in terms of the first order Cauchy transform:

$$\varphi_2\left(\frac{1}{z-a}, \frac{1}{w-a}\right) = G(z,w) = \frac{\partial^2}{\partial z \partial w} \log\left(\frac{F(z)-F(w)}{z-w}\right). \tag{5.24}$$

This applies then in particular to the GUE and Wishart random matrices; that in those cases the second order cumulants vanish follows from equations (5.12) and (5.13); see also Example 33. In the case of Wishart matrices equation (5.24) (in terms of $G(z)$ instead of $F(z)$, via (5.21)) was derived by Bai and Silverstein [14, 15].

However, there are also many important situations where the second order cumulants do not vanish, and we need the full version of (5.19) to understand the fluctuations. The following exercise gives an example for this.

Exercise 10. A circular element in first order is of the form

$$c := \frac{1}{\sqrt{2}}(s_1 + i s_2), \tag{5.25}$$

where s_1 and s_2 are free standard semi-circular elements; see Section 6.8. There we will also show that such a circular element is in $*$-distribution the limit of a complex Gaussian random matrix. Since the same arguments apply also to second order, we define a *circular element of second order* by the Equation (5.25), where now s_1 and s_2 are two semi-circular elements of second order which are also free of second order. This means in particular that all second order cumulants in c and c^* are zero.

We will in the following compare such a second order circular element c with a second order semi-circular element s as defined in Example 21:

(i) Show that the first order distribution of s^2 and cc^* is the same, namely, both are free Poisson elements of rate 1.
(ii) Show that the second order cumulants of s^2 do not vanish.
(iii) Show that the second order cumulants of cc^* are all zero; hence, cc^* is a second order free Poisson element, of rate 1.

This shows that whereas s^2 and cc^* are the same in first order, their second order distributions are different.

5.6 Diagonalization of fluctuations

Consider a sequence of random matrices $(A_1^{(N)}, \ldots, A_s^{(N)})_N$ which has a second order limit $(a_1, \ldots, a_s)$. Then we have for any polynomial $p \in \mathbb{C}\langle x_1, \ldots, x_s \rangle$ that the unnormalized trace of its centred version,

$$\mathrm{Tr}\big(p(A_1^{(N)}, \ldots, A_s^{(N)}) - \mathrm{E}(\mathrm{tr}(p(A_1^{(N)}, \ldots, A_s^{(N)})))\, 1_N\big),$$

converges to a Gaussian variable. Actually, all such traces of polynomials converge jointly to a Gaussian family (this is just the fact that we require in our definition of a second order limit distribution that all third and higher classical cumulants go to zero), and the limiting covariance between two such traces for p_1 and p_2 is given by $\varphi_2(p_1(a_1,\dots,a_s), p_2(a_1,\dots,a_s))$. Often, we have a kind of explicit formula (of a combinatorial nature) for the covariance between monomials in our variables; but only in very rare cases this covariance is diagonal in those monomials. (An important instance where this actually happens is the case of Haar unitary random matrices; see Example 22. Note that there we are dealing with the $*$-distribution and we are getting complex Gaussian distributions.) For a better understanding of the covariance, one usually wants to diagonalize it; this corresponds to going over to Gaussian variables which are independent.

In the case of one GUE random matrix, this diagonalization is one of the main statements in Theorem 1, which was the starting point of this chapter. In the following, we want to see how our description of second order distributions and freeness allows to understand this theorem and its multivariate generalizations.

5.6.1 Diagonalization in the one-matrix case

Let us first look on the one-variable situation. If all second order cumulants are zero (as, e.g for GUE or Wishart random matrices), so that our second order Cauchy transform is given by (5.24), then one can proceed as follows.

In order to extract from $G(z,w)$ some information about the covariance for arbitrary polynomials p_1 and p_2, we use Cauchy's integral formula to write

$$p_1(a) = \frac{1}{2\pi i}\int_{C_1} \frac{p_1(z)}{z-a}dz, \qquad p_2(a) = \frac{1}{2\pi i}\int_{C_2} \frac{p_2(w)}{w-a}dw,$$

where the contour integrals over C_1 and C_2 are in the complex plane around the spectrum of a. We are assuming that a is a bounded self-adjoint operator; thus, we have to integrate around sufficiently large portions of the real line. This gives then, by using Equation (5.24) and integration by parts,

$$\begin{aligned}
\varphi_2(p_1(a), p_2(a)) &= -\frac{1}{4\pi^2}\int_{C_1}\int_{C_2} p_1(z)p_2(w)\varphi_2\left(\frac{1}{z-x}, \frac{1}{w-x}\right)dzdw \\
&= -\frac{1}{4\pi^2}\int_{C_1}\int_{C_2} p_1(z)p_2(w)G(z,w)dzdw \\
&= -\frac{1}{4\pi^2}\int_{C_1}\int_{C_2} p_1(z)p_2(w)\frac{\partial^2}{\partial z\partial w}\log\left(\frac{F(z)-F(w)}{z-w}\right)dzdw \\
&= -\frac{1}{4\pi^2}\int_{C_1}\int_{C_2} p_1'(z)p_2'(w)\log\left(\frac{F(z)-F(w)}{z-w}\right)dzdw.
\end{aligned}$$

We choose now for C_1 and C_2 rectangles with height going to zero; hence, the integration over each of these contours goes to integrals over the real axis, one approaching the real line from above and the other approaching the real line from below. We denote the corresponding limits of $F(z)$, when z is approaching $x \in \mathbb{R}$ from above or from below, by $F(x^+)$ and $F(x^-)$, respectively. Since p_1' and p_2' are continuous at the real axis, we get

$$\varphi_2(p_1(a), p_2(a)) = -\frac{1}{4\pi^2} \int_{\mathbb{R}} \int_{\mathbb{R}} p_1'(x) p_2'(y) \left(\log\left(\frac{F(x^+) - F(y^+)}{x^+ - y^+} \right) \right.$$
$$\left. - \log\left(\frac{F(x^+) - F(y^-)}{x^+ - y^-} \right) - \log\left(\frac{F(x^-) - F(y^+)}{x^- - y^+} \right) + \log\left(\frac{F(x^-) - F(y^-)}{x^- - y^-} \right) \right) dxdy.$$

Note that one has for the reciprocal Cauchy transform $F(\bar{z}) = \overline{F(z)}$; hence, $F(x^-) = \overline{F(x^+)}$. Since the contributions of the denominators cancel, we get in the end

$$\varphi_2(p_1(a), p_2(a)) = -\frac{1}{4\pi^2} \int_{\mathbb{R}} \int_{\mathbb{R}} p_1'(x) p_2'(y) \log \left| \frac{F(x) - F(y)}{F(x) - \overline{F(y)}} \right|^2 dxdy, \tag{5.26}$$

where $F(x)$ denotes now the usual limit $F(x^+)$ coming from the complex upper half-plane.

The diagonalization of this bilinear form (5.26) depends on the actual form of

$$K(x, y) = -\frac{1}{4\pi^2} \log \left| \frac{F(x) - F(y)}{F(x) - \overline{F(y)}} \right|^2 = -\frac{1}{4\pi^2} \log \left| \frac{G(x) - G(y)}{G(x) - \overline{G(y)}} \right|^2. \tag{5.27}$$

Example 41. Consider the GUE case. Then G is the Cauchy transform of the semicircle

$$G(z) = \frac{z - \sqrt{z^2 - 4}}{2}, \qquad \text{thus} \qquad G(x) = \frac{x - i\sqrt{4 - x^2}}{2}.$$

Hence, we have

$$\begin{aligned} K(x, y) &= -\frac{1}{4\pi^2} \log \left| \frac{x - y - i(\sqrt{4 - x^2} - \sqrt{4 - y^2})}{x - y - i(\sqrt{4 - x^2} + \sqrt{4 - y^2})} \right|^2 \\ &= -\frac{1}{4\pi^2} \log \frac{(x - y)^2 + (\sqrt{4 - x^2} - \sqrt{4 - y^2})^2}{(x - y)^2 + (\sqrt{4 - x^2} + \sqrt{4 - y^2})^2} \\ &= -\frac{1}{4\pi^2} \log \frac{4 - xy - \sqrt{(4 - x^2)(4 - y^2)}}{4 - xy + \sqrt{(4 - x^2)(4 - y^2)}}. \end{aligned}$$

In order to relate this to Chebyshev polynomials, let us write $x = 2\cos\theta$ and $y = 2\cos\psi$. Then we have

$$\begin{aligned} K(x,y) &= -\frac{1}{4\pi^2}\log\frac{4(1-\cos\theta\cos\psi-\sin\theta\sin\psi)}{4(1-\cos\theta\cos\psi+\sin\theta\sin\psi)} \\ &= -\frac{1}{4\pi^2}\log\frac{1-\cos(\theta-\psi)}{1-\cos(\theta+\psi)} \\ &= -\frac{1}{4\pi^2}\log(1-\cos(\theta-\psi)) + \frac{1}{4\pi^2}\log(1-\cos(\theta+\psi)) \\ &= \frac{1}{2\pi^2}\sum_{n=1}^{\infty}\frac{1}{n}\left(\cos(n(\theta-\psi)) - \cos(n(\theta+\psi))\right) \\ &= \frac{1}{\pi^2}\sum_{n=1}^{\infty}\frac{1}{n}\sin(n\theta)\sin(n\psi). \end{aligned}$$

In the penultimate step, we have used the expansion (5.31) for $\log(1-\cos\theta)$ from the next exercise.

Similarly as $\cos(n\theta)$ is related to $x = 2\cos\theta$ via the Chebyshev polynomials C_n of the first kind, $\sin(n\theta)$ can be expressed in terms of x via the Chebyshev polynomials U_n of the second kind. Those are defined via

$$U_n(2\cos\theta) = \frac{\sin((n+1)\theta)}{\sin\theta}. \tag{5.28}$$

We will address some of its properties in Exercise 12.

We can then continue our calculation above as follows:

$$\begin{aligned} K(x,y) &= \frac{1}{\pi^2}\sum_{n=1}^{\infty}\frac{1}{n}U_{n-1}(x)\sin\theta\cdot U_{n-1}(y)\sin\psi \\ &= \sum_{n=1}^{\infty}\frac{1}{n}U_{n-1}(x)\frac{1}{2\pi}\sqrt{4-x^2}\cdot U_{n-1}(y)\frac{1}{2\pi}\sqrt{4-y^2}. \end{aligned}$$

We will now use the following two facts about Chebyshev polynomials:

- the Chebyshev polynomials of second kind are orthogonal polynomials with respect to the semi-circular distribution, i.e. for all $m, n \geq 0$

$$\int_{-2}^{+2} U_n(x)U_m(x)\frac{1}{2\pi}\sqrt{4-x^2}dx = \delta_{nm}; \tag{5.29}$$

- the two kinds of Chebyshev polynomials are related by differentiation,

$$C_n' = nU_{n-1} \qquad \text{for } n \geq 0. \tag{5.30}$$

Then we can recover Theorem 1 by checking that the covariance is diagonal for the Chebyshev polynomials of first kind:

$$\begin{aligned}
\varphi_2(C_n(a), C_m(a)) &= \int\int C_n'(x)C_m'(y)K(x,y)dxdy \\
&= \int_{-2}^{+2}\int_{-2}^{+2} nU_{n-1}(x)mU_{m-1}(y)\sum_{k=1}^{\infty}\frac{1}{k}U_{k-1}(x)\frac{1}{2\pi}\sqrt{4-x^2} \\
&\qquad\qquad \cdot U_{k-1}(y)\frac{1}{2\pi}\sqrt{4-y^2}dxdy \\
&= nm\sum_{k=1}^{\infty}\frac{1}{k}\left(\int_{-2}^{+2} U_{n-1}(x)U_{k-1}(x)\frac{1}{2\pi}\sqrt{4-x^2}dx\right) \\
&\qquad\qquad \times \left(\int_{-2}^{+2} U_{m-1}(y)U_{k-1}(y)\frac{1}{2\pi}\sqrt{4-y^2}dy\right) \\
&= nm\sum_{k=1}^{\infty}\frac{1}{k}\delta_{nk}\delta_{mk} \\
&= n\delta_{nm}.
\end{aligned}$$

Note that all our manipulations were formal and we did not address analytic issues, like the justification of the calculations concerning contour integrals. For this and also for extending the formula for the covariance beyond polynomial functions, one should consult the original literature, in particular [15, 104].

Exercise 11. Show the following expansion:

$$-\frac{1}{2}\log(1-1\cos\theta) = \sum_{n=1}^{\infty}\frac{1}{n}\cos(n\theta) + \frac{1}{2}\log 2. \tag{5.31}$$

Exercise 12. Let C_n and U_n be the Chebyshev polynomials, rescaled to the interval $[-2,+2]$, of the first and second kind, respectively (see also Notation 8.33 and subsequent exercises).

(*i*) Show that the definition of the Chebyshev polynomials via recurrence relations, as given in Notation 8.33, is equivalent to the definition via trigonometric functions, as given in the discussion following Theorem 1 and in Equation (5.28).

(*ii*) Show equations (5.29) and (5.30).

(*iii*) Show that the Chebyshev polynomials of first kind are orthogonal with respect to the arc-sine distribution, i.e. for all $n, m \geq 0$ with $(m,n) \neq (0,0)$, we have

$$\int_{-2}^{+2} C_n(x)C_m(y)\frac{dx}{\pi\sqrt{4-x^2}} = \delta_{nm}. \tag{5.32}$$

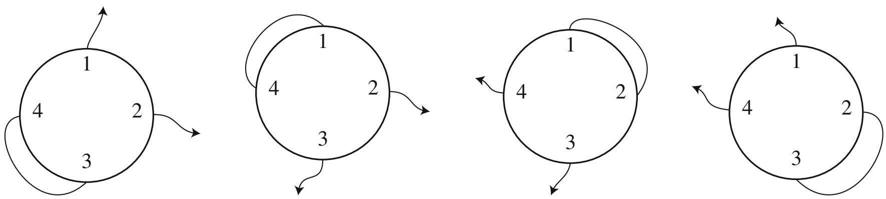


Fig. 5.5 The four non-crossing half-pairings on four points with two through strings are shown

Note that the definition for the case $n = 0$, $C_0 = 2$, is made in order to have it fit with the recurrence relations; to fit the orthonormality relations, $C_0 = 1$ would be the natural choice.

Example 42. By similar calculations as for the GUE, one can show that in the case of Wishart matrices, the diagonalization of the covariance (5.13) is achieved by going over to shifted Chebyshev polynomials of the first kind, $\sqrt{c^n} C_n\left((x-(1+c))/\sqrt{c}\right)$. This result is due to Cabanal-Duvillard [47]; see also [14, 115].

Remark 43. We want to address here a combinatorial interpretation of the fact that the Chebyshev polynomials C_k diagonalize the covariance for a GUE random matrix.

Let s be our second order semi-circular element; hence, $\varphi_2(s^m, s^n)$ is given by the number of annular non-crossing pairings on an (m,n) annulus. This is, of course, not diagonal in m and n because some points on each circle can be paired among themselves, and this pairing on both sides has no correlation; so there is no constraint that m has to be equal to n. However, a quantity which clearly must be the same for both circles is the number of *through-pairs*, i.e. pairs which connect both circles. Thus, in order to diagonalize the covariance, we should go over from the number of points on a circle to the number of through-pairs leaving this circle. A nice way to achieve this is to cut our diagrams in two parts – one part for each circle. These diagrams will be called *non-crossing annular half-pairings*. See Figs. 5.5 and 5.7. We will call what is left in a half-pairing of a through-pair after cutting an *open pair* – as opposed to *closed pairs* which live totally on one circle and are thus not affected by the cutting.

In this pictorial description, s^m corresponds to the sum over non-crossing annular half-pairings on one circle with m points, and s^n corresponds to a sum over non-crossing annular half-pairings on another circle with n points. Then $\varphi_2(s^m, s^n)$ corresponds to pairing the non-crossing annular half-pairings for s^m with the non-crossing annular half-pairings for s^n. A pairing of two non-crossing annular half-pairings consists of glueing together their open pairs in all possible planar ways. This clearly means that both non-crossing annular half-pairings must have the same number of open pairs, and thus our covariance should become diagonal if we go over from the number n of points on a circle to the number k of open pairs. Furthermore, there are clearly k possibilities to pair two sets of k open pairs in a planar way.

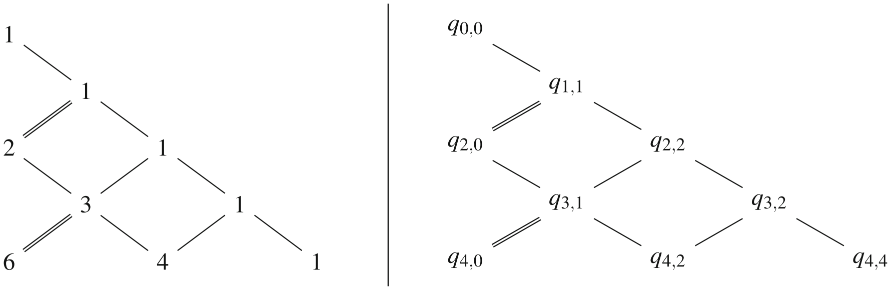


Fig. 5.6 As noted earlier, for the purpose of diagonalizing the fluctuations, the constant term of the polynomials is not important. If we make the small adjustment that $C_0(x) = 1$ and all the others are unchanged, then the recurrence relation becomes $C_{n+1}(x) = xC_n(x) - 2C_{n-1}(x)$ for $n \geq 2$ and $C_2(x) = xC_1(x) - 2C_0(x)$. From this we obtain $q_{n+1,k} = q_{n,k-1} + q_{n,k+1}$ for $k \geq 1$ and $q_{n+1,0} = 2q_{n,1}$. From these relations, we see that for $k \geq 1$ we have $q_{n,k} = \binom{n}{(n-k)/2}$ when $n-k$ is even and 0 when $n-k$ is odd. When $k = 0$, we have $q_{n,0} = 2\binom{n-1}{n/2-1}$ when n is even and $q_{n,0} = 0$ when n is odd

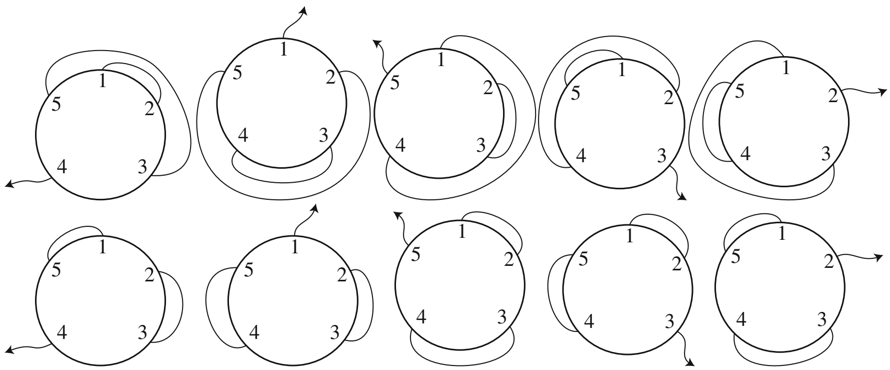


Fig. 5.7 When $n = 5$ and $k = 1$, $q_{5,1} = 10$. The ten non-crossing half-pairings on five points with one through string

From this point of view, the Chebyshev polynomials C_k should describe k open pairs. If we write x^n as a linear combination of the C_k, $x^n = \sum_{k=0}^{n} q_{n,k} C_k(x)$, then the above correspondence suggests that for $k > 0$, the coefficients $q_{n,k}$ are the number of non-crossing annular half-pairings of n points with k open pairs. See Fig. 5.6 and Fig. 5.7.

That this is indeed the correct combinatorial interpretation of the result of Johansson can be found in [115]. There the main emphasis is actually on the case of Wishart matrices and the result of Cabanal-Duvillard from Example 42. The Wishart case can be understood in a similar combinatorial way; instead of non-crossing annular half-pairings and through-pairs, one has to consider *non-crossing annular half-permutations* and *through-blocks*.

5.6.2 Diagonalization in the multivariate case

Consider now the situation of several variables; then we have to diagonalize the bilinear form $(p_1, p_2) \mapsto \varphi_2(p_1(a_1, \ldots, a_s), p_2(a_1, \ldots, a_s))$. For polynomials in just one of the variables, this is the same problem as in the previous section. It remains to understand the mixed fluctuations in more than one variable. If we have that $a_1, \ldots, a_s$ are free of second order, then this is fairly easy. The following theorem from [131] follows directly from Definition 24 of second order freeness.

Theorem 44. *Assume $a_1, \ldots, a_s$ are free of second order in the second order probability space $(\mathcal{A}, \varphi, \varphi_2)$. Let, for each $i = 1, \ldots, s$, $Q_k^{(i)}$ $(k \geq 0)$ be the orthogonal polynomials for the distribution of a_i; i.e. $Q_k^{(i)}$ is a polynomial of degree k such that $\varphi(Q_k^{(i)}(a_i)Q_l^{(i)}(a_i)) = \delta_{kl}$ for all $k, l \geq 0$. Then the fluctuations of mixed words in the a_i's are diagonalized by cyclically alternating products $Q_{k_1}^{(i_1)}(a_{i_1}) \cdots Q_{k_m}^{(i_m)}(a_{i_m})$ (with all $k_r \geq 1$ and $i_1 \neq i_2$, $i_2 \neq i_3$, $\ldots$, $i_m \neq i_1$), and the covariances are given by the number of cyclic matchings of these products:*

$$\begin{aligned} &\varphi_2\left(Q_{k_1}^{(i_1)}(a_{i_1}) \cdots Q_{k_m}^{(i_m)}(a_{i_m}), Q_{l_1}^{(j_1)}(a_{j_1}) \cdots Q_{l_n}^{(j_n)}(a_{j_n})\right) \\ &\qquad = \delta_{mn} \cdot \#\{r \in \{1, \ldots, n\} \mid i_s = j_{s+r}, k_s = l_{s+r}\ \forall s = 1, \ldots, n\}, \end{aligned} \tag{5.33}$$

where we count $s + r$ modulo n.

Remark 45. Note the different nature of the solution for the one-variate and the multivariate case. For example, for independent GUE's, we have that the covariance is diagonalized by the following set of polynomials:

- Chebyshev polynomials C_k of *first* kind in one of the variables
- cyclically alternating products of Chebyshev polynomials U_k of *second* kind for different variables.

Again there is a combinatorial way of understanding the appearance of the two different kinds of Chebyshev polynomials. As we have outlined in Remark 43, the Chebyshev polynomials C_k show up in the one-variate case, because this corresponds to going over to non-crossing annular half-pairings with k through-pairs. In the multivariate case, one has to realize that having several variables breaks the circular symmetry of the circle and thus effectively replaces a circular problem by a linear one. In this spirit, the expansion of x^n in terms of Chebyshev polynomials U_k of second kind counts the number of *non-crossing linear half-pairings* on n points with k open pairs.

In the Wishart case, there is a similar description by replacing non-crossing annular half-permutations by *non-crossing linear half-permutations*, resulting in an analogue appearance of orthogonal polynomials of first and second kind for the one-variate and multivariate situation, respectively.

More details and the proofs of the above statements can be found in [115].

Chapter 6
Free Group Factors and Freeness

The concept of freeness was actually introduced by Voiculescu in the context of operator algebras, more precisely, during his quest to understand the structure of special von Neumann algebras, related to free groups. We wish to recall here the relevant context and show how freeness shows up there very naturally and how it can provide some information about the structure of those von Neumann algebras.

Operator algebras are $*$- algebras of bounded operators on a Hilbert space which are closed in some canonical topologies. (C^*-algebras are closed in the operator norm, and von Neumann algebras are closed in the weak operator topology; the first topology is the operator version of uniform convergence, the latter of pointwise convergence.) Since the group algebra of a group can be represented on itself by bounded operators given by left multiplication (this is the regular representation of a group), one can take the closure in the appropriate topology of the group algebra and get thus C^*-algebras and von Neumann algebras corresponding to the group. The *group von Neumann algebra* arising from a group G in this way is usually denoted by $\mathcal{L}(G)$. This construction, which goes back to the foundational papers of Murray and von Neumann in the 1930s, is, for G an infinite discrete group, a source of important examples in von Neumann algebra theory, and much of the progress in von Neumann algebra theory was driven by the desire to understand the relation between groups and their von Neumann algebras better. The group algebra consists of finite sums over group elements; going over to a closure means that we allow also some infinite sums. One should note that the weak closure, in the case of infinite groups, is usually much larger than the group algebra, and it is very hard to control which infinite sums are added. Von Neumann algebras are quite large objects and their classification is notoriously difficult.

J.A. Mingo, R. Speicher, *Free Probability and Random Matrices*,
Fields Institute Monographs 35, DOI 10.1007/978-1-4939-6942-5_6

6.1 Group (von Neumann) algebras

Let G be a discrete group. We want to consider compactly supported continuous functions $a : G \to \mathbb{C}$, equipped with convolution $(a, b) \mapsto a * b$. Note that compactly supported means just finitely supported in the discrete case, and thus the set of such functions can be identified with the *group algebra* $\mathbb{C}[G]$ of formal finite linear combinations of elements in G with complex coefficients, $a = \sum_{g\in G} a(g)g$, where only finitely many $a(g) \neq 0$. Integration over such functions is with respect to the counting measure; hence, the convolution is then written as

$$a*b = \sum_{g\in G}(a*b)(g)g = \sum_{g\in G}\left(\sum_{h\in G} a(h)b(h^{-1}g)\right)g = \sum_{h\in G} a(h)h \sum_{k\in G} b(k)k = ab,$$

and is hence nothing but the multiplication in $\mathbb{C}[G]$. Note that the function $\delta_e = 1 \cdot e$ is the identity element in the group algebra $\mathbb{C}[G]$, where e is the identity element in G.

Now define an inner product on $\mathbb{C}[G]$ by setting

$$\langle g, h\rangle = \begin{cases} 1, & \text{if } g = h \\ 0, & \text{if } g \neq h \end{cases} \tag{6.1}$$

on G and extending sesquilinearly to $\mathbb{C}[G]$. From this inner product, we define the 2-norm on $\mathbb{C}[G]$ by $\|a\|_2^2 = \langle a, a\rangle$. In this way $(\mathbb{C}[G], \|\cdot\|_2)$ is a normed vector space. However, it is not complete in the case of infinite G (for finite G the following is trivial). The completion of $\mathbb{C}[G]$ with respect to $\|\cdot\|_2$ consists of all functions $a : G \to \mathbb{C}$ satisfying $\sum_{g\in G} |a(g)|^2 < \infty$ and is denoted by $\ell_2(G)$ and is a Hilbert space.

Now consider the unitary group representation $\lambda : G \to \mathcal{U}(\ell_2(G))$ defined by

$$\lambda(g) \cdot \sum_{h\in G} a(h)h := \sum_{h\in G} a(h)gh. \tag{6.2}$$

This is the *left regular representation* of G on the Hilbert space $\ell_2(G)$. It is obvious from the definition that each $\lambda(g)$ is an isometry of $\ell_2(G)$, but we want to check that it is in fact a unitary operator on $\ell_2(G)$. Since clearly $\langle gh, k\rangle = \langle h, g^{-1}k\rangle$, the adjoint of the operator $\lambda(g)$ is $\lambda(g^{-1})$. But then since λ is a group homomorphism, we have $\lambda(g)\lambda(g)^* = I = \lambda(g)^*\lambda(g)$, so that $\lambda(g)$ is indeed a unitary operator on $\ell_2(G)$.

Now extend the domain of λ from G to $\mathbb{C}[G]$ in the obvious way:

$$\lambda(a) = \lambda\left(\sum_{g\in G} a(g)g\right) = \sum_{g\in G} a(g)\lambda(g).$$

This makes λ into an algebra homomorphism $\lambda : \mathbb{C}[G] \to B(\ell_2(G))$, i.e. λ is a representation of the group algebra on $\ell_2(G)$. We define two new (closed) algebras via this representation. The *reduced group C^*-algebra* $C^*_{\text{red}}(G)$ of G is the closure of $\lambda(\mathbb{C}[G]) \subset B(\ell_2(G))$ in the operator norm topology. The *group von Neumann algebra of* G, denoted $\mathcal{L}(G)$, is the closure of $\lambda(\mathbb{C}[G])$ in the strong operator topology on $B(\ell_2(G))$.

One knows that for an infinite discrete group G, $\mathcal{L}(G)$ is a type II_1 von Neumann algebra, i.e. $\mathcal{L}(G)$ is infinite dimensional, but yet there is a trace τ on $\mathcal{L}(G)$ defined by $\tau(a) := \langle ae, e\rangle$ for $a \in \mathcal{L}(G)$, where $e \in G$ is the identity element. To see the trace property of τ, it suffices to check it for group elements; this extends then to the general situation by linearity and normality. However, for $g, h \in G$, the fact that $\tau(gh) = \tau(hg)$ is just the statement that $gh = e$ is equivalent to $hg = e$; this is clearly true in a group. The existence of a trace shows that $\mathcal{L}(G)$ is a proper subalgebra of $B(\ell_2(G))$; this is the case because there does not exist a trace on all bounded operators on an infinite dimensional Hilbert space. An easy fact is that if G is an ICC group, meaning that the conjugacy class of each $g \in G$ with $g \neq e$ has infinite cardinality, then $\mathcal{L}(G)$ is a factor, i.e. has trivial centre (see [106, Theorem 6.75]). Another fact is that if G is an amenable group (e.g. the infinite permutation group $S_\infty = \cup_n S_n$), then $\mathcal{L}(G)$ is the hyperfinite II_1 factor R.

Exercise 1.

(*i*) Show that $\mathcal{L}(G)$ is a factor if and only if G is an ICC group.
(*ii*) Show that the infinite permutation group $S_\infty = \cup_n S_n$ is ICC. (Note that each element from S_∞ moves only a finite number of elements.)

6.2 Free group factors

Now consider the case where $G = \mathbb{F}_n$, the *free group on n generators*; n can here be a natural number $n \geq 1$ or $n = \infty$. Let us briefly recall the definition of $\mathbb{F}_n$ and some of its properties. Consider the set of all words, of arbitrary length, over the $2n+1$-letter alphabet $\{a_1, a_2, \ldots, a_n, a_1^{-1}, a_2^{-1}, \ldots, a_n^{-1}\} \cup \{e\}$, where the letters of the alphabet satisfy no relations other than $ea_i = a_i e = a_i$, $ea_i^{-1} = a_i^{-1}e = a_i^{-1}$, $a_i^{-1}a_i = a_i a_i^{-1} = e$. We say that a word is *reduced* if its length cannot be reduced by applying one of the above relations. Then the set of all reduced words in this alphabet together with the binary operation of concatenating words and reducing constitutes the free group $\mathbb{F}_n$ on n generators. $\mathbb{F}_n$ is the group generated by n symbols satisfying no relations other than those required by the group axioms. Clearly $\mathbb{F}_1$ is isomorphic to the abelian group $\mathbb{Z}$, while $\mathbb{F}_n$ is non-abelian for $n > 1$ and in fact has trivial centre. The integer n is called the *rank* of the free group; it is fairly easy, though not totally trivial, to see (e.g. by reducing it via abelianization to a corresponding question about abelian free groups) that $\mathbb{F}_n$ and $\mathbb{F}_m$ are isomorphic if and only if $m = n$.

Exercise 2. Show that $\mathbb{F}_n$ is, for $n \geq 2$, an ICC group.

Since $\mathbb{F}_n$ has the infinite conjugacy class property, one knows that the group von Neumann algebra $\mathcal{L}(\mathbb{F}_n)$ is a II_1 factor, called a *free group factor*. Murray and von Neumann showed that $\mathcal{L}(\mathbb{F}_n)$ is not isomorphic to the hyperfinite factor, but otherwise nothing was known about the structure of these free group factors, when free probability was invented by Voiculescu to understand them better.

While as pointed out above we have that $\mathbb{F}_n \simeq \mathbb{F}_m$ if and only if $m = n$, the corresponding problem for the free group factors is still unknown; see however some results in this direction in section 6.12.

Free group factor isomorphism problem: Let $m, n \geq 2$ (possibly equal to ∞), $n \neq m$. Are the von Neumann algebras $\mathcal{L}(\mathbb{F}_n)$ and $\mathcal{L}(\mathbb{F}_m)$ isomorphic?

The corresponding problem for the reduced group C^*-algebras was solved by Pimsner and Voiculescu [143] in 1982: they showed that $C^*_{\mathrm{red}}(\mathbb{F}_n) \not\simeq C^*_{\mathrm{red}}(\mathbb{F}_m)$ for $m \neq n$.

6.3 Free product of groups

There is the notion of free product of groups. If G, H are groups, then their free product $G * H$ is defined to be the group whose generating set is the disjoint union of G and H and which has the property that the only relations in $G * H$ are those inherited from G and H and the identification of the neutral elements of G and H. That is, there should be no non-trivial algebraic relations between elements of G and elements of H in $G * H$. In a more abstract language, the free product is the coproduct in the category of groups. For example, in the category of groups, the n-fold direct product of n copies of $\mathbb{Z}$ is the lattice $\mathbb{Z}^n$; the n-fold coproduct (free product) of n copies of $\mathbb{Z}$ is the free group $\mathbb{F}_n$ on n generators.

In the category of groups, we can understand $\mathbb{F}_n$ via the decomposition $\mathbb{F}_n = \mathbb{Z} * \mathbb{Z} * \cdots * \mathbb{Z}$. Is there a similar *free product of von Neumann algebras* that will help us to understand the structure of $\mathcal{L}(\mathbb{F}_n)$? The notion of *freeness* or *free independence* makes this precise. In order to understand what it means for elements in $\mathcal{L}(G)$ to be *free*, we need to deal with infinite sums, so the algebraic notion of freeness will not do: we need a state.

6.4 Moments and isomorphism of von Neumann algebras

We will try to understand a von Neumann algebra with respect to a state. Let M be a von Neumann algebra and let $\varphi : M \to \mathbb{C}$ be a state defined on M, i.e. a positive linear functional. Select finitely many elements $a_1, \ldots, a_k \in M$. Let us first recall the notion of ($*$-)moments and ($*$-)distribution in such a context.

Definition 1. 1) The collection of numbers gotten by applying the state to words in the alphabet $\{a_1, \ldots, a_k\}$ is called the collection of *joint moments* of $a_1, \ldots, a_k$, or the *distribution* of $a_1, \ldots, a_k$.

2) The collection of numbers gotten by applying the state to words in the alphabet $\{a_1, \ldots, a_k, a_1^*, \ldots, a_k^*\}$ is called the collection of *joint* $*$*-moments* of $a_1, \ldots, a_k$, or the $*$*-distribution* of $a_1, \ldots, a_k$.

Theorem 2. *Let* $M = \mathrm{vN}(a_1, \ldots, a_k)$ *be generated as von Neumann algebra by elements* $a_1, \ldots, a_k$ *and let* $N = \mathrm{vN}(b_1, \ldots, b_k)$ *be generated as von Neumann algebra by elements* $b_1, \ldots, b_k$. *Let* $\varphi : M \to \mathbb{C}$ *and* $\psi : N \to \mathbb{C}$ *be faithful normal states. If* $a_1, \ldots, a_k$ *and* $b_1, \ldots, b_k$ *have the same* $*$*-distributions with respect to* φ *and* ψ*, respectively, then the map* $a_i \mapsto b_i$ *extends to a* $*$*-isomorphism of* M *and* N.

Exercise 3. Prove Theorem 2 by observing that the assumptions imply that the GNS-constructions with respect to φ and ψ are isomorphic.

Though the theorem is not hard to prove, it conveys the important message that all information about a von Neumann algebra is, in principle, contained in the *-moments of a generating set with respect to a faithful normal state.

In the case of the group von Neumann algebras $\mathcal{L}(G)$, the canonical state is the trace τ. This is defined as a vector state, so it is automatically normal. It is worth to notice that it is also faithful (and hence $(\mathcal{L}(G), \tau)$ is a tracial W^*-probability space).

Proposition 3. *The trace* τ *on* $\mathcal{L}(G)$ *is a faithful state.*

Proof: Suppose that $a \in \mathcal{L}(G)$ satisfies $0 = \tau(a^*a) = \langle a^*ae, e\rangle = \langle ae, ae\rangle$, thus $ae = 0$. So we have to show that $ae = 0$ implies $a = 0$. To show that $a = 0$, it suffices to show that $\langle a\xi, \eta\rangle = 0$ for any $\xi, \eta \in \ell_2(G)$. It suffices to consider vectors of the form $\xi = g, \eta = h$ for $g, h \in G$, since we can get the general case from this by linearity and continuity. Now, by using the traciality of τ, we have

$$\langle ag, h\rangle = \langle age, he\rangle = \langle h^{-1}age, e\rangle = \tau(h^{-1}ag) = \tau(gh^{-1}a) = \langle gh^{-1}ae, e\rangle = 0,$$

since the first argument to the last inner product is 0. □

6.5 Freeness in the free group factors

We now want to see that the algebraic notion of freeness of subgroups in a free product of groups translates with respect to the canonical trace τ to our notion of free independence.

Let us say that a product in an algebra $\mathcal{A}$ is *alternating* with respect to subalgebras $\mathcal{A}_1, \ldots, \mathcal{A}_s$ if adjacent factors come from different subalgebras. Recall that our definition of free independence says: the subalgebras $\mathcal{A}_1, \ldots, \mathcal{A}_s$ are free if any product in centred elements over these algebras which alternates is centred.

Proposition 4. *Let* G *be a group containing subgroups* $G_1, \ldots, G_s$ *such that* $G = G_1 * \cdots * G_s$. *Let* τ *be the state* $\tau(a) = \langle ae, e\rangle$ *on* $\mathbb{C}[G]$. *Then the subalgebras* $\mathbb{C}[G_1], \ldots, \mathbb{C}[G_s] \subset \mathbb{C}[G]$ *are free with respect to* τ.

Proof: Let $a_1a_2\cdots a_k$ be an element in $\mathbb{C}[G]$ which alternates with respect to the subalgebras $\mathbb{C}[G_1],\ldots,\mathbb{C}[G_s]$, and assume the factors of the product are centred with respect to τ. Since τ is the "coefficient of the identity" state, this means that if $a_j \in \mathbb{C}[G_{i_j}]$, then a_j looks like $a_j = \sum_{g\in G_{i_j}} a_j(g)g$ and $a_j(e) = 0$. Thus we have

$$\tau(a_1a_2\cdots a_k) = \sum_{g_1\in G_{i_1},\ldots,g_k\in G_{i_k}} a_1(g_1)a_2(g_2)\cdots a_k(g_k)\tau(g_1g_2\cdots g_k).$$

Now, $\tau(g_1g_2\cdots g_k) \neq 0$ only if $g_1g_2\cdots g_k = e$. But $g_1g_2\cdots g_k$ is an alternating product in G with respect to the subgroups $G_1, G_2,\ldots,G_s$, and since $G = G_1 * G_2 * \cdots * G_s$, this can happen only when at least one of the factors, let's say g_j, is equal to e; but in this case $a_j(g_j) = a_j(e) = 0$. So each summand in the sum for $\tau(a_1a_2\cdots a_k)$ vanishes, and we have $\tau(a_1a_2\cdots a_k) = 0$, as required. □

Thus freeness of the subgroup algebras $\mathbb{C}[G_1],\ldots,\mathbb{C}[G_s]$ with respect to τ is just a simple reformulation of the fact that $G_1,\ldots,G_s$ are free subgroups of G. However, a non-trivial fact is that this reformulation carries over to closures of the subalgebras.

Proposition 5. (1) *Let A be a C^*-algebra, $\varphi : A \to \mathbb{C}$ a state. Let $B_1,\ldots,B_s \subset A$ be unital $*$-subalgebras which are free with respect to φ. Put $A_i := \overline{B_i}^{\|\cdot\|}$, the norm closure of B_i. Then $A_1,\ldots,A_s$ are also free.*

(2) *Let M be a von Neumann algebra, $\varphi : M \to \mathbb{C}$ a normal state. Let $B_1,\ldots,B_s$ be unital $*$-subalgebras which are free. Put $M_i :=$ vN(B_i). Then $M_1,\ldots,M_s$ are also free.*

Proof: (1) Consider $a_1,\ldots,a_k$ with $a_i \in A_{j_i}$, $\varphi(a_i) = 0$, and $j_i \neq j_{i+1}$ for all i. We have to show that $\varphi(a_1\cdots a_k) = 0$. Since B_i is dense in A_i, we can, for each i, approximate a_i in operator norm by a sequence $(b_i^{(n)})_{n\in\mathbb{N}}$, with $b_i^{(n)} \in B_i$, for all n. Since we can replace $b_i^{(n)}$ by $b_i^{(n)} - \varphi(b_i^{(n)})$ (note that $\varphi(b_i^{(n)})$ converges to $\varphi(a_i) = 0$), we can assume, without restriction, that $\varphi(b_i^{(n)}) = 0$. But then we have

$$\varphi(a_1\cdots a_k) = \lim_{n\to\infty} \varphi(b_1^{(n)}\cdots b_k^{(n)}) = 0,$$

since, by the freeness of $B_1,\ldots,B_s$, we have $\varphi(b_1^{(n)}\cdots b_k^{(n)}) = 0$ for each n.

(2) Consider $a_1,\ldots,a_k$ with $a_i \in M_{j_i}$, $\varphi(a_i) = 0$, and $j_i \neq j_{i+1}$ for all i. We have to show that $\varphi(a_1\cdots a_k) = 0$. We approximate essentially as in the C^*-algebra case; we only have to take care that the multiplication of our k factors is still continuous in the appropriate topology. More precisely, we can now approximate, for each i, the operator a_i in the strong operator topology by a sequence (or a net, if you must) $b_i^{(n)}$. By invoking Kaplansky's density theorem, we can choose those such that we keep everything bounded, namely, $\|b_i^{(n)}\| \leq \|a_i\|$ for all n. Again we can centre the sequence, so that we can

assume that all $\varphi(b_i^{(n)}) = 0$. Since the joint multiplication is on bounded sets continuous in the strong operator topology, we have then still the convergence of $b_1^{(n)} \cdots b_k^{(n)}$ to $a_1 \cdots a_k$ and, thus, since φ is normal, also the convergence of $0 = \varphi(b_1^{(n)} \cdots b_k^{(n)})$ to $\varphi(a_1 \cdots a_k)$.

□

6.6 The structure of free group factors

What does this tell us for the free group factors? It is clear that each generator of the free group gives a Haar unitary element in $(\mathcal{L}(\mathbb{F}_n), \tau)$. By the discussion above, those elements are $*$-free. Thus the free group factor $\mathcal{L}(\mathbb{F}_n)$ is generated by n $*$-free Haar unitaries $u_1, \dots, u_n$. Note that, by Theorem 2, we will get the free group factor $\mathcal{L}(\mathbb{F}_n)$ whenever we find somewhere n Haar unitaries which are $*$-free with respect to a faithful normal state. Furthermore, since we are working inside von Neumann algebras, we have at our disposal measurable functional calculus, which means that we can also deform the Haar unitaries into other, possibly more suitable, generators.

Theorem 6. *Let M be a von Neumann algebra and τ a faithful normal state on M. Assume that $x_1, \dots, x_n \in M$ generate M,* $\mathrm{vN}(x_1, \dots, x_n) = M$ *and that*

- *$x_1, \dots, x_n$ are $*$-free with respect to τ,*
- *each x_i is normal, and its spectral measure with respect to τ is diffuse (i.e. has no atoms).*

Then $M \simeq \mathcal{L}(\mathbb{F}_n)$.

Proof: Let x be a normal element in M which is such that its spectral measure with respect to τ is diffuse. Let $A = \mathrm{vN}(x)$ be the von Neumann algebra generated by x. We want to show that there is a Haar unitary $u \in A$ that generates A as a von Neumann algebra. A is a commutative von Neumann algebra and the restriction of τ to A is a faithful state. A cannot have any minimal projections as that would mean that the spectral measure of x with respect to τ was not diffuse. Thus there is a normal $*$-isomorphism $\pi : A \to L^\infty[0,1]$ where we put Lebesgue measure on $[0,1]$. This follows from the well-known fact that any commutative von Neumann algebra is $*$-isomorphic to $L^\infty(\mu)$ for some measure μ and that all spaces $L^\infty(\mu)$ for μ without atoms are $*$-isomorphic; see, for example, [170, Chapter III, Theorem 1.22].

Under π the trace τ becomes a normal state on $L^\infty[0,1]$. Thus there is a positive function $h \in L^1[0,1]$ such that for all $a \in A$, $\tau(a) = \int_0^1 \pi(a)(t)h(t)\,dt$. Since τ is faithful, the set $\{t \in [0,1] \mid h(t) = 0\}$ has Lebesgue measure 0. Thus $H(s) = \int_0^s h(t)\,dt$ is a continuous positive strictly increasing function on $[0,1]$ with range $[0,1]$. So by the Stone-Weierstrass theorem, the C^*-algebra generated by 1 and H is all of $C[0,1]$. Hence the von Neumann algebra generated by 1 and H is all of $L^\infty[0,1]$. Let $v(t) = \exp(2\pi i H(t))$. Then H is in the von Neumann algebra generated by v, so the von Neumann algebra generated by v is $L^\infty[0,1]$. Also,

$$\int_0^1 v(t)^n h(t)\,dt = \int_0^1 \exp(2\pi i n H(t)) H'(t)\,dt = \int_0^1 e^{2\pi i n s}\,ds = \delta_{0,n}.$$

Thus v is Haar unitary with respect to h. Finally let $u \in A$ be such that $\pi(u) = v$. Then the von Neumann algebra generated by u is A and u is a Haar unitary with respect to the trace τ.

This means that for each i we can find in $\mathrm{vN}(x_i)$ a Haar unitary u_i which generates the same von Neumann algebra as x_i. By Proposition 5, freeness of the x_i goes over to freeness of the u_i. So we have found n Haar unitaries in M which are $*$-free and which generate M. Thus M is isomorphic to the free group factor $\mathcal{L}(\mathbb{F}_n)$. □

Example 7. Instead of generating $\mathcal{L}(\mathbb{F}_n)$ by n $*$-free Haar unitaries, it is also very common to use n free semi-circular elements. (Note that for self-adjoint elements $*$-freeness is of course the same as freeness.) This is of course covered by the theorem above. But let us be a bit more explicit on deforming a semi-circular element into a Haar unitary. Let $s \in M$ be a semi-circular operator. The spectral measure of s is $\sqrt{4-t^2}/(2\pi)\,dt$, i.e.

$$\tau(f(s)) = \frac{1}{2\pi}\int_{-2}^{2} f(t)\sqrt{4-t^2}\,dt.$$

If

$$H(t) = \frac{t}{4\pi}\sqrt{4-t^2} + \frac{1}{\pi}\sin^{-1}(t/2) \qquad \text{then} \qquad H'(t) = \frac{1}{2\pi}\sqrt{4-t^2},$$

and $u = \exp(2\pi i H(s))$ is a Haar unitary, i.e.

$$\tau(u^k) = \int_{-2}^{2} e^{2\pi i k H(t)} H'(t)\,dt = \int_{-1/2}^{1/2} e^{2\pi i k r}\,dr = \delta_{0,k},$$

which generates the same von Neumann subalgebra as s.

6.7 Compression of free group factors

Let M be any II_1 factor with faithful normal trace τ and e a projection in M. Let $eMe = \{exe \mid x \in M\}$; eMe is again a von Neumann algebra, actually a II_1 factor, with e being its unit, and it is called the *compression* of M by e. It is an elementary fact in von Neumann algebra theory that the isomorphism class of eMe depends only on $t = \tau(e)$, and we denote this isomorphism class by M_t. A deeper fact of Murray and von Neumann is that $(M_s)_t = M_{st}$. We can define M_t for all $t > 0$ as follows. For a positive integer n, let $M_n = M \otimes M_n(\mathbb{C})$, and for any t, let $M_t = eM_ne$ for any sufficiently large n and any projection e in M_n with trace t, where here we use the non-normalized trace $\tau \otimes \mathrm{Tr}$ on M_n. Murray and von

Neumann then defined the *fundamental group* of M, $\mathcal{F}(M)$, to be $\{t \in \mathbb{R}^+ \mid M \simeq M_t\}$ and showed that it is a multiplicative subgroup of $\mathbb{R}^+$. (See [106, Ex. 13.4.5 and 13.4.6].) It is a theorem that when G is an amenable ICC group, we have that $\mathcal{L}(G)$ is the hyperfinite II_1 factor and $\mathcal{F}(\mathcal{L}(G)) = \mathbb{R}^+$; see [170].

Rădulescu showed that $\mathcal{F}(\mathcal{L}(\mathbb{F}_\infty)) = \mathbb{R}^+$; see [144]. For finite n, $\mathcal{F}(\mathcal{L}(\mathbb{F}_n))$ is unknown; but it is known to be either $\mathbb{R}^+$ or $\{1\}$. In the rest of this chapter, we will give the key ideas about those compression results for free group factors.

The first crucial step was taken by Voiculescu who showed in 1990 in [179] that for integer m, n, k, we have $\mathcal{L}(\mathbb{F}_n)_{1/k} \simeq \mathcal{L}(\mathbb{F}_m)$, where $(m-1)/(n-1) = k^2$, or equivalently

$$\mathcal{L}(\mathbb{F}_n) \simeq M_k(\mathbb{C}) \otimes \mathcal{L}(\mathbb{F}_m), \qquad \text{where} \qquad \frac{m-1}{n-1} = k^2. \tag{6.3}$$

So if we embed $\mathcal{L}(\mathbb{F}_m)$ into $M_k(\mathbb{C}) \otimes \mathcal{L}(\mathbb{F}_m) \simeq \mathcal{L}(\mathbb{F}_n)$ as $x \mapsto 1 \otimes x$, then $\mathcal{L}(\mathbb{F}_m)$ is a subfactor of $\mathcal{L}(\mathbb{F}_n)$ of Jones index k^2; see [105, Example 2.3.1]. Thus, $(m-1)/(n-1) = [\mathcal{L}(\mathbb{F}_n) : \mathcal{L}(\mathbb{F}_m)]$. Notice the similarity to Schreier's index formula for free groups. Indeed, suppose G is a free group of rank n and H is a subgroup of G of finite index. Then H is necessarily a free group, say of rank m, and Schreier's index formula says that $(m-1)/(n-1) = [G : H]$.

Rather than proving Voiculescu's theorem, Equation (6.3), in full generality, we shall first prove a special case which illustrates the main ideas of the proof and then sketch the general case.

Theorem 8. *We have* $\mathcal{L}(\mathbb{F}_3)_{1/2} \simeq \mathcal{L}(\mathbb{F}_9)$.

To prove this theorem, we must find in $\mathcal{L}(\mathbb{F}_3)_{1/2}$ nine free normal elements with diffuse spectral measure which generate $\mathcal{L}(\mathbb{F}_3)_{1/2}$. In order to achieve this, we will start with normal elements x_1, x_2, x_3, together with a faithful normal state φ, such that

- the spectral measure of each x_i is diffuse (i.e. no atoms) and
- x_1, x_2, x_3 are $*$-free.

Let N be the von Neumann algebra generated by x_1, x_2, and x_3. Then $N \simeq \mathcal{L}(\mathbb{F}_3)$. We will then show that there is a projection p in N such that

- $\varphi(p) = 1/2$
- there are 9 free and diffuse elements in pNp which generate pNp.

Thus $\mathcal{L}(\mathbb{F}_3)_{1/2} \simeq pNp \simeq \mathcal{L}(\mathbb{F}_9)$.

The crucial issue above is that we will be able to choose our elements x_1, x_2, x_3 in such a form that we can easily recognize p and the generating elements of pNp. (Just starting abstractly with three $*$-free normal diffuse elements will not be very helpful, as we have then no idea how to get p and the required nine free elements.)

Actually, since our claim is equivalent to $\mathcal{L}(\mathbb{F}_3) \simeq M_2(\mathbb{C}) \otimes \mathcal{L}(\mathbb{F}_9)$, it will surely be a good idea to try to realize x_1, x_2, x_3 as 2×2 matrices. This will be achieved in the next section with the help of circular operators.

6.8 Circular operators and complex Gaussian random matrices

To construct the elements x_1, x_2, x_3 as required above, we need to make a digression into circular operators. Let X be an $2N \times 2N$ GUE random matrix. Let

$$P = \begin{pmatrix} I_n & 0_n \\ 0_n & 0_n \end{pmatrix} \qquad \text{and} \qquad G = \sqrt{2}\, PX(1-P).$$

Then G is a $N \times N$ matrix with independent identically distributed entries which are centred complex Gaussian random variables with complex variance $1/N$; such a matrix we call a *complex Gaussian random matrix*. We can determine the limiting $*$-moments of G as follows.

Write $Y_1 = (G+G^*)/\sqrt{2}$ and $Y_2 = -i(G-G^*)/\sqrt{2}$ then $G = (Y_1 + iY_2)/\sqrt{2}$ and Y_1 and Y_2 are independent $N \times N$ GUE random matrices. Therefore by the asymptotic freeness of independent GUE (see section 1.11), Y_1 and Y_2 converge as $N \to \infty$ to free standard semi-circulars s_1 and s_2.

Definition 9. Let s_1 and s_2 be free and standard semi-circular. Then we call $c = (s_1 + is_2)/\sqrt{2}$ a *circular operator*.

Since s_1 and s_2 are free, we can easily calculate the free cumulants of c. If $\varepsilon = \pm 1$ let us adopt the following notation for $x^{(\varepsilon)}$: $x^{(-1)} = x^*$ and $x^{(1)} = x$. Recall that for a standard semi-circular operator s

$$\kappa_n(s, \dots, s) = \begin{cases} 1, & n = 2 \\ 0, & n \neq 2 \end{cases}.$$

Thus

$$\begin{aligned} \kappa_n(c^{(\varepsilon_1)}, \dots, c^{(\varepsilon_n)}) &= 2^{-n/2}\kappa_n(s_1 + \varepsilon_1 i s_2, \dots, s_1 + i\varepsilon_n s_2) \\ &= 2^{-n/2}\big(\kappa_n(s_1, \dots, s_1) + i^n \varepsilon_1 \cdots \varepsilon_n \kappa_n(s_2, \dots, s_2)\big) \end{aligned}$$

since all mixed cumulants in s_1 and s_2 are 0. Thus $\kappa_n(c^{(\varepsilon_1)}, \dots, c^{(\varepsilon_n)}) = 0$ for $n \neq 2$, and

$$\kappa_2(c^{(\varepsilon_1)}, c^{(\varepsilon_2)}) = 2^{-1}\big(\kappa_2(s_1, s_1) - \varepsilon_1\varepsilon_2\kappa_2(s_2, s_2)\big) = \frac{1 - \varepsilon_1\varepsilon_2}{2} = \begin{cases} 1 & \varepsilon_1 \neq \varepsilon_2 \\ 0 & \varepsilon_1 = \varepsilon_2 \end{cases}.$$

Hence, $\kappa_2(c,c^*) = \kappa_2(c^*,c) = 1$, $\kappa_2(c,c) = \kappa_2(c^*,c^*) = 0$, and all other $*$-cumulants are 0. Thus

$$\tau((c^*c)^n) = \sum_{\pi\in NC(2n)} \kappa_\pi(c^*,c,c^*,c,\dots,c^*,c) = \sum_{\pi\in NC_2(2n)} \kappa_\pi(c^*,c,c^*,c,\dots,c^*,c).$$

Now note that any $\pi \in NC_2(2n)$ connects, by parity reasons, automatically only c with c^*, hence $\kappa_\pi(c^*,c,c^*,c,\dots,c^*,c) = 1$ for all $\pi \in NC_2(2n)$, and we have

$$\tau((c^*c)^n) = |NC_2(2n)| = \tau(s^{2n}),$$

where s is a standard semi-circular element. Since $t \mapsto \sqrt{t}$ is a uniform limit of polynomials in t, we have that the moments of $|c| = \sqrt{c^*c}$ and $|s| = \sqrt{s^2}$ are the same and $|c|$ and $|s|$ have the same distribution. The operator $|c| = |s|$ is called a *quarter-circular operator* and has moments

$$\tau(|c|^k) = \frac{1}{\pi}\int_0^2 t^k\sqrt{4-t^2}\,dt.$$

An additional result which we will need is Voiculescu's theorem on the polar decomposition of a circular operator.

Theorem 10. *Let (M,τ) be a W^*-probability space and $c \in M$ a circular operator. If $c = u\,|c|$ is its polar decomposition in M, then*

(*i*) *u and $|c|$ are $*$-free,*
(*ii*) *u is a Haar unitary,*
(*iii*) *$|c|$ is a quarter circular operator.*

The proof of (*i*) and (*ii*) can either be done using random matrix methods (as was done by Voiculescu [180]) or by showing that if u is a Haar unitary and q is a quarter-circular operator such that u and q are $*$-free, then uq has the same $*$-moments as a circular operator (this was done by Nica and Speicher [137]). The latter can be achieved, for example, by using the formula for cumulants of products, equation (2.23). For the details of this approach, see [137, Theorem 15.14].

Theorem 11. *Let $(\mathcal{A},\varphi)$ be a unital $*$-algebra with a state φ. Suppose $s_1, s_2, c \in \mathcal{A}$ are $*$-free and s_1 and s_2 semi-circular and c circular. Then*

$$x = \frac{1}{\sqrt{2}}\begin{pmatrix} s_1 & c \\ c^* & s_2 \end{pmatrix} \in (M_2(\mathcal{A}),\varphi_2)$$

is semi-circular.

Here we have used the standard notation $M_2(\mathcal{A}) = M_2(\mathbb{C})\otimes\mathcal{A}$ for 2×2 matrices with entries from $\mathcal{A}$ and $\varphi_2 = \mathrm{tr}\otimes\varphi$ for the composition of the normalized trace with φ.

Proof: Let $\mathbb{C}\langle x_{11}, x_{12}, x_{21}, x_{22}\rangle$ be the polynomials in the non-commuting variables $x_{11}, x_{12}, x_{21}, x_{22}$. Let

$$p_k(x_{11}, x_{12}, x_{21}, x_{22}) = \frac{1}{2}\mathrm{Tr}\left(\begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix}^k\right).$$

Now let $\mathcal{A}_N = M_N(L^{\infty-}(\Omega))$ be the $N \times N$ matrices with entries in $L^{\infty-}(\Omega) := \bigcap_{p\geq 1} L^p(\Omega)$, for some classical probability space Ω. On $\mathcal{A}_N$ we have the state $\varphi_N(X) = \mathrm{E}(N^{-1}\mathrm{Tr}(X))$. Now suppose in $\mathcal{A}_N$ we have S_1, S_2, and C, with S_1 and S_2 GUE random matrices and C a complex Gaussian random matrix and with the entries of S_1, S_2, C independent. Then we know that S_1, S_2, C converge in $*$-distribution to s_1, s_2, c, i.e. for any polynomial p in four non-commuting variables, we have $\varphi_N(p(S_1, C, C^*, S_2)) \to \varphi(p(s_1, c, c^*, s_2))$. Now let

$$X = \frac{1}{\sqrt{2}}\begin{pmatrix} S_1 & C \\ C^* & S_2 \end{pmatrix}.$$

Then X is in $\mathcal{A}_{2N}$, and

$$\begin{aligned}\varphi_{2N}(X^k) = \varphi_N\big(p_k(S_1, C, C^*, S_2)\big) \to \varphi\big(p_k(s_1, c, c^*, s_2)\big) &= \varphi(\frac{1}{2}\mathrm{Tr}(x^k)) \\ &= \mathrm{tr}\otimes\varphi(x^k).\end{aligned}$$

On the other hand, X is a $2N \times 2N$ GUE random matrix; so $\varphi_{2N}(X^k)$ converges to the k^{th} moment of a semi-circular operator. Hence x in $M_2(\mathcal{A})$ is semi-circular. □

Exercise 4. Suppose s_1, s_2, c, and x are as in Theorem 11. Show that x is semi-circular by computing $\varphi(\mathrm{tr}(x^n))$ directly using the methods of Lemma 1.9.

We can now present the realization of the three generators x_1, x_2, x_3 of $\mathcal{L}(\mathbb{F}_3)$ which we need for the proof of the compression result.

Lemma 12. *Let $\mathcal{A}$ be a unital $*$-algebra and φ a state on $\mathcal{A}$. Suppose s_1, s_2, s_3, s_4, c_1, c_2, u in $\mathcal{A}$ are $*$-free, with s_1, s_2, s_3, and s_4 semi-circular, c_1 and c_2 circular, and u a Haar unitary. Let*

$$x_1 = \begin{pmatrix} s_1 & c_1 \\ c_1^* & s_2 \end{pmatrix}, \quad x_2 = \begin{pmatrix} s_3 & c_2 \\ c_2^* & s_4 \end{pmatrix}, \quad x_3 = \begin{pmatrix} u & 0 \\ 0 & 2u \end{pmatrix}.$$

Then x_1, x_2, x_3 are $*$-free in $M_2(\mathcal{A})$ with respect to the state $\mathrm{tr}\otimes\varphi$*; x_1 and x_2 are semi-circular and x_3 is normal and diffuse.*

Proof: We model x_1 by X_1, x_2 by X_2, and x_3 by X_3 where

$$X_1 = \begin{pmatrix} S_1 & C_1 \\ C_1^* & S_2 \end{pmatrix}, \quad X_2 = \begin{pmatrix} S_3 & C_2 \\ C_2^* & S_3 \end{pmatrix}, \quad X_3 = \begin{pmatrix} U & 0 \\ 0 & 2U \end{pmatrix}$$

and S_1, S_2, S_3, S_4 are $N \times N$ GUE random matrices, C_1 and C_2 are $N \times N$ complex Gaussian random matrices, and U is a diagonal deterministic unitary matrix, chosen so that the entries of X_1 are independent from those of X_2 and that the diagonal entries of U converge in distribution to the uniform distribution on the unit circle. Then X_1, X_2, X_3 are asymptotically $*$-free by Theorem 4.4. Thus x_1, x_2, and x_3 are $*$-free because they have the same distribution as the limiting distribution of X_1, X_2, and X_3. By the previous Theorem 11, x_1 and x_2 are semi-circular. x_3 is clearly normal, and its spectral distribution is given by the uniform distribution on the union of the circle of radius 1 and the circle of radius 2. □

6.9 Proof of $\mathcal{L}(\mathbb{F}_3)_{1/2} \simeq \mathcal{L}(\mathbb{F}_9)$

We will now present the proof of Theorem 8.

Proof: We have shown that if we take four semi-circular operators s_1 s_2, s_3, s_4, two circular operators c_1, c_2, and a Haar unitary u in a von Neumann algebra M with trace τ such that $s_1, s_2, s_3, s_4, c_1, c_2, u$ are $*$-free, then

- the elements

$$x_1 = \begin{pmatrix} s_1 & c_1 \\ c_1^* & s_2 \end{pmatrix}, \qquad x_2 = \begin{pmatrix} s_3 & c_2 \\ c_2^* & s_4 \end{pmatrix}, \qquad x_3 = \begin{pmatrix} u & 0 \\ 0 & 2u \end{pmatrix}$$

  are $*$-free in $(M_2(M), \mathrm{tr} \otimes \tau)$,
- x_1 and x_2 are semi-circular, and x_3 is normal and has diffuse spectral measure.

Let $N = \mathrm{vN}(x_1, x_2, x_3) \subseteq M_2(M)$. Then, by Theorem 6, $N \simeq \mathcal{L}(\mathbb{F}_3)$. Since

$$\begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix} = x_3^* x_3 \in N, \quad \text{we also have the spectral projection} \quad p = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \in N,$$

and thus $px_1(1-p) \in N$ and $px_2(1-p) \in N$. We have the polar decompositions

$$\begin{pmatrix} 0 & c_1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & v_1 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 0 \\ 0 & |c_1| \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 & c_2 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & v_2 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 0 \\ 0 & |c_2| \end{pmatrix},$$

where $c_1 = v_1|c_1|$ and $c_2 = v_2|c_2|$ are the polar decompositions of c_1 and c_2, respectively, in M.

Hence we see that $N = \mathrm{vN}(x_1, x_2, x_3)$ is generated by the ten elements

$$y_1 = \begin{pmatrix} s_1 & 0 \\ 0 & 0 \end{pmatrix} \quad y_2 = \begin{pmatrix} 0 & 0 \\ 0 & s_2 \end{pmatrix} \quad y_3 = \begin{pmatrix} 0 & v_1 \\ 0 & 0 \end{pmatrix} \quad y_4 = \begin{pmatrix} 0 & 0 \\ 0 & |c_1| \end{pmatrix} \quad y_5 = \begin{pmatrix} s_3 & 0 \\ 0 & 0 \end{pmatrix}$$

$$y_6 = \begin{pmatrix} 0 & 0 \\ 0 & s_4 \end{pmatrix} \quad y_7 = \begin{pmatrix} 0 & v_2 \\ 0 & 0 \end{pmatrix} \quad y_8 = \begin{pmatrix} 0 & 0 \\ 0 & |c_2| \end{pmatrix} \quad y_9 = \begin{pmatrix} u & 0 \\ 0 & 0 \end{pmatrix} \quad y_{10} = \begin{pmatrix} 0 & 0 \\ 0 & u \end{pmatrix}.$$

Let us put

$$v := \begin{pmatrix} 0 & v_1 \\ 0 & 0 \end{pmatrix}; \quad \text{then} \quad v^* v = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad v v^* = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = p = p^2.$$

Since we can write now any $py_{i_1} \cdots y_{i_n} p$ in the form $py_{i_1} 1 y_{i_2} 1 \cdots 1 y_{i_n} p$ and replace each 1 by $p^2 + v^* v$, it is clear that $\bigcup_{i=1}^{10} \{py_i p, py_i v^*, vy_i p, vy_i v^*\}$ generate pNp. This gives for pNp the generators

$$s_1, \quad v_1 s_2 v_1^*, \quad v_1 v_1^*, \quad v_1 \, |c_1| \, v_1^*, \quad s_3, \quad v_1 s_4 v_1^*, \quad v_2 v_1^*, \quad v_1 \, |c_2| \, v_1^*, \quad u, \quad v_1 u v_1^*.$$

Note that $v_1 v_1^* = 1$ can be removed from the set of generators. To check that the remaining nine elements are $*$-free and diffuse, we recall a few elementary facts about freeness.

Exercise 5. Show the following:

(*i*) if $\mathcal{A}_1$ and $\mathcal{A}_2$ are free subalgebras of $\mathcal{A}$, if $\mathcal{A}_{11}$ and $\mathcal{A}_{12}$ are free subalgebras of $\mathcal{A}_1$, and if $\mathcal{A}_{21}$ and $\mathcal{A}_{22}$ are free subalgebras of $\mathcal{A}_2$; then $\mathcal{A}_{11}, \mathcal{A}_{12}, \mathcal{A}_{21}, \mathcal{A}_{22}$ are free;

(*ii*) if u is a Haar unitary $*$-free from $\mathcal{A}$, then $\mathcal{A}$ is $*$-free from $u\mathcal{A}u^*$;

(*iii*) if u_1 and u_2 are Haar unitaries and u_2 is $*$-free from $\{u_1\} \cup \mathcal{A}$ then $u_2 u_1^*$ is a Haar unitary and is $*$-free from $u_1 \mathcal{A} u_1^*$.

By construction $s_1, s_2, s_3, s_4, |c_1|, |c_2|, v_1, v_2, u$ are $*$-free. Thus, in particular, $s_2, s_4, |c_1|, |c_2|, v_2, u$ are $*$-free. Hence, by (*ii*), $v_1 s_2 v_1^*, v_1 s_4 v_1^*, v_1 |c_1| v_1^*, v_1 |c_2| v_1^*$, $v_1 u v_1^*$ are $*$-free and, in addition, $*$-free from u, s_1, s_3, v_2. Thus

$$u, \quad s_1, \quad s_3, \quad v_1 s_2 v_1^*, \quad v_1 s_4 v_1^*, \quad v_1 |c_1| v_1^*, \quad v_1 |c_2| v_1^*, \quad v_1 u v_1^*, \quad v_2$$

are $*$-free. Let $\mathcal{A} = \mathrm{alg}(s_2, s_4, |c_1|, |c_2|, u)$. We have that v_2 is $*$-free from $\{v_1\} \cup \mathcal{A}$, so by (*iii*), $v_2 v_1^*$ is $*$-free from $v_1 \mathcal{A} v_1^*$. Thus, $v_2 v_1^*$ is $*$-free from

$$v_1 s_2 v_1^*, \quad v_1 s_4 v_1^*, \quad v_1 |c_1| v_1^*, \quad v_1 |c_2| v_1^*, \quad v_1 u v_1^*$$

and it was already $*$-free from s_1, s_3 and u. Thus by (*i*) our nine elements

$$s_1, \quad s_3, \quad v_1 s_2 v_1^*, \quad v_1 s_4 v_1^*, \quad v_1 |c_1| v_1^*, \quad v_1 |c_2| v_1^*, \quad u, \quad v_1 u v_1^*, \quad v_2 v_1^*$$

are $*$-free. Since they are either semi-circular, quarter-circular, or Haar elements, they are all normal and diffuse; as they generate pNp, we have that pNp is generated by nine $*$-free normal and diffuse elements and thus, by Theorem 6, $pNp \simeq \mathcal{L}(\mathbb{F}_9)$. Hence $\mathcal{L}(\mathbb{F}_3)_{1/2} \simeq \mathcal{L}(\mathbb{F}_9)$. □

6.10 The general case $\mathcal{L}(\mathbb{F}_{\mathbf{n}})_{\mathbf{1/k}} \simeq \mathcal{L}(\mathbb{F}_{\mathbf{1+(n-1)k^2}})$

Sketch We sketch now the proof for the general case of Equation (6.3). We write $\mathcal{L}(\mathbb{F}_n) = \mathrm{vN}(x_1, \ldots, x_n)$ where for $1 \le i \le n-1$ each x_i is a semi-circular element of the form

$$x_i = \frac{1}{\sqrt{k}} \begin{pmatrix} s_1^{(i)} & c_{12}^{(i)} & \ldots & c_{1k}^{(i)} \\ c_{12}^{(i)*} & s_2^{(i)} & \ldots & c_{2k}^{(i)} \\ \vdots & & \ddots & \vdots \\ c_{1k}^{(i)*} & \ldots & \ldots & s_k^{(i)} \end{pmatrix} \quad \text{and where} \quad x_n = \begin{pmatrix} u & 0 & \ldots & 0 \\ 0 & 2u & \ldots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \ldots & ku \end{pmatrix},$$

with all $s_j^{(i)}$ ($j = 1, \ldots, k; i = 1, \ldots, n-1$) semi-circular, all $c_{pq}^{(i)}$ ($1 \le p < q \le k$; $i = 1, \ldots, n-1$) circular, and u a Haar unitary, so that all elements are $*$-free.

So we have $(n-1)k$ semi-circular operators, $(n-1)\binom{k}{2}$ circular operators, and one Haar unitary. Each circular operator produces two free elements, so we have in total

$$(n-1)k + 2(n-1)\binom{k}{2} + 1 = (n-1)k^2 + 1$$

free and diffuse generators. Thus $\mathcal{L}(\mathbb{F}_n)_{1/k} \simeq \mathcal{L}(\mathbb{F}_{1+(n-1)k^2})$. □

6.11 Interpolating free group factors

The formula $\mathcal{L}(\mathbb{F}_n)_{1/k} \simeq \mathcal{L}(\mathbb{F}_m)$, which up to now makes sense only for integer m, n, and k, suggests that one might try to define $\mathcal{L}(\mathbb{F}_r)$ also for noninteger r by compression. A crucial issue is that, by the above formula, different compressions should give the same result. That this really works and is consistent was shown, independently, by Dykema [67] and Rădulescu [145].

Theorem 13. *Let R be the hyperfinite II_1 factor and $\mathcal{L}(\mathbb{F}_\infty) = \mathrm{vN}(s_1, s_2, \ldots)$ be a free group factor generated by countably many free semi-circular elements s_i, such that R and $\mathcal{L}(\mathbb{F}_\infty)$ are free in some W^*-probability space (M, τ). Consider orthogonal projections $p_1, p_2, \cdots \in R$ and put $r := 1 + \sum_j \tau(p_j)^2 \in [1, \infty]$. Then the von Neumann algebra*

$$\mathcal{L}(\mathbb{F}_r) := \mathrm{vN}(R, p_j s_j p_j (j \in \mathbb{N})) \tag{6.4}$$

is a factor and depends, up to isomorphism, only on r.

These $\mathcal{L}(\mathbb{F}_r)$ for $r \in \mathbb{R}$, $1 \le r \le \infty$ are the *interpolating free group factors*. Note that we do not claim to have noninteger free groups $\mathbb{F}_r$. The notation $\mathcal{L}(\mathbb{F}_r)$ cannot be split into smaller components.

Dykema and Rădulescu showed the following results.

Theorem 14. *1) For $r \in \{2, 3, 4, \dots, \infty\}$ the interpolating free group factor $\mathcal{L}(\mathbb{F}_r)$ is the usual free group factor.*
2) We have for all $r, s \geq 1$: $\mathcal{L}(\mathbb{F}_r) \star \mathcal{L}(\mathbb{F}_s) \simeq \mathcal{L}(\mathbb{F}_{r+s})$.
3) We have for all $r \geq 1$ and all $t \in (0, \infty)$ the same compression formula as in the integer case:

$$\big(\mathcal{L}(\mathbb{F}_r)\big)_t \simeq \mathcal{L}(\mathbb{F}_{1+t^{-2}(r-1)}). \tag{6.5}$$

The compression formula above is also valid in the case $r = \infty$; since then $1 + t^{-2}(r - 1) = \infty$, it yields in this case that any compression of $\mathcal{L}(\mathbb{F}_\infty)$ is isomorphic to $\mathcal{L}(\mathbb{F}_\infty)$; or in other words, we have that the fundamental group of $\mathcal{L}(\mathbb{F}_\infty)$ is equal to $\mathbb{R}^+$.

6.12 The dichotomy for the free group factor isomorphism problem

Whereas for $r = \infty$, the compression of $\mathcal{L}(\mathbb{F}_r)$ gives the same free group factor (and thus we know that the fundamental group is maximal in this case); for $r < \infty$ we get some other free group factors. Since we do not know whether these are isomorphic to the original $\mathcal{L}(\mathbb{F}_r)$, we cannot decide upon the fundamental group in this case. However, on the positive side, we can connect different free group factors by compressions; this yields that some isomorphisms among the free group factors will imply other isomorphisms. For example, if we would know that $\mathcal{L}(\mathbb{F}_2) \simeq \mathcal{L}(\mathbb{F}_3)$, then this would imply that also

$$\mathcal{L}(\mathbb{F}_5) \simeq \big(\mathcal{L}(\mathbb{F}_2)\big)_{1/2} \simeq \big(\mathcal{L}(\mathbb{F}_3)\big)_{1/2} \simeq \mathcal{L}(\mathbb{F}_9).$$

The possibility of using arbitrary $t \in (0, \infty)$ in our compression formulas allows to connect any two free group factors by compression, which gives then the following dichotomy for the free group isomorphism problem. This is again due to Dykema and Rădulescu.

Theorem 15. *We have exactly one of the following two possibilities.*

(i) All interpolating free group factors are isomorphic: $\mathcal{L}(\mathbb{F}_r) \simeq \mathcal{L}(\mathbb{F}_s)$ for all $1 < r, s \leq \infty$. In this case the fundamental group of each $\mathcal{L}(\mathbb{F}_r)$ is equal to $\mathbb{R}^+$.
(ii) The interpolating free group factors are pairwise non-isomorphic: $\mathcal{L}(\mathbb{F}_r) \not\simeq \mathcal{L}(\mathbb{F}_s)$ for all $1 < r \neq s \leq \infty$. In this case the fundamental group of each $\mathcal{L}(\mathbb{F}_r)$, for $r \neq \infty$, is equal to $\{1\}$.

Chapter 7
Free Entropy χ: The Microstates Approach via Large Deviations

An important concept in classical probability theory is Shannon's notion of entropy. Having developed the analogy between free and classical probability theory, one hopes to find that a notion of *free entropy* exists in counterpart to the Shannon entropy. In fact there is a useful notion of free entropy. However, the development of this new concept is at present far from complete. The current state of affairs is that there are two distinct approaches to free entropy. These should give isomorphic theories, but at present we only know that they coincide in a limited number of situations.

The first approach to a theory of free entropy is via *microstates*. This is rooted in the concept of large deviations. The second approach is *microstates free*. This draws its inspiration from the statistical approach to classical entropy via the notion of Fisher information. The unification problem in free probability theory is to prove that these two theories of free entropy are consistent. We will in this chapter only talk about the first approach via microstates; the next chapter will address the microstates free approach.

7.1 Motivation

Let us return to the connection between random matrix theory and free probability theory which we have been developing. We know that a p-tuple $(A_N^{(1)}, \ldots, A_N^{(p)})$ of $N \times N$ matrices chosen independently at random with respect to the GUE density (compare Exercise 1.8), $P_N(A) = \text{const} \cdot \exp(-N\mathrm{Tr}(A^2)/2)$, on the space of $N \times N$ Hermitian matrices converges almost surely (in moments with respect to the normalized trace) to a freely independent family $(s_1, \ldots, s_p)$ of semi-circular elements lying in a non-commutative probability space; see Theorem 4.4. The von Neumann algebra generated by p freely independent semi-circulars is the von Neumann algebra $L(\mathbb{F}_p)$ of the free group on p generators.

J.A. Mingo, R. Speicher, *Free Probability and Random Matrices*,
Fields Institute Monographs 35, DOI 10.1007/978-1-4939-6942-5_7

We ask now the following question: How likely is it to observe other distributions/operators for large N?

Let us consider the case $p = 1$ more closely. For a random Hermitian matrix $A = A^*$ (distribution as above) with real random eigenvalues $\lambda_1 \leq \cdots \leq \lambda_N$, denote by

$$\mu_A = \frac{1}{N}(\delta_{\lambda_1} + \cdots + \delta_{\lambda_N}) \tag{7.1}$$

the eigenvalue distribution of A (also known as the *empirical eigenvalue distribution*), which is a random measure on $\mathbb{R}$. Wigner's semi-circle law states that as $N \to \infty$, $P_N(\mu_A \approx \mu_W) \to 1$, where μ_W is the (non-random) semi-circular distribution and $\mu_A \approx \mu_W$ means that the measures are close in a sense that can be made precise. We are now interested in the deviations from this. What is the rate of decay of the probability $P_N(\mu_A \approx \nu)$, where ν is some measure (not necessarily the semi-circle)? We expect that

$$P_N(\mu_A \approx \nu) \sim e^{-N^2 I(\nu)} \tag{7.2}$$

for some *rate function* I vanishing at μ_W. By analogy with the classical theory of large deviations, I should correspond to a suitable notion of free entropy.

We used in the above the notion "$\approx$" for meaning "being close" and "$\sim$" for "behaves asymptotically (in N) like"; here they should just be taken on an intuitive level, later, in the actual theorems they will be made more precise.

In the next two sections, we will recall some of the basic facts of the classical theory of large deviations and, in particular, Sanov's theorem; this standard material can be found, for example, in the book [64]. In Section 7.4 we will come back to the random matrix question.

7.2 Large deviation theory and Cramér's theorem

Consider a real-valued random variable X with distribution μ. Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed random variables with the same distribution as X, and put $S_n = (X_1 + \cdots + X_n)/n$. Let $m = E[X]$ and $\sigma^2 = \mathrm{var}(X) = E[X^2] - m^2$. Then the law of large numbers asserts that $S_n \to m$, if $\mathrm{E}[|X|] < \infty$; while if $\mathrm{E}[X^2] < \infty$, the central limit theorem tells us that for large n

$$S_n \approx m + \frac{\sigma}{\sqrt{n}} N(0, 1). \tag{7.3}$$

For example, if $\mu = N(0, 1)$ is Gaussian, then $m = 0$ and S_n has the Gaussian distribution $N(0, 1/n)$, and hence

$$P(S_n \approx x) = P(S_n \in [x, x + dx]) \approx e^{-nx^2/2} dx \frac{\sqrt{n}}{\sqrt{2\pi}} \sim e^{-nI(x)} dx.$$

Thus the probability that S_n is near the value x decays exponentially in n at a rate determined by x, namely the *rate function* $I(x) = x^2/2$. Note that the convex function $I(x)$ has a global minimum at $x = 0$, the minimum value there being 0, which corresponds to the fact that S_n approaches the mean 0 in probability.

This behaviour is described in general by the following theorem of Cramér. Let X, μ, $\{X_i\}_i$, and S_n be as above. There exists a function $I(x)$, the rate function, such that

$$\begin{aligned} P(S_n > x) &\sim e^{-nI(x)}, \qquad x > m \\ P(S_n < x) &\sim e^{-nI(x)}, \qquad x < m. \end{aligned}$$

How does one calculate the rate function I for a given distribution μ? We shall let X be a random variable with the same distribution as the X_i's. For arbitrary $x > m$, one has for all $\lambda \geq 0$

$$\begin{aligned} P(S_n > x) &= P(nS_n > nx) \\ &= P(e^{\lambda(nS_n - nx)} \geq 1) \\ &\leq E[e^{\lambda(nS_n - nx)}] \qquad \text{(by Markov's inequality)} \\ &= e^{-\lambda nx} E[e^{\lambda(X_1 + \cdots + X_n)}] \\ &= (e^{-\lambda x} E[e^{\lambda X}])^n. \end{aligned}$$

Here we are allowing that $\mathrm{E}[e^{\lambda X}] = +\infty$. Now put

$$\Lambda(\lambda) := \log E[e^{\lambda X}], \tag{7.4}$$

the *cumulant generating series* of μ; c.f. Section 1.1. We consider Λ to be an extended real-valued function but here only consider μ for which $\Lambda(\lambda)$ is finite for all real λ in some open set containing 0; however, Cramér's theorem (Theorem 1) holds without this assumption. With this assumption Λ has a power series expansion with radius of convergence $\lambda_0 > 0$, and in particular all moments exist.

Exercise 1. Suppose that X is a real random variable and there is $\lambda_0 > 0$ so that for all $|\lambda| \leq \lambda_0$ we have $E(e^{\lambda X}) < \infty$. Then X has moments of all orders, and the function $\lambda \mapsto E(e^{\lambda X})$ has a power series expansion with a radius of convergence of at least λ_0.

Then the inequality above reads

$$P(S_n > x) \leq e^{-\lambda nx + n\Lambda(\lambda)} = e^{-n(\lambda x - \Lambda(\lambda))}, \tag{7.5}$$

which is valid for all $0 \leq \lambda$. By Jensen's inequality we have, for all $\lambda \in \mathbb{R}$,

$$\Lambda(\lambda) = \log E[e^{\lambda X}] \geq E[\log e^{\lambda X}] = \lambda m. \tag{7.6}$$

This implies that for $\lambda < 0$ and $x > m$ we have $-n(\lambda x - \Lambda(\lambda)) \geq 0$, and so equation (7.5) is valid for all λ. Thus

$$P(S_n > x) \leq \inf_{\lambda} e^{-n(\lambda x - \Lambda(\lambda))} = \exp\left(-n \sup_{\lambda}(\lambda x - \Lambda(\lambda))\right).$$

The function $\lambda \mapsto \Lambda(\lambda)$ is convex, and the *Legendre transform* of Λ defined by

$$\Lambda^*(x) := \sup_{\lambda}(\lambda x - \Lambda(\lambda)) \tag{7.7}$$

is also a convex function of x, as it is the supremum of a family of convex functions of x.

Exercise 2. Show that $(E(Xe^{\lambda X}))^2 \leq E(e^{\lambda X})E(Xe^{\lambda X})$. Show that $\lambda \mapsto \Lambda(\lambda)$ is convex.

Note that $\Lambda(0) = \log 1 = 0$; thus, $\Lambda^*(x) \geq (0x - \Lambda(0)) = 0$ is non-negative, and hence equation (7.6) implies that $\Lambda^*(m) = 0$.

Thus, we have proved that, for $x > m$,

$$P(S_n > x) \leq e^{-n\Lambda^*(x)}, \tag{7.8}$$

where Λ^* is the Legendre transform of the cumulant generating function Λ. In the same way, one proves the same estimate for $P(S_n < x)$ for $x < m$. This gives Λ^* as a candidate for the rate function. Moreover we have by Exercise 3 that $\lim_n \log[P(S_n > x)]^{1/n}$ exists and by Equation (7.8) this limit is less than $\exp(-\Lambda^*(x))$. If we assume that neither $P(X>x)$ nor $P(X<x)$ is 0, $\exp(-\Lambda^*(x))$ will be the limit. In general we have

$$-\inf_{y>x} \Lambda^*(y) \leq \liminf_{n} \frac{1}{n} \log P(S_n > x) \leq \limsup_{n} \frac{1}{n} \log P(S_n \geq x) \leq -\inf_{y \geq x} \Lambda^*(y).$$

Exercise 3. Let $a_n = \log P(S_n > a)$. Show that

(*i*) for all m, n: $a_{m+n} \geq a_m + a_n$;
(*ii*) for all m

$$\liminf_{n \to \infty} \frac{a_n}{n} \geq \frac{a_m}{m};$$

(*iii*) $\lim_{n \to \infty} a_n/n$ exists.

However, in preparation for the vector-valued version, we will show that $\exp(-n\Lambda^*(x))$ is asymptotically a lower bound; more precisely, we need to verify that

$$\liminf_{n\to\infty} \frac{1}{n} \log P(x-\delta < S_n < x+\delta) \geq -\Lambda^*(x)$$

for all x and all $\delta > 0$. By replacing X_i by $X_i - x$, we can reduce this to the case $x = 0$, namely, showing that

$$-\Lambda^*(0) \leq \liminf_{n\to\infty} \frac{1}{n} \log P(-\delta < S_n < \delta). \tag{7.9}$$

Note that $-\Lambda^*(0) = \inf_\lambda \Lambda(\lambda)$. The idea of the proof of (7.9) is then to perturb the distribution μ to $\tilde{\mu}$ such that $x = 0$ is the mean of $\tilde{\mu}$. Let us only consider the case where Λ has a global minimum at some point η. This will always be the case if μ has compact support and both $P(X > 0)$ and $P(X < 0)$ are not 0. The general case can be reduced to this by a truncation argument. With this reduction $\Lambda(\lambda)$ is finite for all λ, and thus Λ has an infinite radius of convergence (c.f. Exercise 1), and thus Λ is differentiable. So we have $\Lambda'(\eta) = 0$. Now let $\tilde{\mu}$ be the measure on $\mathbb{R}$ such that

$$d\tilde{\mu}(x) = e^{\eta x - \Lambda(\eta)} d\mu(x). \tag{7.10}$$

Note that

$$\int_{\mathbb{R}} d\tilde{\mu}(x) = e^{-\Lambda(\eta)} \int_{\mathbb{R}} e^{\eta x} d\mu(x) = e^{-\Lambda(\eta)} E[e^{\eta X}] = e^{-\Lambda(\eta)} e^{\Lambda(\eta)} = 1,$$

which verifies that $\tilde{\mu}$ is a probability measure. Consider now i.i.d. random variables $\{\tilde{X}_i\}_i$ with distribution $\tilde{\mu}$, and put $\tilde{S}_n = (\tilde{X}_1 + \cdots + \tilde{X}_n)/n$. Let $\tilde{X}$ have the distribution $\tilde{\mu}$. We have

$$\begin{aligned} E[\tilde{X}] = \int_{\mathbb{R}} x\, d\tilde{\mu}(x) &= e^{-\Lambda(\eta)} \int_{\mathbb{R}} x e^{\eta x} d\mu(x) = e^{-\Lambda(\eta)} \frac{d}{d\lambda} \int_{\mathbb{R}} e^{\lambda x} d\mu(x) \Big|_{\lambda=\eta} \\ &= e^{-\Lambda(\eta)} \frac{d}{d\lambda} e^{\Lambda(\lambda)} \Big|_{\lambda=\eta} = e^{-\Lambda(\eta)} \Lambda'(\eta) e^{\Lambda(\eta)} = \Lambda'(\eta) = 0. \end{aligned}$$

Now, for all $\epsilon > 0$, we have $\exp(\eta \sum x_i) \leq \exp(n\epsilon|\eta|)$ whenever $|\sum x_i| \leq n\epsilon$ and so

$$\begin{aligned} P(-\epsilon < S_n < \epsilon) &= \int_{|\sum_{i=1}^n x_i| < n\epsilon} d\mu(x_1) \cdots d\mu(x_n) \\ &\geq e^{-n\epsilon|\eta|} \int_{|\sum_{i=1}^n x_i| < n\epsilon} e^{\eta \sum x_i} d\mu(x_1) \cdots d\mu(x_n) \end{aligned}$$

$$= e^{-n\epsilon|\eta|} e^{n\Lambda(\eta)} \int_{|\sum_{i=1}^{n} x_i| < n\epsilon} d\tilde{\mu}(x_1) \cdots d\tilde{\mu}(x_n)$$
$$= e^{-n\epsilon|\eta|} e^{n\Lambda(\eta)} P(-\epsilon < \tilde{S}_n < \epsilon).$$

By the weak law of large numbers, $\tilde{S}_n \to E[\tilde{X}_i] = 0$ in probability, i.e. we have $\lim_{n\to\infty} P(-\epsilon < \tilde{S}_n < \epsilon) = 1$ for all $\epsilon > 0$. Thus for all $0 < \epsilon < \delta$

$$\begin{aligned}
\liminf_{n\to\infty} \frac{1}{n} \log P(-\delta < S_n < \delta) &\geq \liminf_{n\to\infty} \frac{1}{n} \log P(-\epsilon < S_n < \epsilon) \\
&\geq \Lambda(\eta) - \epsilon|\eta|, \qquad \text{for all } \epsilon > 0 \\
&\geq \Lambda(\eta) \\
&= \inf \Lambda(\lambda) \\
&= -\Lambda^*(0).
\end{aligned}$$

This sketches the proof of Cramér's theorem for $\mathbb{R}$. The higher-dimensional form of Cramér's theorem can be proved in a similar way.

Theorem 1 (Cramér's Theorem for $\mathbb{R}^d$). *Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random vectors, i.e. independent $\mathbb{R}^d$-valued random variables with common distribution μ (a probability measure on $\mathbb{R}^d$). Put*

$$\Lambda(\lambda) := \mathrm{E}[e^{\langle \lambda, X_i \rangle}], \ \lambda \in \mathbb{R}^d, \tag{7.11}$$

and

$$\Lambda^*(x) := \sup_{\lambda \in \mathbb{R}^d} \{\langle \lambda, x \rangle - \Lambda(\lambda)\}. \tag{7.12}$$

Assume that $\Lambda(\lambda) < \infty$ for all $\lambda \in \mathbb{R}^d$, and put $S_n := (X_1 + \cdots + X_n)/n$.

Then the distribution μ_{S_n} of the random variable S_n satisfies a large deviation principle with rate function Λ^**, i.e.*

- *$x \mapsto \Lambda^*(x)$ is lower semicontinuous (actually convex)*
- *Λ^* is* good*, i.e. $\{x \in \mathbb{R}^d : \Lambda^*(x) \leq \alpha\}$ is compact for all $\alpha \in \mathbb{R}$*
- *For any closed set $F \subset \mathbb{R}^d$,*

$$\limsup_{n\to\infty} \frac{1}{n} \log P(S_n \in F) \leq -\inf_{x\in F} \Lambda^*(x) \tag{7.13}$$

- *For any open set $G \subset \mathbb{R}^d$,*

$$\liminf_{n\to\infty} \frac{1}{n} \log P(S_n \in G) \geq -\inf_{x\in G} \Lambda^*(x). \tag{7.14}$$

7.3 Sanov's theorem and entropy

We have seen Cramér's theorem for $\mathbb{R}^d$; in an informal way, it says $P(S_n \approx x) \sim \exp(-n\Lambda^*(x))$. Actually, we are interested not in S_n, but in the empirical distribution $(\delta_{X_1} + \cdots + \delta_{X_n})/n$.

Let us consider this in the special case of random variables $X_i : \Omega \to A$, taking values in a finite alphabet $A = \{a_1, \ldots, a_d\}$, with $p_k := P(X_i = a_k)$. As $n \to \infty$, the empirical distribution of the X_i's should converge to the "most likely" probability measure $(p_1, \ldots, p_d)$ on A.

Now define the vector of indicator functions $Y_i : \Omega \to \mathbb{R}^d$ by

$$Y_i := (1_{\{a_1\}}(X_i), \ldots, 1_{\{a_d\}}(X_i)), \tag{7.15}$$

so that in particular p_k is equal to the probability that Y_i will have a 1 in the k-th spot and 0's elsewhere. Then the averaged sum $(Y_1 + \cdots + Y_n)/n$ gives the relative frequency of $a_1, \ldots, a_d$, i.e. it contains the same information as the empirical distribution of $(X_1, \ldots, X_n)$.

A probability measure on A is given by a d-tuple $(q_1, \ldots, q_d)$ of positive real numbers satisfying $q_1 + \cdots + q_d = 1$. By Cramér's theorem,

$$\begin{aligned} P\left\{\frac{1}{n}(\delta_{X_1} + \cdots + \delta_{X_n}) \approx (q_1, \ldots, q_d)\right\} &= P\left\{\frac{Y_1 + \cdots + Y_n}{n} \approx (q_1, \ldots, q_d)\right\} \\ &\sim e^{-n\Lambda^*(q_1, \ldots, q_d)}. \end{aligned}$$

Here

$$\Lambda(\lambda_1, \ldots, \lambda_d) = \log E[e^{\langle \lambda, Y_i \rangle}] = \log(p_1 e^{\lambda_1} + \cdots + p_d e^{\lambda_d}).$$

Thus the Legendre transform is given by

$$\Lambda^*(q_1, \ldots, q_d) = \sup_{(\lambda_1, \ldots, \lambda_d)} \{\lambda_1 q_1 + \cdots + \lambda_d q_d - \Lambda(\lambda_1, \ldots, \lambda_d)\}.$$

We compute the supremum over all tuples $(\lambda_1, \ldots, \lambda_d)$ by finding the partial derivative $\partial/\partial\lambda_i$ of $\lambda_1 q_1 + \cdots + \lambda_d q_d - \Lambda(\lambda_1, \ldots, \lambda_d)$ to be

$$q_i - \frac{1}{p_1 e^{\lambda_1} + \cdots + p_d e^{\lambda_d}} p_i e^{\lambda_i}.$$

By concavity the maximum occurs when

$$\lambda_i = \log \frac{q_i}{p_i} + \log(p_1 e^{\lambda_1} + \cdots + p_d e^{\lambda_d}) = \log \frac{q_i}{p_i} + \Lambda(\lambda_1, \ldots, \lambda_d),$$

and we compute

$$
\begin{aligned}
&\Lambda^*(q_1, \ldots, q_d) \\
&= q_1 \log \frac{q_1}{p_1} + \cdots + q_d \log \frac{q_d}{p_d} + (q_1 + \cdots + q_d)\Lambda(\lambda_1, \ldots, \lambda_d) - \Lambda(\lambda_1, \ldots, \lambda_d) \\
&= q_1 \log \frac{q_1}{p_1} + \cdots + q_d \log \frac{q_d}{p_d}.
\end{aligned}
$$

The latter quantity is Shannon's relative entropy, $H((q_1, \ldots, q_d)|(p_1, \ldots, p_d))$, of $(q_1, \ldots, q_d)$ with respect to $(p_1, \ldots, p_d)$. Note that $H((q_1, \ldots, q_d)|(p_1, \ldots, p_d)) \geq 0$, with equality holding if and only if $q_1 = p_1, \ldots, q_d = p_d$.

Thus $(p_1, \ldots, p_d)$ is the most likely realization, with other realizations exponentially unlikely; their unlikelihood is measured by the rate function Λ^*; and this rate function is indeed Shannon's relative entropy. This is the statement of Sanov's theorem. We have proved it here for a finite alphabet; it also holds for continuous distributions.

Theorem 2 (Sanov's Theorem). *Let $X_1, X_2, \ldots$ be i.i.d. real-valued random variables with common distribution μ, and let*

$$\nu_n = \frac{1}{n}(\delta_{X_1} + \cdots + \delta_{X_n}) \tag{7.16}$$

be the empirical distribution of $X_1, \ldots, X_n$, which is a random probability measure on $\mathbb{R}$. Then $\{\nu_n\}_n$ satisfies a large deviation principle with rate function $I(\nu) = S(\nu, \mu)$ (called the relative entropy*) given by*

$$I(\nu) = \begin{cases} \int p(t) \log p(t) d\mu(t), & \text{if } d\nu = p\, d\mu \\ +\infty, & \text{otherwise.} \end{cases} \tag{7.17}$$

Concretely, this means the following. Consider the set $\mathcal{M}$ of probability measures on $\mathbb{R}$ with the weak topology (which is a metrizable topology, e.g. by the Lévy metric). Then for closed F and open G in $\mathcal{M}$, we have

$$\limsup_{n\to\infty} \frac{1}{n} \log P(\nu_n \in F) \leq - \inf_{\nu \in F} S(\nu, \mu) \tag{7.18}$$

$$\liminf_{n\to\infty} \frac{1}{n} \log P(\nu_n \in G) \geq - \inf_{\nu \in G} S(\nu, \mu). \tag{7.19}$$

7.4 Back to random matrices and one-dimensional free entropy

Consider again the space $\mathcal{H}_N$ of Hermitian matrices equipped with the probability measure P_N having density

$$dP_N(A) = \text{const} \cdot e^{-\frac{N}{2}\operatorname{Tr}(A^2)} dA. \tag{7.20}$$

We let $\mathbb{R}^N_{\geq} = \{(x_1, \ldots, x_N) \in \mathbb{R}^N \mid x_1 \leq \cdots \leq x_N\}$. For a self-adjoint matrix A, we write the eigenvalues of A as $\lambda_1(A) \leq \cdots \leq \lambda_N(A)$. The joint eigenvalue distribution $\tilde{P}_N$ on $\mathbb{R}^N_{\geq}$ is defined by

$$\tilde{P}_N(B) := P_N\{A \in \mathcal{H}_N \mid (\lambda_1(A), \ldots, \lambda_N(A)) \in B\}. \tag{7.21}$$

The permutation group S_N acts on $\mathbb{R}^N$ by permuting the coordinates, with $\mathbb{R}^N_{\geq}$ as a fundamental domain (ignoring sets of measure 0). So we can use this action to transport $\tilde{P}_N$ around $\mathbb{R}^N$ to get a probability measure on $\mathbb{R}^N$.

One knows (e.g. see [7, Thm. 2.5.2]) that $\tilde{P}_N$ is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^N$ and has density

$$d\tilde{P}_N(\lambda_1, \ldots, \lambda_N) = C_N \cdot e^{-\frac{N}{2}\sum_{i=1}^N \lambda_i^2} \prod_{i<j} (\lambda_i - \lambda_j)^2 \prod_{i=1}^N d\lambda_i, \tag{7.22}$$

where

$$C_N = \frac{N^{N^2/2}}{(2\pi)^{N/2} \prod_{j=1}^N j!}. \tag{7.23}$$

We want to establish a large deviation principle for the empirical eigenvalue distribution $\mu_A = (\delta_{\lambda_1(A)} + \cdots + \delta_{\lambda_N(A)})/N$ of a random matrix in $\mathcal{H}_N$.

One can argue heuristically as follows for the expected form of the rate function. We have

$$\begin{aligned} P_N\{\mu_A \approx \nu\} &= \tilde{P}_N\left\{\frac{1}{N}(\delta_{\lambda_1} + \cdots + \delta_{\lambda_N}) \approx \nu\right\} \\ &= C_N \cdot \int_{\{\frac{1}{N}(\delta_{\lambda_1} + \cdots + \delta_{\lambda_N}) \approx \nu\}} e^{-\frac{N}{2}\sum \lambda_i^2} \prod_{i<j} (\lambda_i - \lambda_j)^2 \prod_{i=1}^N d\lambda_i. \end{aligned}$$

Now for $(\delta_{\lambda_1(A)} + \cdots + \delta_{\lambda_N(A)})/N \approx \nu$,

$$-\frac{N}{2}\sum_{i=1}^N \lambda_i^2 = -\frac{N^2}{2}\frac{1}{N}\sum_{i=1}^N \lambda_i^2$$

is a Riemann sum for the integral $\int t^2 d\nu(t)$. Moreover

$$\prod_{i<j} (\lambda_i - \lambda_j)^2 = \exp\left(\sum_{i<j} \log|\lambda_i - \lambda_j|^2\right) = \exp\left(\sum_{i \neq j} \log|\lambda_i - \lambda_j|\right)$$

is a Riemann sum for $N^2 \int\int \log|s-t| d\nu(s) d\nu(t)$.

Hence, heuristically, we expect that $P_N(\mu_A \approx \nu) \sim \exp(-N^2 I(\nu))$, with

$$I(\nu) = -\int\int \log|s-t| d\nu(s)d\nu(t) + \frac{1}{2}\int t^2 d\nu(t) - \lim_{N\to\infty}\frac{1}{N^2}\log C_N. \qquad (7.24)$$

The value of the limit can be explicitly computed as $3/4$. Note that by writing

$$s^2 + t^2 - 4\log|s-t| = s^2 + t^2 - 2\log(s^2+t^2) + 4\log\frac{\sqrt{s^2+t^2}}{|s-t|}$$

and using the inequalities

$$t - 2\log t \geq 2 - 2\log 2 \quad \text{for } t > 0 \qquad \text{and} \qquad 2(s^2+t^2) \geq (s-t)^2$$

we have for $s \neq t$ that $s^2 + t^2 - 4\log|s-t| \geq 2 - 4\log 2$. This shows that if ν has a finite second moment, the integral $\int\int(s^2+t^2-4\log|s-t|)\, d\nu(s)d\nu(t)$ is always defined as an extended real number, possibly $+\infty$, in which case we set $I(\nu) = +\infty$, otherwise $I(\nu)$ is finite and is given by (7.24).

Voiculescu was thus motivated to use the integral $\int\int \log|s-t|\, d\mu_x(s)d\mu_x(t)$ to define in [181] the free entropy $\chi(x)$ for one self-adjoint variable x with distribution μ_x; see equation (7.30).

The large deviation argument was then made rigorous in the following theorem of Ben Arous and Guionnet [26].

Theorem 3. *Put*

$$I(\nu) = -\int\int \log|s-t| d\nu(s)d\nu(t) + \frac{1}{2}\int t^2 d\nu(t) - \frac{3}{4}. \qquad (7.25)$$

Then,

(*i*) $I : \mathcal{M} \to [0,\infty]$ *is a well-defined, convex, good function on the space, $\mathcal{M}$, of probability measures on $\mathbb{R}$. It has unique minimum value of* 0 *which occurs at the Wigner semi-circle distribution μ_W with variance* 1.

(*ii*) *The empirical eigenvalue distribution satisfies a large deviation principle with respect to $\tilde{P}_N$ with rate function I: we have for any open set G in $\mathcal{M}$*

$$\liminf_{N\to\infty}\frac{1}{N^2}\log\tilde{P}_N\Big(\frac{\delta_{\lambda_1}+\cdots+\delta_{\lambda_N}}{N} \in G\Big) \geq -\inf_{\nu\in G} I(\nu), \qquad (7.26)$$

and for any closed set F in $\mathcal{M}$

$$\limsup_{N\to\infty}\frac{1}{N^2}\log\tilde{P}_N\Big(\frac{\delta_{\lambda_1}+\cdots+\delta_{\lambda_N}}{N} \in F\Big) \leq -\inf_{\nu\in F} I(\nu). \qquad (7.27)$$

Exercise 4. The above theorem includes in particular the statement that for a Wigner semi-circle distribution μ_W with variance 1, we have

$$-\iint \log|s-t|\, d\mu_W(s)d\mu_W(t) = \frac{1}{4}. \tag{7.28}$$

Prove this directly!

Exercise 5.

(i) Let μ be a probability measure with support in $[-2,2]$. Show that we have

$$\int_{\mathbb{R}}\int_{\mathbb{R}} \log|s-t|d\mu(s)d\mu(t) = -\sum_{n=1}^{\infty}\frac{1}{2n}\left(\int_{\mathbb{R}} C_n(t)d\mu(t)\right)^2,$$

where C_n are the Chebyshev polynomials of the first kind.

(ii) Use part (i) to give another derivation of (7.28).

7.5 Definition of multivariate free entropy

Let (M,τ) be a tracial W^*-probability space and $x_1,\dots,x_n$ self-adjoint elements in M. Recall that by definition the joint distribution of the non-commutative random variables $x_1,\dots,x_n$ is the collection of all mixed moments

$$\operatorname{distr}(x_1,\dots,x_n) = \{\tau(x_{i_1}x_{i_2}\cdots x_{i_k}) \mid k\in\mathbb{N}, i_1,\dots,i_k\in\{1,\dots,n\}\}.$$

In this section we want to examine the probability that the distribution of $(x_1,\dots,x_n)$ occurs in Voiculescu's multivariable generalization of Wigner's semi-circle law.

Let $A_1,\dots,A_n$ be independent Gaussian random matrices: $A_1,\dots,A_n$ are chosen independently at random from the sample space $M_N(\mathbb{C})_{sa}$ of $N\times N$ self-adjoint matrices over $\mathbb{C}$, equipped with Gaussian probability measure having density proportional to $\exp(-\operatorname{Tr}(A^2)/2)$ with respect to Lebesgue measure on $M_N(\mathbb{C})_{sa}$. We know that as $N\to\infty$ we have almost sure convergence $(A_1,\dots,A_n) \xrightarrow{\text{distr}} (s_1,\dots,s_n)$ with respect to the normalized trace, where $(s_1,\dots,s_n)$ is a free semi-circular family. Large deviations from this limit should be given by

$$P_N\{(A_1,\dots,A_n)\mid \operatorname{distr}(A_1,\dots,A_n)\approx \operatorname{distr}(x_1,\dots,x_n)\}\sim e^{-N^2 I(x_1,\dots,x_n)},$$

where $I(x_1,\dots,x_n)$ is the *free entropy* of $x_1,\dots,x_n$. The problem is that this has to be made more precise and that, in contrast to the one-dimensional case, there is no analytical formula to calculate this quantity.

We use the equation above as motivation to define free entropy as follows. This is essentially the definition of Voiculescu from [182]; the only difference is that he also included a cut-off parameter R and required in the definition of the "microstate set" Γ that $\|A_i\| \le R$ for all $i = 1, \ldots, n$. Later it was shown by Belinschi and Bercovici [20] that removing this cut-off condition gives the same quantity.

Definition 4. Given a tracial W^*-probability space (M, τ) and an n-tuple $(x_1, \ldots, x_n)$ of self-adjoint elements in M, we define the *(microstates) free entropy* $\chi(x_1, \ldots, x_n)$ of the variables $x_1, \ldots, x_n$ as follows. First, we put

$$\begin{aligned}
&\Gamma(x_1, \ldots, x_n; N, r, \epsilon) \\
&\quad := \big\{(A_1, \ldots, A_n) \in M_N(\mathbb{C})^n_{sa} \mid |\mathrm{tr}(A_{i_1} \cdots A_{i_k}) - \tau(x_{i_1} \cdots x_{i_k})| \le \epsilon \\
&\qquad\qquad\qquad\qquad \text{for all } 1 \le i_1, \ldots, i_k \le n, 1 \le k \le r\big\}.
\end{aligned}$$

In words, $\Gamma(x_1, \ldots, x_n; N, r, \epsilon)$, which we call the *set of microstates*, is the set of all n-tuples of $N \times N$ self-adjoint matrices which approximate the mixed moments of the self-adjoint elements $x_1, \ldots, x_n$ of length at most r to within ϵ.

Let Λ denote Lebesgue measure on $M_N(\mathbb{C})^n_{sa} \simeq \mathbb{R}^{nN^2}$. Then we define

$$\chi(x_1, \ldots, x_n; r, \epsilon) := \limsup_{N \to \infty} \left(\frac{1}{N^2} \log \big(\Lambda(\Gamma(x_1, \ldots, x_n; N, r, \epsilon))\big) + \frac{n}{2} \log(N) \right),$$

and finally put

$$\chi(x_1, \ldots, x_n) := \lim_{\substack{r \to \infty \\ \epsilon \to 0}} \chi(x_1, \ldots, x_n; r, \epsilon). \tag{7.29}$$

It is an important open problem whether the lim sup in the definition above of $\chi(x_1, \ldots, x_n; r, \epsilon)$ is actually a limit.

We want to elaborate on the meaning of Λ, the Lebesgue measure on $M_N(\mathbb{C})^n_{sa} \simeq \mathbb{R}^{nN^2}$, and the normalization constant $n \log(N)/2$. Let us consider the case $n = 1$. For a self-adjoint matrix $A = (a_{ij})_{i,j=1}^N \in M_N(\mathbb{C})_{sa}$, we identify the elements on the diagonal (which are real) and the real and imaginary part of the elements above the diagonal (which are the adjoints of the corresponding elements below the diagonals) with an $N + 2\frac{N(N-1)}{2} = N^2$ dimensional vector of real numbers. The actual choice of this mapping is determined by the fact that we want the Euclidean inner product in $\mathbb{R}^{N^2}$ to correspond on the side of the matrices to the form $(A, B) \mapsto \mathrm{Tr}(AB)$. Note that

$$\mathrm{Tr}(A^2) = \sum_{i,j=1}^N a_{ij} a_{ji} = \sum_{i=1}^N (\mathrm{Re} a_{ii})^2 + 2 \sum_{1 \le i < j \le N} \big((\mathrm{Re} a_{ij})^2 + (\mathrm{Im} a_{ij})^2\big).$$

This means that there is a difference of a factor $\sqrt{2}$ between the diagonal and the off-diagonal elements. (The same effect made its appearance in Chapter 1, Exercise 8, when we defined the GUE by assigning different values for the covariances for variables on and off the diagonal – in order to make this choice invariant under conjugation by unitary matrices.) So our specific choice of a map between $M_N(\mathbb{C})$ and $\mathbb{R}^{N^2}$ means that we map the set $\{A \in M_N(\mathbb{C})_{sa} \mid \mathrm{Tr}(A^2) \leq R^2\}$ to the ball $B_{N^2}(R)$ of radius R in N^2 real dimensions. The pull back under this map of the Lebesgue measure on $\mathbb{R}^{N^2}$ is what we call Λ, the Lebesgue measure on $M_N(\mathbb{C})_{sa}$. The situation for general n is given by taking products.

Note that a microstate $(A_1, \ldots, A_n) \in \Gamma(x_1, \ldots, x_n; N, r, \epsilon)$ satisfies for $r \geq 2$

$$\frac{1}{N}\mathrm{Tr}(A_1^2 + \cdots + A_n^2) \leq \tau(x_1^2 + \cdots + x_n^2) + n\epsilon =: c^2,$$

and thus the set of microstates $\Gamma(x_1, \ldots, x_n; N, r, \epsilon)$ is contained in the ball $B_{nN^2}(\sqrt{N}c)$. The fact that the latter grows logarithmically like

$$\frac{1}{N^2}\log \Lambda\left(B_{nN^2}(\sqrt{N}c)\right) = \frac{1}{N^2}\log\frac{(\sqrt{N}c\sqrt{\pi})^{nN^2}}{\Gamma(1 + nN^2/2)} \sim -\frac{n}{2}\log N,$$

is the reason for adding the term $n \log N/2$ in the definition of $\chi(x_1, \ldots, x_n; r, \epsilon)$.

7.6 Some important properties of χ

The free entropy has the following properties:

(i) For $n = 1$, much more can be said than for general n. In particular, one can show that the lim sup in the definition of χ is indeed a limit and that we have the explicit formula

$$\chi(x) = \int\int \log|s - t| d\mu_x(s) d\mu_x(t) + \frac{1}{2}\log(2\pi) + \frac{3}{4}. \tag{7.30}$$

Thus the definition of χ reduces in this case to the quantity from the previous section. Our discussion before Theorem 3 shows then that $\chi(x) \in [-\infty, \infty)$. For $n \geq 2$, no formula of this sort is known.

When x is a semi-circular operator with variance 1, we know the value of the double integral by (7.28); hence, for a semi-circular operator s with variance 1, we have

$$\chi(s) = \frac{1}{2}(1 + \log(2\pi)). \tag{7.31}$$

(ii) χ is subadditive:

$$\chi(x_1, \ldots, x_n) \leq \chi(x_1) + \cdots + \chi(x_n). \tag{7.32}$$

This is an easy consequence of the fact that

$$\Gamma(x_1,\ldots,x_n;N,r,\epsilon)\subset\prod_{i=1}^{n}\Gamma(x_i;N,r,\epsilon).$$

Thus, in particular, by using the corresponding property from (*i*), we always have $\chi(x_1,\ldots,x_n)\in[-\infty,\infty)$.

(*iii*) χ is upper semicontinuous: if $(x_1^{(m)},\ldots,x_n^{(m)})\stackrel{\text{distr}}{\longrightarrow}(x_1,\ldots,x_n)$ for $m\to\infty$, then

$$\chi(x_1,\ldots,x_n)\geq\limsup_{m\to\infty}\chi(x_1^{(m)},\ldots,x_n^{(m)}). \tag{7.33}$$

This is because if, for arbitrary words of length k with $1\leq k\leq r$, we have

$$|\tau(x_{i_1}^{(m)}\cdots x_{i_k}^{(m)})-\tau(x_{i_1}\cdots x_{i_k})|<\frac{\epsilon}{2}$$

for sufficiently large m, then

$$\Gamma(x_1^{(m)},\ldots,x_n^{(m)};N,r,\frac{\epsilon}{2})\subset\Gamma(x_1,\ldots,x_n;N,r,\epsilon).$$

(*iv*) If $x_1,\ldots,x_n$ are free, then $\chi(x_1,\ldots,x_n)=\chi(x_1)+\cdots+\chi(x_n)$.

(*v*) $\chi(x_1,\ldots,x_n)$, under the constraint $\sum\tau(x_i^2)=n$, has a unique maximum when $x_1,\ldots,x_n$ is a free semi-circular family $(s_1,\ldots,s_n)$ with $\tau(s_i^2)=1$. In this case

$$\chi(s_1,\ldots,s_n)=\frac{n}{2}(1+\log(2\pi)). \tag{7.34}$$

(*vi*) Consider $y_j=F_j(x_1,\ldots,x_n)$, for some "convergent" non-commutative power series F_j, such that the mapping $(x_1,\ldots,x_n)\mapsto(y_1,\ldots,y_n)$ can be inverted by some other power series. Then

$$\chi(y_1,\ldots,y_n)=\chi(x_1,\ldots,x_n)+n\log(|\det|\mathcal{J}(x_1,\ldots,x_n)), \tag{7.35}$$

where $\mathcal{J}$ is a non-commutative Jacobian and $|\det|$ is the Fuglede-Kadison determinant. (We will provide more information on the Fuglede-Kadison determinant in Chapter 11.)

With the exception of (*ii*) and (*iii*), the statements above are quite non-trivial; for the proofs we refer to the original papers of Voiculescu [182, 186].

Exercise 6. (*i*) For an n-tuple $(x_1,\ldots,x_n)$ of self-adjoint elements in M and an invertible real matrix $T=(t_{ij})_{i,j=1}^n\in M_n(\mathbb{R})$, we put $y_i:=\sum_{j=1}^n t_{ij}x_j\in M$

($i = 1, \dots, n$). Part (*vi*) of the above says then (by taking into account the meaning of the Fuglede-Kadison determinant for matrices, see (11.4)) that

$$\chi(y_1, \dots, y_n) = \chi(x_1, \dots, x_n) + \log |\det T|. \tag{7.36}$$

Prove this directly from the definitions.

(*ii*) Show that $\chi(x_1, \dots, x_n) = -\infty$ if $x_1, \dots, x_n$ are linearly dependent.

7.7 Applications of free entropy to operator algebras

One hopes that χ can be used to construct invariants for von Neumann algebras. In particular, we define the *free entropy dimension* of the n-tuple $x_1, \dots, x_n$ by

$$\delta(x_1, \dots, x_n) = n + \limsup_{\epsilon \searrow 0} \frac{\chi(x_1 + \epsilon s_1, \dots, x_n + \epsilon s_n)}{|\log \epsilon|}, \tag{7.37}$$

where $s_1, \dots, s_n$ is a free semi-circular family, free from $\{x_1, \dots, x_n\}$.

One of the main problems in this context is to establish the validity (or falsehood) of the following implication (or some variant thereof): if $\mathrm{vN}(x_1, \dots, x_n) = \mathrm{vN}(y_1, \dots, y_n)$, does this imply that $\delta(x_1, \dots, x_n) = \delta(y_1, \dots, y_n)$?

In recent years there have been a number of results which allow one to infer some properties of a von Neumann algebra from knowledge of the free entropy dimension for some generators of this algebra. Similar statements can be made on the level of the free entropy. However, there the actual value of χ is not important; the main issue is to distinguish finite values of χ from the situation $\chi = -\infty$.

Let us note that in the case of free group factors $\mathcal{L}(\mathbb{F}_n) = \mathrm{vN}(s_1, \dots, s_n)$, we have of course for the canonical generators $\chi(s_1, \dots, s_n) > -\infty$ and $\delta(s_1, \dots, s_n) = n$. (For the latter one should notice that the sum of two free semi-circulars is just another semi-circular, where the variances add; hence the numerator in (7.37) stays bounded for $\epsilon \to 0$ in this case.)

We want now to give the idea how to use free entropy to get statements about a von Neumann algebra. For this, let P be some property that a von Neumann algebra M may or may not have. Assume that we can verify that "M has P" implies that $\chi(x_1, \dots, x_n) = -\infty$ for any generating set $\mathrm{vN}(x_1, \dots, x_n) = M$. Then a von Neumann algebra for which we have at least one generating set with finite free entropy cannot have this property P. In particular, $\mathcal{L}(\mathbb{F}_n)$ cannot have P.

Three such properties where this approach was successful are property Γ, the existence of a Cartan subalgebra, and the property of being prime.

Let us first recall the definition of property Γ. We will use here the usual non-commutative L^2-norm, $\|x\|_2 := \sqrt{\tau(x^*x)}$, for elements x in our tracial W^*-probability space (M, τ).

Definition 5. A bounded sequence $(t_k)_{k \geq 0}$ in (M, τ) is *central* if $\lim_{k\to\infty} \|[x, t_k]\|_2 = 0$ for all $x \in M$, where $[\cdot, \cdot]$ denotes the commutator of two elements, i.e. $[x, t_k] = xt_k - t_kx$. If $(t_k)_k$ is a central sequence and $\lim_{k\to\infty} \|t_k - \tau(t_k)1\|_2 = 0$, then $(t_k)_k$ is said to be a *trivial central sequence*. (M, τ) has *property Γ* if there exists a non-trivial central sequence in M.

Note that elements from the centre of an algebra always give central sequences; hence if M does not have property Γ, then it is a factor.

Definition 6. 1) Given any von Neumann subalgebra N of a von Neumann algebra M, we let the normalizer of N be the von Neumann subalgebra of M generated by all the unitaries $u \in M$ which normalize N, i.e. $uNu^* = N$. A von Neumann subalgebra N of M is said to be maximal abelian if it is abelian and is not properly contained in any other abelian subalgebra. A maximal abelian subalgebra is a *Cartan subalgebra* of M if its normalizer generates M.

2) Finally we recall that a finite von Neumann algebra M is *prime* if it cannot be decomposed as $M = M_1 \overline{\otimes} M_2$ for II_1 factors M_1 and M_2. Here $\overline{\otimes}$ denotes the von Neumann tensor product of M_1 and M_2; see [170, Ch. IV].

The above-mentioned strategy is the basis of the proof of the following theorem:

Theorem 7. *Let M be a finite von Neumann algebra with trace τ generated by self-adjoint operators $x_1, \dots, x_n$, where $n \geq 2$. Assume that $\chi(x_1, \dots, x_n) > -\infty$, where the free entropy is calculated with respect to the trace τ. Then*

(i) *M does not have property Γ. In particular, M is a factor.*
(ii) *M does not have a Cartan subalgebra.*
(iii) *M is prime.*

Corollary 8. *All this applies in the case of the free group factor $\mathcal{L}(\mathbb{F}_n)$ for $2 \leq n < \infty$; thus,*

(i) *$\mathcal{L}(\mathbb{F}_n)$ does not have property Γ.*
(ii) *$\mathcal{L}(\mathbb{F}_n)$ does not have a Cartan subalgebra.*
(iii) *$\mathcal{L}(\mathbb{F}_n)$ is prime.*

Parts (i) and (ii) of the theorem above are due to Voiculescu [185]; part (iii) was proved by Liming Ge [76]. In particular, the absence of Cartan subalgebras for $\mathcal{L}(\mathbb{F}_n)$ was a spectacular result, as it falsified the conjecture, which had been open for decades, that every II_1 factor should possess a Cartan subalgebra. Such a conjecture was suggested by the fact that von Neumann algebras obtained from ergodic measurable relations always have Cartan subalgebras, and for a while there was the hope that all von Neumann algebras might arise in this way.

In order to give a more concrete idea of this approach, we will present the essential steps in the proof for part (i) (which is the simplest part of the theorem above) and say a few words about the proof of part (iii). However, one should note that the absence of property Γ for $\mathcal{L}(\mathbb{F}_n)$ is an old result of Murray and von Neumann which can be proved more directly without using free entropy. The following follows quite closely the exposition of Biane [36].

7.7.1 The proof of Theorem 7, part (*i*)

We now give the main arguments and estimates for the proof of part (i) of Theorem 7. So let $M = \mathrm{vN}(x_1, \ldots, x_n)$ have property Γ; we must prove that this implies $\chi(x_1, \ldots, x_n) = -\infty$.

Let $(t_k)_k$ be a non-trivial central sequence in M. Then its real and imaginary parts are also central sequences (at least one of them non-trivial), and, by applying functional calculus to this sequence, we may replace the t_k's with a non-trivial central sequence of orthogonal projections $(p_k)_k$, and assume the existence of a real number θ in the open interval $(0, 1/2)$ such that $\theta < \tau(p_k) < 1 - \theta$ for all k and $\lim_{k\to\infty} \|[x, p_k]\|_2 = 0$ for all $x \in M$.

We then prove the following key lemma.

Lemma 9. *Let (M, τ) be a tracial W^*-probability space generated by self-adjoint elements $x_1, \ldots, x_n$ satisfying $\tau(x_i^2) \le 1$. Let $0 < \theta < \frac{1}{2}$ be a constant and $p \in M$ a projection such that $\theta < \tau(p) < 1 - \theta$. If there is $\omega > 0$ such that $\|[p, x_i]\|_2 < \omega$ for $1 \le i \le n$, then there exist positive constants C_1, C_2 depending only on n and θ such that $\chi(x_1, \ldots, x_n) \le C_1 + C_2 \log \omega$.*

Assuming this is proved, choose $p = p_k$. We can take $\omega_k \to 0$ as $k \to \infty$. Thus we get $\chi(x_1, \ldots, x_n) \le C_1 + C_2 \log \omega$ for all $\omega > 0$, implying $\chi(x_1, \ldots, x_n) = -\infty$. (Note that we can achieve the assumption $\tau(x_i^2) \le 1$ by rescaling our generators.) It remains to prove the lemma.

Proof: Take $(A_1, \ldots, A_n) \in \Gamma(x_1, \ldots, x_n; N, r, \epsilon)$ for N, r sufficiently large and ϵ sufficiently small. As p can be approximated by polynomials in $x_1, \ldots, x_n$ and by an application of the functional calculus, we find a projection matrix $Q \in M_N(\mathbb{C})$ whose range is a subspace of dimension $q = \lfloor N\tau(p) \rfloor$ and such that we have (where the $\|\cdot\|_2$-norm is now with respect to tr in $M_N(\mathbb{C})$) $\|[A_i, Q]\|_2 < 2\omega$ for all $i = 1, \ldots, n$. This Q is of the form

$$Q = U \begin{pmatrix} I_q & 0 \\ 0 & 0_{N-q} \end{pmatrix} U^*$$

for some $U \in \mathcal{U}(N)/\mathcal{U}(q) \times \mathcal{U}(N-q)$. Write

$$U^* A_i U = \begin{pmatrix} B_i & C_i^* \\ C_i & D_i \end{pmatrix}.$$

Then $\|[A_i, Q]\|_2 \le 2\omega$ implies the same for the conjugated matrices, i.e.

$$\sqrt{\frac{2}{N}\mathrm{Tr}(C_i C_i^*)} = \left\| \begin{pmatrix} 0 & -C_i^* \\ C_i & 0 \end{pmatrix} \right\|_2 = \left\| \left[\begin{pmatrix} B_i & C_i^* \\ C_i & D_i \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \right] \right\|_2 = \|[A_i, Q]\|_2 < 2\omega,$$

and thus we have for all $i = 1, \ldots, n$

$$\mathrm{Tr}(C_i C_i^*) < \frac{N}{2}(2\omega)^2 = 2N\omega^2.$$

Furthermore, $\tau(x_i^2) \leq 1$ implies that $\mathrm{tr}(A_i^2) \leq 1+\epsilon$ and hence $\mathrm{Tr}(A_i^2) \leq (1+\epsilon)N \leq 2N$, since we can take $\epsilon \leq 1$. Thus, in particular, we also have $\mathrm{Tr}(B_i^2) \leq 2N$ and $\mathrm{Tr}(D_i^2) \leq 2N$.

Denote now by $B_p(R)$ the ball of radius R in $\mathbb{R}^p$ centred at the origin and consider the map which sends our matrices $A_i \in M_N(\mathbb{C})$ to the Euclidean space $\mathbb{R}^{N^2}$. Then the latter conditions mean that each B_i is contained in a ball $B_{q^2}(\sqrt{2N})$ and that each D_i is contained in a ball $B_{(N-q)^2}(\sqrt{2N})$. For the rectangular $q \times (N-q)$ matrix $C_i \in M_{q,N-q}(\mathbb{C}) \simeq \mathbb{R}^{2q(N-q)}$, the condition $\mathrm{Tr}(CC^*) \leq 2N\omega^2$ means that C is contained in a ball $B_{2q(N-q)}(\sqrt{4N}\omega)$. (Here we get an extra factor $\sqrt{2}$, because all elements from C_i correspond to upper triangular elements from A_i.)

Thus, the estimates above show that we can cover $\Gamma(x_1, \ldots, x_n; N, r, \epsilon)$ by a union of products of balls:

$$\Gamma(x_1, \ldots, x_n; N, r, \epsilon) \subseteq \bigcup_{\substack{U \in \\ \mathcal{U}(N)/\mathcal{U}(q)\times\mathcal{U}(N-q)}} \left[U \left(B_{q^2}(\sqrt{2N}) \times B_{2q(N-q)}(\omega\sqrt{4N}) \times B_{(N-q)^2}(\sqrt{2N}) \right) U^* \right]^n.$$

This does not give directly an estimate for the volume of our set Γ, as we have here a covering by infinitely many sets. However, we can reduce this to a finite cover by approximating the U's which appear by elements from a finite δ-net.

By a result of Szarek [169], for any $\delta > 0$, there exists a δ-net $(U_s)_{s\in S}$ in the Grassmannian $\mathcal{U}(N)/\mathcal{U}(q) \times \mathcal{U}(N-q)$ with $|S| \leq (C\delta^{-1})^{N^2-q^2-(N-q)^2}$ with C a universal constant.

For $(A_1, \ldots, A_n)$, Q, and U as above, we have that there exists $s \in S$ such that $\|U - U_s\| \leq \delta$ implies $\|[U_s^* A_i U_s, U^* Q U]\|_2 \leq 2\omega + 8\delta$. Repeating the arguments above for $U_s^* A_i U_s$ instead of $U^* A_i U$ (where we have to replace 2ω by $2\omega + 8\delta$), we get

$$\begin{aligned}&\Gamma(x_1, \ldots, x_n; N, r, \epsilon)\\ &\subseteq \bigcup_{s\in S} \left[U_s \left(B_{q^2}(\sqrt{2N}) \times B_{2q(N-q)}\big((\omega + 4\delta)\sqrt{4N}\big) \times B_{(N-q)^2}(\sqrt{2N}) \right) U_s^* \right]^n,\end{aligned} \tag{7.38}$$

and hence

$$\Lambda(\Gamma(x_1,\dots,x_n;N,r,\epsilon)) \le (C\delta^{-1})^{N^2-q^2-(N-q)^2}$$
$$\times\left[\Lambda\left(B_{q^2}(\sqrt{2N})\right)\Lambda\left(B_{2q(N-q)}\big((\omega+4\delta)\sqrt{4N}\big)\right)\Lambda\left(B_{(N-q)^2}(\sqrt{2N})\right)\right]^n.$$

By using the explicit form of the Lebesgue measure of $B_p(R)$ as

$$\Lambda(B_p(R)) = \frac{R^p\pi^{p/2}}{\Gamma(1+\frac{p}{2})},$$

this simplifies to the bound

$$(C\delta^{-1})^{2q(N-q)}\left[\frac{(2N\pi)^{N^2/2}[\sqrt{2}(\omega+4\delta)]^{2q(N-q)}}{\Gamma(1+q^2/2)\Gamma(1+q(N-q))\Gamma(1+(N-q)^2/2)}\right]^n.$$

Thus

$$\frac{1}{N^2}\log\Lambda(\Gamma(x_1,\dots,x_n;N,r,\epsilon))+\frac{n}{2}\log N \le \tilde{C}_1+\tilde{C}_2\big(\log\delta^{-1}+n\log(\omega+4\delta)\big),$$

for positive constants $\tilde{C}_1, \tilde{C}_2$ depending only on n and θ. Taking now $\delta=\omega$ gives the claimed estimate with $C_1 := \tilde{C}_1 + n\log 5$ and $C_2 := (n-1)\tilde{C}_2$. □

One should note that our estimates work for all n. However, in order to have C_2 strictly positive, we need $n>1$. For $n=1$ we only get an estimate against a constant C_1, which is not very useful. This corresponds to the fact that for each i the smallness of the off-diagonal block C_i of U^*A_iU in some basis U is not very surprising; however, if we have the smallness of all such blocks $C_1,\dots,C_n$ of $U^*A_1U,\dots,U^*A_nU$ for a common U, then this is a much stronger constraint.

7.7.2 The proof of Theorem 7, part (*iii*)

The proof of part (*iii*) proceeds in a similar, though technically more complicated, fashion. Let us assume that our II_1 factor $M=\mathrm{vN}(x_1,\dots,x_n)$ has a Cartan subalgebra N. We have to show that this implies $\chi(x_1,\dots,x_n)=-\infty$.

First one has to rewrite the property of having a Cartan subalgebra in a more algebraic way, encoding a kind of "smallness". Voiculescu showed the following. For each $\epsilon>0$, there exist a finite-dimensional C^*-subalgebra N_0 of N; $k(j)\in\mathbb{N}$ for all $1\le j\le n$; orthogonal projections $p_j^{(i)}, q_j^{(i)}\in N_0$ and elements $x_j^{(i)}\in M$ for all $j = 1, \ldots, n$ and $1 \leq i \leq k(j)$ such that the following holds: $x_j^{(i)} = p_j^{(i)} x_j^{(i)} q_j^{(i)}$ for all $j = 1, \ldots, n$ and $1 \leq i \leq k(j)$,

$$\|x_j - \sum_{1 \leq i \leq k(j)} (x_j^{(i)} + x_j^{(i)*})\|_2 < \epsilon \qquad \text{for all } j = 1, \ldots, n, \tag{7.39}$$

and

$$\sum_{1 \leq j \leq n} \sum_{1 \leq i \leq k(j)} \tau(p_j^{(i)}) \tau(q_j^{(i)}) < \epsilon.$$

Consider now a microstate $(A_1, \ldots, A_n) \in \Gamma(x_1, \ldots, x_n; N, r, \epsilon)$. Since polynomials in the generators $x_1, \ldots, x_n$ approximate the given projections $p_j^{(i)}, q_j^{(i)} \in N_0 \subset M$, the same polynomials in the matrices $A_1, \ldots, A_n$ will approximate versions of these projections in finite matrices. Thus we find a unitary matrix such that $(UA_1U^*, \ldots, UA_nU^*)$ is of a special form with respect to fixed matrix versions of the projections. This gives some constraints on the volume of possible microstates. Again, in order to get rid of the freedom of conjugating by an arbitrary unitary matrix, one covers the unitary $N \times N$ matrices by a δ-net S and gets so in the end a similar bound as in (7.38). Invoking from [169] the result that one can choose a δ-net with $|S| < (C/\delta)^{N^2}$ leads finally to an estimate for $\chi(x_1, \ldots, x_n)$ as in Lemma 9. The bound in this estimate goes to $-\infty$ for $\epsilon \to 0$, which proves that $\chi(x_1, \ldots, x_n) = -\infty$. □

Chapter 8
Free Entropy χ^*: The Non-microstates Approach via Free Fisher Information

In classical probability theory, there exist two important concepts which measure the amount of "information" of a given distribution. These are the Fisher information and the entropy. There exist various relations between these quantities, and they form a cornerstone of classical probability theory and statistics. Voiculescu introduced free probability analogues of these quantities, called free Fisher information and free entropy, denoted by Φ and χ, respectively. However, there remain some gaps in our present understanding of these quantities. In particular, there exist two different approaches, each of them yielding a notion of entropy and Fisher information. One hopes that finally one will be able to prove that both approaches give the same result, but at the moment this is not clear. Thus, for the time being, we have to distinguish the entropy χ and the free Fisher information Φ coming from the first approach (via microstates) and the free entropy χ^* and the free Fisher information Φ^* coming from the second non-microstates approach (via conjugate variables).

Whereas we considered the microstates approach for χ in the previous chapter, we will in this chapter deal with the second approach, which fits quite nicely with the combinatorial theory of freeness. In this approach, the Fisher information is the basic quantity (in terms of which the free entropy χ^* is defined), so we will restrict our attention mainly to Φ^*.

The concepts of information and entropy are only useful when we consider states (so that we can use the positivity of φ to get estimates for the information or entropy). Thus, in this section, we will always work in the framework of a W^*-probability space. Furthermore, it is crucial that we work with a faithful normal trace. The extension of the present theory to non-tracial situations is unclear.

J.A. Mingo, R. Speicher, *Free Probability and Random Matrices*,
Fields Institute Monographs 35, DOI 10.1007/978-1-4939-6942-5_8

8.1 Non-commutative derivatives

In Chapter 2 we already encountered non-commutative derivatives, on an informal level, in connection with the subordination property of free convolution. Here we will introduce and investigate these non-commutative derivatives more thoroughly.

Definition 1. We denote by $\mathbb{C}\langle X_1, \dots, X_n\rangle$ the algebra of polynomials in n non-commuting variables $X_1, \dots, X_n$. On this we define the *partial non-commutative derivatives* ∂_i $(i = 1, \dots, n)$ as linear mappings

$$\partial_i : \mathbb{C}\langle X_1, \dots, X_n\rangle \to \mathbb{C}\langle X_1, \dots, X_n\rangle \otimes \mathbb{C}\langle X_1, \dots, X_n\rangle$$

by

$$\partial_i 1 = 0, \qquad \partial_i X_j = \delta_{ij} 1 \otimes 1 \quad (j = 1, \dots, n),$$

and by the Leibniz rule

$$\partial_i (P_1 P_2) = \partial_i (P_1) \cdot 1 \otimes P_2 + P_1 \otimes 1 \cdot \partial_i (P_2) \qquad (P_1, P_2 \in \mathbb{C}\langle X_1, \dots, X_n\rangle).$$

This means that ∂_i is given on monomials by

$$\partial_i (X_{i(1)} \cdots X_{i(m)}) = \sum_{k=1}^{m} \delta_{i,i(k)} X_{i(1)} \cdots X_{i(k-1)} \otimes X_{i(k+1)} \cdots X_{i(m)}. \tag{8.1}$$

Example 2. Consider the monomial $P(X_1, X_2, X_3) = X_2 X_1^3 X_3 X_1$. Then, we have

$$\begin{aligned}
\partial_1 P &= X_2 \otimes X_1^2 X_3 X_1 + X_2 X_1 \otimes X_1 X_3 X_1 + X_2 X_1^2 \otimes X_3 X_1 + X_2 X_1^3 X_3 \otimes 1\\
\partial_2 P &= 1 \otimes X_1^3 X_3 X_1\\
\partial_3 P &= X_2 X_1^3 \otimes X_1.
\end{aligned}$$

Exercise 1.

(*i*) Prove, for $i \in \{1, \dots, n\}$, the co-associativity of ∂_i

$$(id \otimes \partial_i) \circ \partial_i = (\partial_i \otimes id) \circ \partial_i. \tag{8.2}$$

(*ii*) If one mixes different partial derivatives, the situation becomes more complicated. Show that $(id \otimes \partial_i) \circ \partial_j = (\partial_j \otimes id) \circ \partial_i$, but in general for $i \neq j$ $(id \otimes \partial_i) \circ \partial_j \neq (\partial_i \otimes id) \circ \partial_j$.

Proposition 3. *In the case $n = 1$, we can identify $\mathbb{C}\langle X\rangle \otimes \mathbb{C}\langle X\rangle$ with the polynomials $\mathbb{C}[X,Y]$ in two commuting variables X and Y, via $X \hat{=} X \otimes 1$ and $Y \hat{=} 1 \otimes X$. With this identification, $\partial := \partial_1$ is given by the* free difference quotient

$$\partial P(X) \hat{=} \frac{P(X) - P(Y)}{X - Y}.$$

Proof: It suffices to consider $P(X) = X^m$; then we have

$$\partial P(X) = 1 \otimes X^{m-1} + X \otimes X^{m-2} + X^2 \otimes X^{m-3} + \cdots + X^{m-1} \otimes 1$$

and

$$\frac{X^m - Y^m}{X - Y} = X^{m-1} + X^{m-2}Y + X^{m-3}Y^2 + \cdots + Y^{m-1}.$$

□

One should note that in the non-commutative world, there exists another canonical derivation into the tensor product, namely, the mapping $P \mapsto P \otimes 1 - 1 \otimes P$. Actually, there is an important relation between this derivation and our partial derivatives.

Lemma 4. *For all $P \in \mathbb{C}\langle X_1, \ldots, X_n\rangle$, we have*

$$\sum_{j=1}^{n} \partial_j P \cdot X_j \otimes 1 - 1 \otimes X_j \cdot \partial_j P = P \otimes 1 - 1 \otimes P. \tag{8.3}$$

Exercise 2. Prove Lemma 4 by checking it for monomials P.

This allows an easy proof of the following free version of a Poincaré inequality. This is an unpublished result of Voiculescu and can be found in [63].

In this inequality, we will apply our non-commutative polynomials to operators $x_1, \ldots, x_n \in M$. If $P = P(X_1, \ldots, X_n) \in \mathbb{C}\langle X_1, \ldots, X_n\rangle$, then $P(x_1, \ldots, x_n) \in M$ is obtained by replacing each of the variables X_i by the corresponding x_i. Note in particular that this applies also to the right-hand side of the inequality. There $\partial_i P$ is an element in $\mathbb{C}\langle X_1, \ldots, X_n\rangle^{\otimes 2}$, and $\partial_i P(x_1, \ldots, x_n)$ is to be understood as $(\partial_i P)(x_1, \ldots, x_n)$. As usual, $\|a\|_2 := \sqrt{\tau(a^*a)}$ denotes the non-commutative L^2-norm given by τ, and with $L^2(M)$, we denote the completion of M with respect to this norm. The L^2-norm on the right-hand side of the inequality is of course with respect to $\tau \otimes \tau$.

Theorem 5 (Free Poincaré Inequality). *Let (M, τ) be a tracial W^*-probability space. Consider self-adjoint $x_1, \ldots, x_n \in M$. Then, we have for all $P = P^* \in \mathbb{C}\langle X_1, \ldots, X_n\rangle$ the inequality*

$$\|P(x_1,\dots,x_n) - \tau(P(x_1,\dots,x_n))\|_2 \le C \cdot \sum_{i=1}^{n} \|\partial_i P(x_1,\dots,x_n)\|_2, \qquad (8.4)$$

where $C := \sqrt{2}\max_{j=1,\dots,n}\|x_j\|$.

Proof: Let us put $p := P(x_1,\dots,x_n)$ and $q_i := (\partial_i P)(x_1,\dots,x_n)$. It suffices to consider P with $\tau(p) = 0$. Then, we get from Lemma 4

$$\begin{aligned}\|p \otimes 1 - 1 \otimes p\|_2 &= \|\sum_{i=1}^{n} q_i \cdot x_i \otimes 1 - 1 \otimes x_i \cdot q_i\|_2\\ &\le \sum_{i=1}^{n}\big(\underbrace{\|q_i \cdot x_i \otimes 1\|_2}_{\le \|q_i\|_2 \cdot \|x_i \otimes 1\|} + \|1 \otimes x_i \cdot q_i\|_2\big)\\ &\le 2\max_{j=1,\dots,n}\|x_j\|\sum_{i=1}^{n}\|q_j\|_2.\end{aligned}$$

On the other hand, we have (recall that $\tau(p) = 0$)

$$\begin{aligned}\|p \otimes 1 - 1 \otimes p\|_2^2 &= \tau \otimes \tau\left[(p \otimes 1 - 1 \otimes p)^2\right]\\ &= \tau \otimes \tau[p^2 \otimes 1 + 1 \otimes p^2 - 2p \otimes p]\\ &= 2\tau(p^2)\\ &= 2\|p\|_2^2.\end{aligned}$$

□

Corollary 6. *Let (M,τ) be a tracial W^*-probability space and $x_i = x_i^* \in M$ for $i=1,\dots,n$. Consider $P = P^* \in \mathbb{C}\langle X_1,\dots,X_n\rangle$. Assume that $(\partial_i P)(x_1,\dots,x_n) = 0$ for all $i = 1,\dots,n$. Then, $p := P(x_1,\dots,x_n)$ is a constant, $p = \tau(p)\cdot 1$.*

8.2 ∂_i as unbounded operator on $\mathbb{C}\langle x_1,\dots,x_n\rangle$

Let (M,τ) be a tracial W^*-probability space and consider $x_i = x_i^* \in M$ ($i = 1,\dots,n$), and let $\mathbb{C}\langle x_1,\dots,x_n\rangle$ be the $*$-subalgebra of M generated by $x_1.\dots,x_n$. We shall continue to denote by $\mathbb{C}\langle X_1,\dots,X_n\rangle$ the algebra generated by the non-commuting variables $X_1,\dots,X_n$. We always have an evaluation map

$$eval : \mathbb{C}\langle X_1,\dots,X_n\rangle \to \mathbb{C}\langle x_1,\dots,x_n\rangle$$

which sends $X_{i_1}\cdots X_{i_k}$ to $x_{i_1}\cdots x_{i_k}$.

If the evaluation map extends to an algebra isomorphism (i.e. has a trivial kernel), then we say that the operators $x_1,\dots,x_n$ are *algebraically free*.

In the case that $x_1, \ldots, x_n$ are algebraically free, the operators ∂_i can also be defined as derivatives on $\mathbb{C}\langle x_1, \ldots, x_n \rangle \subset M$, according to the commutative diagram

$$\begin{array}{ccc}
\mathbb{C}\langle X_1, \ldots, X_n \rangle & \xrightarrow{\partial_i} & \mathbb{C}\langle X_1, \ldots, X_n \rangle \otimes \mathbb{C}\langle X_1, \ldots, X_n \rangle \\
\downarrow & & \downarrow \\
eval & & eval \\
\downarrow & & \downarrow \\
\mathbb{C}\langle x_1, \ldots, x_n \rangle & \longrightarrow & \mathbb{C}\langle x_1, \ldots, x_n \rangle \otimes \mathbb{C}\langle x_1, \ldots, x_n \rangle
\end{array}$$

In that case, we can consider ∂_i as unbounded operator on L^2.

Notation 7. *We denote by*

$$L^p(x_1, \ldots, x_n) := \overline{\mathbb{C}\langle x_1, \ldots, x_n \rangle}^{\|\cdot\|_p} \subset L^p(M)$$

the closure of $\mathbb{C}\langle x_1, \ldots, x_n \rangle \subset M$ *with respect to the* L^p *norms* ($1 \leq p < \infty$)

$$\|a\|_p^p := \tau\left(|a|^p\right) = \tau\left((a^* a)^{p/2}\right).$$

Hence, in the case where $x_1, \ldots, x_n$ are algebraically free, ∂_i is then also an unbounded operator on L^2,

$$\partial_i : L^2(x_1, \ldots, x_n) \supset D(\partial_i) \to L^2(x_1, \ldots, x_n) \otimes L^2(x_1, \ldots, x_n)$$

with domain $D(\partial_i) = \mathbb{C}\langle x_1, \ldots, x_n \rangle$. In order that unbounded operators have a nice analytic structure, they should be closable. In terms of the adjoint, this means that the adjoint operator

$$\partial_i^* : L^2(x_1, \ldots, x_n) \otimes L^2(x_1, \ldots, x_n) \supset D(\partial_i^*) \to L^2(x_1, \ldots, x_n)$$

should be densely defined. One simple way to guarantee this is to have $1 \otimes 1$ in the domain $D(\partial_i^*)$. The following theorem shows that this then implies that all of $\mathbb{C}\langle x_1, \ldots, x_n \rangle \otimes \mathbb{C}\langle x_1, \ldots, x_n \rangle$ (which is by definition dense in $L^2(x_1, \ldots, x_n) \otimes L^2(x_1, \ldots, x_n)$) is in the domain of the adjoint. The proof of this is a direct calculation, which we leave as an exercise.

Theorem 8. *Assume* $1 \otimes 1 \in D(\partial_i^*)$. *Then* ∂_i *is closable. We have*

$$\mathbb{C}\langle x_1, \ldots, x_n \rangle \otimes \mathbb{C}\langle x_1, \ldots, x_n \rangle \subset D(\partial_i^*) \tag{8.5}$$

and for elementary tensors $p \otimes q$ *with* $p, q \in \mathbb{C}\langle x_1, \ldots, x_n \rangle$, *the action of* ∂_i^* *is given by*

$$\partial_i^*(p \otimes q) = p \cdot \partial_i^*(1 \otimes 1) \cdot q - p \cdot (\tau \otimes id)(\partial_i q) - (id \otimes \tau)(\partial_i p) \cdot q. \tag{8.6}$$

In the following, we will use the notation $\xi_i := \partial_i^*(1 \otimes 1)$ $(i = 1, \dots, n)$. In the next section, we will see that the vectors ξ_i actually play a quite prominent role in the definition of the free Fisher information.

Exercise 3.

(i) On $L^2(x_1, \dots, x_n)$, we may extend the map $x \mapsto x^*$ to a bounded conjugate linear operator J, called the modular conjugation operator. For $\eta \in L^2(x_1, \dots, x_n)$ and $p \in \mathbb{C}\langle x_1, \dots, x_n \rangle$, we have $\langle J(\eta), p \rangle = \overline{\langle \eta, J(p) \rangle} = \overline{\langle \eta, p^* \rangle}$. Show that we have $\langle \xi_i, p \rangle = \overline{\langle \xi_i, p^* \rangle}$ for all $p \in \mathbb{C}\langle x_1, \dots, x_n \rangle$, and thus ξ_i is self-adjoint, i.e. $J(\xi_i) = \xi_i$.

(ii) Show that we have for all $p \in \mathbb{C}\langle x_1, \dots, x_n \rangle$ the identity

$$(\tau \otimes id)[(\partial_i p^*)^*] = (id \otimes \tau)(\partial_i p).$$

(iii) Recall that the domain of ∂_i^* is

$$D(\partial_i^*) = \{\eta \in L^2 \otimes L^2 \mid \exists \eta' \in L^2 \text{ such that } \langle \eta', r \rangle = \langle \eta, \partial_i r \rangle \ \forall r \in \mathbb{C}\langle x_1, \dots, x_n \rangle\}.$$

For such an η, we set $\partial_i^*(\eta) = \eta'$. Prove Theorem 8 by showing that for all $r \in \mathbb{C}\langle x_1, \dots, x_n \rangle$, we have $\langle \partial_i^*(p \otimes q), r \rangle = \langle p \otimes q, \partial_i r \rangle$ when we use the right-hand side of (8.6) as the definition of $\partial_i^*(p \otimes q)$.

(iv) Show that

$$\langle (id \otimes \tau)(\partial_i p), (id \otimes \tau)(\partial_i q) \rangle = \langle 1 \otimes \xi_i - \xi_i \otimes 1, \partial_i p^* \cdot 1 \otimes q \rangle.$$

for all $p, q \in \mathbb{C}\langle x_1, \dots, x_n \rangle$.

(v) Show that also the unbounded operator $(id \otimes \tau) \circ \partial_i$, with domain $\mathbb{C}\langle x_1, \dots, x_n \rangle$, is a closable operator on $L^2(x_1, \dots, x_n)$.

Although ∂_i is an unbounded operator from L^2 to L^2, it turns out that this has some surprising boundedness properties in an appropriate sense. This observation is due to Dabrowski [63]. Our presentation follows essentially his arguments.

Proposition 9. *Assume that* $1 \otimes 1 \in D(\partial_i^*)$. *Then we have for all* $p, q \in \mathbb{C}\langle x_1, \dots, x_n \rangle$ *the identity*

$$\langle \partial_i^*(p \otimes 1), \partial_i^*(q \otimes 1) \rangle = \langle \partial_i^*(1 \otimes 1), \partial_i^*(p^* q \otimes 1) \rangle \tag{8.7}$$

and thus

$$\|(id \otimes \tau)(\partial_i p) - p\xi_i\|_2^2 = \|p\xi_i\|_2^2 - \langle \xi_i \otimes 1, \partial_i(p^* p) \rangle. \tag{8.8}$$

Proof: By Eq. (8.6), we have

$$\partial_i^*(p \otimes 1) = p\xi_i - (id \otimes \tau)(\partial_i p), \qquad \partial_i^*(q \otimes 1) = q\xi_i - (id \otimes \tau)(\partial_i q)$$

and

$$\begin{aligned}\partial_i^*(p^*q \otimes 1) &= p^*q\xi_i - (id \otimes \tau)[\partial_i(p^*q)]\\ &= p^*q\xi_i - (id \otimes \tau)[\partial_i p^* \cdot 1 \otimes q] - p^* \cdot (id \otimes \tau)[\partial_i q].\end{aligned}$$

Hence, our assertion (8.7) is equivalent to

$$\begin{aligned}&\langle p\xi_i - (id \otimes \tau)(\partial_i p), q\xi_i - (id \otimes \tau)(\partial_i q)\rangle\\ &\qquad = \langle \xi_i, p^*q\xi_i - (id \otimes \tau)[\partial_i p^* \cdot 1 \otimes q)] - p^* \cdot (id \otimes \tau)[\partial_i q]\rangle.\end{aligned}$$

There are two terms which show up obviously on both sides, and thus we are left with showing

$$-\langle (id\otimes\tau)(\partial_i p), q\xi_i\rangle + \langle (id\otimes\tau)(\partial_i p), (id\otimes\tau)(\partial_i q)\rangle = -\langle \xi_i, (id\otimes\tau)[\partial_i p^* \cdot 1\otimes q]\rangle.$$

If we interpret τ as the operator from L^2 to $\mathbb{C}$ given by $\tau(\xi) = \langle \xi, 1\rangle$, then we have

$$(id \otimes \tau)^*(\xi) = \xi \otimes 1.$$

Thus,

$$\langle \xi_i, (id \otimes \tau)[\partial_i p^* \cdot 1 \otimes q]\rangle = \langle \xi_i \otimes 1, \partial_i p^* \cdot 1 \otimes q\rangle$$

and

$$\begin{aligned}\langle (id \otimes \tau)(\partial_i p), q\xi_i\rangle &= \langle \xi_i q^*, ((id \otimes \tau)[\partial_i p])^*\rangle\\ &= \langle \xi_i q^*, (\tau \otimes id)[\partial_i p^*]\rangle\\ &= \langle \xi_i, (\tau \otimes id)[\partial_i p^*] \cdot 1 \otimes q\rangle,\end{aligned}$$

then (8.7) follows from Exercise 3.

A similar calculation shows that for $r \in \mathbb{C}\langle x_1, \ldots, x_n\rangle$, we have

$$\langle p\xi_i - (id \otimes \tau)(\partial_i p), r\rangle = \langle \partial_i^*(p \otimes 1), r\rangle.$$

Thus, $p\xi_i - (id \otimes \tau)(\partial_i p) = \partial_i^*(p \otimes 1)$. This then implies Eq. (8.8) as follows:

$$\begin{aligned}\|(id \otimes \tau)(\partial_i p) - p\xi_i\|_2^2 &= \langle \partial_i^*(p \otimes 1), \partial_i^*(p \otimes 1)\rangle\\ &= \langle \xi_i, \partial_i^*(p^*p \otimes 1)\rangle\\ &= \langle \xi_i, (p^*p)\xi_i - (id \otimes \tau)[\partial_i(p^*p)]\rangle\\ &= \langle p\xi_i, p\xi_i\rangle - \langle \xi_i, (id \otimes \tau)[\partial_i(p^*p)]\rangle\\ &= \langle p\xi_i, p\xi_i\rangle - \langle \xi_i \otimes 1, \partial_i(p^*p)\rangle.\end{aligned}$$

□

Theorem 10. *Assume that* $1 \otimes 1 \in D(\partial_i^*)$. *Then we have for all* $p \in \mathbb{C}\langle x_1, \dots, x_n\rangle$ *the inequality*

$$\|(id \otimes \tau)(\partial_i p) - p\xi_i\|_2 \leq \|\xi_i\|_2 \cdot \|p\|. \tag{8.9}$$

Hence, with $M = \mathrm{vN}(x_1, \dots, x_n)$, *the mapping* $(id \otimes \tau) \circ \partial_i$ *extends to a bounded mapping* $M \to L^2(M)$, *and we have*

$$\|(id \otimes \tau) \circ \partial_i\|_{M \to L^2(M)} \leq 2\|\xi_i\|_2. \tag{8.10}$$

Proof: Assume that inequality (8.9) has been proved. Then we have

$$\|(id \otimes \tau)\partial_i p\|_2 \leq \|\xi_i\|_2 \cdot \|p\| + \|p\xi_i\|_2 \leq 2\|\xi_i\|_2 \cdot \|p\|$$

for all $p \in \mathbb{C}\langle x_1, \dots, x_n\rangle$. This says that $(id \otimes \tau) \circ \partial_i$ as a linear mapping from $\mathbb{C}\langle x_1, \dots, x_n\rangle \subset M$ to $L^2(M)$ has norm less or equal to $2\|\xi_i\|_2$. It is also easy to check (see Exercise 3) that $(id \otimes \tau) \circ \partial_i$ is closable as an unbounded operator from L^2 to L^2, and, hence, by the following Proposition 11, it can be extended to a bounded mapping on M, with the same bound: $2\|\xi_i\|_2$.

So it remains to prove (8.9). By (8.8), we have

$$\begin{aligned}\|(id \otimes \tau)\partial_i p - p\xi_i\|_2^2 &= \langle \partial_i^*(p \otimes 1), \partial_i^*(p \otimes 1)\rangle \\ &= \langle \xi_i, (p^* p)\xi_i - (id \otimes \tau)(\partial_i(p^* p))\rangle \\ &\leq \|\xi_i\|_2 \cdot \|(id \otimes \tau)(\partial_i(p^* p)) - (p^* p)\xi_i\|_2.\end{aligned}$$

So, by iteration we get

$$\begin{aligned}&\|(id \otimes \tau)(\partial_i p) - p\xi_i\|_2 \leq \|\xi_i\|_2^{1/2} \cdot \|(id \otimes \tau)(\partial_i(p^* p)) - (p^* p)\xi_i\|_2^{1/2} \\ &\quad \leq \|\xi_i\|_2^{1/2} \cdot \|\xi_i\|_2^{1/4} \cdot \|(id \otimes \tau)(\partial_i(p^* p)^2) - (p^* p)^2\xi_i\|_2^{1/4} \\ &\quad \leq \|\xi_i\|_2^{1/2+1/4+\cdots+1/2^n} \cdot \|(id \otimes \tau)(\partial_i(p^* p)^{2^{n-1}}) - (p^* p)^{2^{n-1}}\xi_i\|_2^{1/2^n}.\end{aligned}$$

Now note that the first factor converges, for $n \to \infty$, to $\|\xi_i\|_2$, whereas for the second factor, we can bound as follows:

$$\begin{aligned}&\|(id \otimes \tau)[\partial_i((p^* p)^{2^{n-1}})] - (p^* p)^{2^{n-1}}\xi_i\|_2^{1/2^n} \\ &\qquad \leq \left(\|\partial_i((p^* p)^{2^{n-1}})\|_2 + \|p^* p\|^{2^{n-1}} \cdot \|\xi_i\|_2\right)^{1/2^n} \\ &\qquad \leq \|p\| \cdot \left(2^{n-1}\frac{\|\partial_i(p^* p)\|_2}{\|p^* p\|} + \|\xi_i\|_2\right)^{1/2^n},\end{aligned}$$

where we have used the inequality

$$\|\partial_i (p^* p)^{2^{n-1}}\|_2 \le 2^{n-1} \|p^* p\|^{2^{n-1}-1} \|\partial_i (p^* p)\|_2.$$

Sending $n \to \infty$ gives now the assertion. □

Proposition 11. *Let (M, τ) be a tracial W^*-probability space with separable predual and $\Delta : L^2(M, \tau) \supset D(\Delta) \to L^2(M, \tau)$ be a closable linear operator. Assume that $D(\Delta) \subset M$ is a $*$-algebra and that we have $\|\Delta(x)\|_2 \le c\|x\|$ for all $x \in D(\Delta)$. Then Δ extends to a bounded mapping $\Delta : M \to L^2(M, \tau)$ with $\|\Delta\|_{M \to L^2(M)} \le c$.*

Proof: Since the extension of Δ to the norm closure of $D(\Delta)$ is trivial, we can assume without restriction that $D(\Delta)$ is a C^*-algebra. Consider $y \in M$. By Kaplansky's density theorem, there exists a sequence $(x_n)_{n\in\mathbb{N}}$ with $x_n \in D(\Delta)$, $\|x_n\| \le \|y\|$ for all n such that $(x_n)_n$ converges to y in the strong operator topology. By assumption we know that the sequence $(\Delta(x_n))_n$ is bounded by $c\|y\|$ in the L^2-norm. By the Banach-Saks theorem, we have then a subsequence $(\Delta(x_{n_k}))_k$ of which the Cesàro means converge in the L^2-norm, say to some $z \in L^2(M)$:

$$z_m := \frac{1}{m} \sum_{l=1}^{m} \Delta(x_{n_l}) \to z \in L^2(M).$$

Now put $y_m := \sum_{l=1}^{m} x_{n_l}/m$. Then, we have a sequence $(y_m)_{m\in\mathbb{N}}$ that converges to y in the strong operator topology, hence also in the L^2-norm, such that $(\Delta(y_m))_m = (z_m)_m$ converges to some $z \in L^2(M)$. Since Δ is closable, this z is independent of the chosen sequences, and putting $\Delta(y) := z$ gives the extension to M we seek. Since we have $\|\Delta(y_m)\|_2 \le c\|y\|$ for all m, this goes also over to the limit: $\|\Delta(y)\|_2 = \|z\|_2 \le c\|y\|$. □

8.3 Conjugate variables and free Fisher information Φ^*

Before we give the definition of the free Fisher information, we want to motivate the form of this by having a look at the classical Fisher information.

In classical probability theory, the Fisher information $I(X)$ of a random variable X is the derivative of the entropy of a Brownian motion starting in X. Assume the probability distribution μ_X has a density p; then, the density p_t at time t of such a Brownian motion is given by the solution of the diffusion equation

$$\frac{\partial p_t(u)}{\partial t} = \frac{\partial^2 p_t(u)}{\partial u^2}$$

subject to the initial condition $p_0(u) = p(u)$. Let us calculate the derivative of the classical entropy $S(p_t)$ at $t = 0$, where we use the explicit formula for classical entropy

$$S(p_t) = -\int p_t(u)\log p_t(u)du.$$

We will in the following just do formal calculations, but all steps can be justified rigorously. We will also use the notations

$$\dot{p} := \frac{\partial}{\partial t}p, \qquad p' := \frac{\partial}{\partial u}p,$$

where $p(t,u) = p_t(u)$. Then we have

$$\frac{dS(p_t)}{dt} = -\int \frac{\partial}{\partial t}[p_t(u)\cdot\log p_t(u)]\,du = -\int [\dot{p}_t \log p_t + \dot{p}_t]\,du.$$

The second term vanishes,

$$\int \dot{p}_t\,du = \frac{d}{dt}\int p_t(u)\,du = 0$$

(because p_t is a probability density for all t); by invoking the diffusion equation and by integration by parts, the first term gives

$$-\int \dot{p}_t \log p_t\,du = -\int p_t'' \log p_t\,du = \int p_t'(\log p_t)'\,du = \int \frac{(p_t'(u))^2}{p_t(u)}\,du.$$

Taking this at $t = 0$ gives the explicit formula

$$I(X) = \int \frac{(p'(u))^2}{p(u)}\,du \qquad \text{if } d\mu_X(u) = p(u)\,du$$

for the Fisher information of X.

To get a non-commutative version of this, one first needs a conceptual understanding of this formula. For this let us rewrite it in the form

$$I(X) = \int \frac{(p'(u))^2}{p(u)}\,du = \mathrm{E}\Big[\Big(-\frac{p'}{p}(X)\Big)^2\Big] = \mathrm{E}(\xi^2),$$

where the random variable ξ (usually called the *score function*) is defined by

$$\xi := -\frac{p'}{p}(X) \qquad (\text{which is in } L^2(X) \text{ if } I(X) < \infty).$$

The advantage of this is that the score ξ has some conceptual meaning. Consider a nice $f(X) \in L^2(X)$, and calculate

$$\mathrm{E}(\xi f(X)) = -\mathrm{E}\Big[\frac{p'}{p}(X)f(X)\Big] = -\int \frac{p'(u)}{p(u)} f(u)p(u)\,du$$
$$= -\int p'(u)f(u)\,du = \int p(u)f'(u)\,du = \mathrm{E}(f'(X)).$$

In terms of the derivative operator $\dfrac{d}{du}$ and its adjoint, we can also write this in L^2 as

$$\langle \xi, f(X)\rangle = \mathrm{E}(\xi\overline{f(X)}) = \mathrm{E}(\overline{f'(X)}) = \langle 1, f'(X)\rangle = \Big\langle\Big(\frac{d}{du}\Big)^* 1, f(X)\Big\rangle,$$

implying that

$$\xi = \Big(\frac{d}{du}\Big)^* 1.$$

The above formulas were for the case $n = 1$ of one variable, but doing the same in the multivariate case is no problem in the classical case.

Exercise 4. Repeat this formal proof in the multivariate case to show that for a random vector $(X_1, \ldots, X_n)$ with density p on $\mathbb{R}^n$ and a function $f : \mathbb{R}^n \to \mathbb{R}$, we have

$$\mathrm{E}\bigg(\Big(\frac{\partial}{\partial u_i} f\Big)(X_1, \ldots, X_n)\bigg) = -\mathrm{E}\Big(\big(\frac{\frac{\partial}{\partial u_i} p}{p}\big)(X_1, \ldots, X_n) \cdot f(X_1, \ldots, X_n)\Big).$$

This can now be made non-commutative by replacing the commutative derivative $\partial/\partial u_i$ by the non-commutative derivative ∂_i. The following definitions are due to Voiculescu [187].

Definition 12. Let (M, τ) be a tracial W^*-probability space and $x_i = x_i^* \in M$ for $i = 1, \ldots, n$.

1) We say $\xi_1, \ldots, \xi_n \in L^2(M)$ *satisfy the conjugate relations* for $x_1, \ldots, x_n$ if we have for all $P \in \mathbb{C}\langle X_1, \ldots, X_n\rangle$

$$\tau(\xi_i P(x_1, \ldots, x_n)) = \tau \otimes \tau\big((\partial_i P)(x_1, \ldots, x_n)\big) \tag{8.11}$$

where for $\eta \in L^2(M)$ we set $\tau(\eta) = \langle \eta, 1\rangle$ or, more explicitly,

$$\tau(\xi_i x_{i(1)} \cdots x_{i(m)}) = \sum_{k=1}^{m} \delta_{i\,i(k)} \tau(x_{i(1)} \cdots x_{i(k-1)}) \tau(x_{i(k+1)} \cdots x_{i(m)}) \tag{8.12}$$

for all $m \geq 0$ and all $1 \leq i, i(1), \ldots, i(m) \leq n$.
($m = 0$ means here of course: $\tau(\xi_i) = 0$.)

2) $\xi_1, \ldots, \xi_n$ is a *conjugate system* for $x_1, \ldots, x_n$, if they satisfy the conjugate relations (8.11) and if in addition $\xi_i \in L^2(x_1, \ldots, x_n)$ for all $i = 1, \ldots, n$.
3) The *free Fisher information* of $x_1, \ldots, x_n$ is defined by

$$\Phi^*(x_1, \ldots, x_n) = \begin{cases} \sum_{i=1}^n \|\xi_i\|_2^2, & \text{if } \xi_1, \ldots, \xi_n \text{ is a conjugate system for } x_1, \ldots, x_n \\ +\infty, & \text{if no conjugate system exists} \end{cases}. \tag{8.13}$$

Note the conjugate relations prescribe the inner products of the ξ_i with a dense subset in $L^2(x_1, \ldots, x_n)$; thus, a conjugate system is unique if it exists.

If there exist $\xi_1, \ldots, \xi_n \in L^2(M)$ which satisfy the conjugate relations, then there exists a conjugate system; this is given by $p\xi_1, \ldots, p\xi_n$ where p is the orthogonal projection from $L^2(M)$ onto $L^2(x_1, \ldots, x_n)$. This holds because the left-hand side of (8.11) is unchanged by replacing ξ_i by $p\xi_i$. Furthermore, we have in such a situation

$$\Phi^*(x_1, \ldots, x_n) = \sum_{i=1}^n \|p\xi_i\|_2^2 \leq \sum_{i=1}^n \|\xi_i\|_2^2,$$

with equality if and only if $\xi_1, \ldots, \xi_n$ is already a conjugate system.

If x and y are free and x has a conjugate variable ξ, then ξ satisfies the conjugate relation (1) in Definition 12 for $x + y$. This means that

$$\tau(\xi(x+y)^n) = \sum_{l=1}^n \tau((x+y)^{l-1})\tau((x+y)^{n-l}).$$

This can be verified from the definition, but there is an easier way to do this using free cumulants. See Exercise 7 following Remark 21 below. By projecting ξ onto $L^2(x+y)$, we get η a conjugate vector whose length has not increased. Thus, when x and y are free, we have $\Phi^*(x+y) \leq \min\{\Phi^*(x), \Phi^*(y)\}$. However, the free Stam inequality (see Theorem 19) is sharper.

Formally, the definition of ξ_i could also be written as $\xi_i = \partial_i^*(1 \otimes 1)$. However, in order that this makes sense, we need ∂_i as an unbounded operator on $L^2(x_1, \ldots, x_n)$, which is the case if and only if $x_1, \ldots, x_n$ are algebraically free. The next proposition by Mai, Speicher, and Weber [121] shows that the existence of a conjugate system excludes algebraic relations between the x_i, and hence the conjugate variables are, if they exist, always of the form $\xi_i = \partial_i^*(1 \otimes 1)$. This implies then also, by Theorem 8, that the ∂_i are closable.

Theorem 13. *Let (M, τ) be a tracial W^*-probability space and $x_i = x_i^* \in M$ for $i = 1, \ldots, n$. Assume that a conjugate system $\xi_1, \ldots, \xi_n$ for $x_1, \ldots, x_n$ exists. Then $x_1, \ldots, x_n$ are algebraically free.*

Proof: Consider $P \in \mathbb{C}\langle X_1, \dots, X_n\rangle$ with $P(x_1, \dots, x_n) = 0$. We claim that then also $q_i := (\partial_i P)(x_1, \dots, x_n) = 0$ for all $i = 1, \dots, n$. In order to see this, let us consider $R_1 P R_2$ for $R_1, R_2 \in \mathbb{C}\langle X_1, \dots, X_n\rangle$. We have $(R_1 P R_2)(x_1, \dots, x_n) = 0$ and, because of

$$\partial_i(R_1 P R_2) = \partial_i R_1 \cdot 1 \otimes P R_2 + R_1 \otimes 1 \cdot \partial_i P \cdot 1 \otimes R_2 + R_1 P \otimes 1 \cdot \partial_i R_2,$$

we get, by putting $r_1 := R_1(x_1, \dots, x_n)$ and $r_2 := R_2(x_1, \dots, x_n)$,

$$\big(\partial_i(R_1 P R_2)\big)(x_1, \dots, x_n) = r_1 \otimes 1 \cdot q_i \cdot 1 \otimes r_2.$$

Thus, we have

$$\begin{aligned} 0 = \tau[\xi_i \cdot (R_1 P R_2)(x_1, \dots, x_n)] &= \tau \otimes \tau\big[\big(\partial_i(R_1 P R_2)\big)(x_1, \dots, x_n)\big] \\ &= \tau \otimes \tau[r_1 \otimes 1 \cdot q_i \cdot 1 \otimes r_2] = \tau \otimes \tau[q_i \cdot r_1 \otimes r_2]. \end{aligned}$$

Hence, $\tau \otimes \tau[q_i \cdot r_1 \otimes r_2] = 0$ for all $r_1, r_2 \in \mathbb{C}\langle x_1, \dots, x_n\rangle$, which implies that $q_i = 0$.

So we can get from a given relation new ones by formal differentiation. We prefer to have relations in $\mathbb{C}\langle x_1, \dots, x_n\rangle$ and not in the tensor product; this can be achieved by applying $id \otimes \tau$ to the q_i. Thus, we have seen that a relation of the form $P(x_1, \dots, x_n) = 0$ implies also the relation $(\partial_i P)(x_1, \dots, x_n) = 0$ and in particular $id \otimes \tau[(\partial_i P)(x_1, \dots, x_n)] = 0$.

Assume now that we have an algebraic relation between the x_i of the form $P(x_1, \dots, x_n) = 0$ for $P \in \mathbb{C}\langle x_1, \dots, x_n\rangle$. Let m be the degree of P. This means that P has a highest order term of the form $\alpha\, X_{i(1)} \cdots X_{i(m)}$ ($\alpha \in \mathbb{C}$); note that there might be other terms of highest order. Denote by D the operator

$$D := (id \otimes \tau) \circ \partial_{i(1)} \circ (id \otimes \tau) \circ \partial_{i(2)} \circ \cdots \circ (id \otimes \tau) \circ \partial_{i(m)}.$$

As an application of $(id \otimes \tau) \circ \partial_i$ reduces the degree of a word $X_{j(1)} \cdots X_{j(k)}$ by at least 1, and exactly 1 only when $j(k) = i$, we have $DX_{i(1)} \cdots X_{i(m)} = 1$, and the application of D on other monomials of length m, as well as on monomials of smaller length, gives 0. This implies that $DP = \alpha$. On the other hand, we know that $DP(x_1, \dots, x_n) = 0$. Hence, we get $\alpha = 0$. By dealing with all highest order terms of P in this fashion, we get in the end that all highest order terms of P are equal to zero; hence, $P = 0$. This means there are no non-trivial algebraic relations for the x_i. □

Let us now look on the free Fisher information Φ^*. As in the case of the free entropy χ, one has again quite explicit formulas in the one-dimensional case, but not in higher dimensions. Before stating the theorem, let us review two basic properties of the Hilbert transform H. Suppose $1 \le p < \infty$ and $f \in L^p(\mathbb{R})$, with respect to Lebesgue measure. For each $\epsilon > 0$, let

$$h_\epsilon(s) = \frac{1}{\pi}\int f(t)\frac{s-t}{(s-t)^2+\epsilon^2}\,dt.$$

Then $h_\epsilon \in L^P(\mathbb{R})$, h_ϵ converges almost everywhere to a function $h \in L^p(\mathbb{R})$, and $\|h_\epsilon - h\|_p \to 0$ as $\epsilon \to 0^+$. We call h the *Hilbert transform* of f and denote it $H(f)$. We can also write $H(f)$ as a Cauchy principal value integral

$$H(f)(s) = \frac{1}{\pi}\int \frac{f(t)}{s-t}\,dt = \frac{1}{2\pi}\int \frac{f(s-t)-f(s+t)}{t}\,dt.$$

When $p=2$, H is an isometry, and for general p, there is a constant C_p such that $\|H(f)\|_p \le C_p\|f\|_p$. See Stein and Weiss [168, Ch. VI, §6, paragraph 6.13].

The Hilbert transform is also related to the Cauchy transform as follows. Recall from Notation 3.4 that the Poisson kernel P and the *conjugate* Poisson kernel Q are given by

$$P_t(s) = \frac{1}{\pi}\frac{t}{s^2+t^2} \qquad \text{and} \qquad Q_t(s) = \frac{1}{\pi}\frac{s}{s^2+t^2}.$$

We have $P_t(s) + iQ_t(s) = i(\pi(s+it))^{-1}$. Let $G(z) = \int f(t)(z-t)^{-1}\,dt$, then

$$h_\epsilon(s) = (Q_\epsilon * f)(s) = \frac{1}{\pi}\mathrm{Re}(G(s+i\epsilon)) \text{ and } (P_\epsilon * f)(s) = \frac{-1}{\pi}\mathrm{Im}(G(s+i\epsilon)). \tag{8.14}$$

The first term converges to $H(f)$ and the second to f as $\epsilon \to 0^+$.

The following result is due to Voiculescu [187].

Theorem 14. *Consider $x = x^* \in M$ and assume that μ_x has a density p which is in $L^3(\mathbb{R})$. Then a conjugate variable exists and is given by*

$$\xi = 2\pi H(p)(x), \qquad \textit{where} \qquad H(p)(v) = \frac{1}{\pi}\int \frac{p(u)}{v-u}\,du$$

is the Hilbert transform. The free Fisher information is then

$$\Phi^*(x) = \frac{4}{3}\pi^2\int p(u)^3\,du. \tag{8.15}$$

Proof: We just give a sketch by providing the formal calculations. If we put $\xi = 2\pi H(p)(x)$, then we have

$$\begin{aligned}\tau(\xi f(x)) &= \tau(2\pi H(p)(x)f(x))\\ &= 2\pi\int H(p)(v)f(v)\,p(v)dv\end{aligned}$$

$$
\begin{aligned}
&= 2\int\int \frac{f(v)}{v-u} p(u)p(v)\,du dv \\
&= \int\int \frac{f(v)}{v-u} p(u)p(v)\,du dv + \int\int \frac{f(u)}{u-v} p(v)p(u)\,dv du \\
&= \int\int \frac{f(u)-f(v)}{u-v} p(u)p(v)\,du dv \\
&= \tau\otimes\tau(\partial f(x)).
\end{aligned}
$$

So we have

$$
\Phi^*(x) = \tau\left((2\pi H(p)(x))^2\right) = 4\pi^2 \int (H(p)(u))^2 p(u)\,du = \frac{4}{3}\pi^2 \int p(u)^3\,du.
$$

The last equality is a general property of the Hilbert transform which follows from Equation (8.14); see Exercise 5. □

Exercise 5.

(i) By replacing $H(p)$ by h_ϵ, make the formal argument rigorous.

(ii) Show, by doing a contour integral, that with $p \in L^3(\mathbb{R})$, we have for the Cauchy transform $G(z) = \int (z-t)^{-1} p(t)\,dt$ that $\int G(t+i\epsilon)^3\,dt = 0$ for all $\epsilon > 0$. Then use Equation (8.14) to prove the last step in the proof of Theorem 14.

After [187] it remained open for a while whether the condition on the density in the last theorem is also necessary. That this is indeed the case is the content of the next proposition, which is an unpublished result of Belinschi and Bercovici. Before we get to this, we need to consider briefly freeness for unbounded operators.

The notion of freeness we have given so far assumes that our random variables have moments of all orders. We now see that the use of conjugate variables requires us to use unbounded operators and these might only have a first and second moment, so our current definition of freeness cannot be applied. For classical independence, there is no need for the random variables to have any moments; the usual definition of independence relies on spectral projections. In the non-commutative picture, we also use spectral projections, except now they may not commute. To describe this we need to review the idea of an operator affiliated to a von Neumann algebra.

Let M be a von Neumann algebra acting on a Hilbert space H, and suppose that t is a closed operator on H. Let $t = u|t|$ be the polar decomposition of t; see, for example, Reed and Simon [150, Ch. VIII]. Now $|t|$ is a closed self-adjoint operator and thus has a spectral resolution $E_{|t|}$. This means that $E_{|t|}$ is a projection-valued measure on $\mathbb{R}$, i.e. we require that for each Borel set $B \subseteq \mathbb{R}$ we have that $E_{|t|}(B)$ is a projection on H and for each pair $\eta_1, \eta_2 \in H$ the measure μ_{η_1,η_2}, defined by $\mu_{\eta_1,\eta_2}(B) = \langle E_{|t|}(B)\eta_1, \eta_2\rangle$, is a complex measure on $\mathbb{R}$. Returning to our t, if both u and $E_{|t|}(B)$ belong to M for every Borel set B, we say that t is *affiliated* with M.

Suppose now that M has a faithful trace τ and $H = L^2(M)$. For t self-adjoint and affiliated with M, we let μ_t, the distribution of t, be given by $\mu_t(B) =$

$\tau(E_t(B))$. If $t \geq 0$ and $\int \lambda \, d\mu_t(\lambda) < \infty$, we say that t is *integrable*. For a general closed operator affiliated with M, we say that t is p-integrable if $|t|^p$ is integrable, i.e. $\int \lambda^p \, d\mu_{|t|}(\lambda) < \infty$. In this picture, $L^2(M)$ is the space of square integrable operators affiliated with M.

Definition 15. Suppose M is a von Neumann algebra with a faithful trace τ and $t_1, \ldots, t_s$ are closed operators affiliated with M. For each i, let A_i be the von Neumann subalgebra of M generated by u_i and the spectral projections $E_{|t_i|}(B)$ where $B \subset \mathbb{R}$ is a Borel set and $t_i = u_i|t_i|$ is the polar decomposition of t_i. If the subalgebras $A_1, \ldots, A_s$ are free with respect to τ, then we say that the operators $t_1, \ldots, t_s$ are *free* with respect to τ.

Remark 16. In [134, Thm. XV] Murray and von Neumann showed that the operators affiliated with M form a $*$-algebra. So if t_1 and t_2 are self-adjoint operators affiliated with M, we can form the spectral measure $\mu_{t_1+t_2}$. When t_1 and t_2 are free, this is the free additive convolution of μ_{t_1} and μ_{t_2}. Indeed this was the definition of $\mu_{t_1} \boxplus \mu_{t_2}$ given by Bercovici and Voiculescu [30]. This shows that by passing to self-adjoint operators affiliated to a von Neumann algebra, one can obtain the free additive convolution of two probability measures on $\mathbb{R}$ from the addition of two free random variables; see Remark 3.48.

Remark 17. If $x = x^* \in M$ and $|z| > \|x\|$, then both

$$\sum_{n \geq 0} z^{-(n+1)} x^n \qquad \text{and} \qquad \sum_{n \geq 1} z^{-(n+1)} \sum_{k=0}^{n-1} x^k \otimes x^{n-k-1}$$

converge in norm to elements of M and $M \otimes M$, respectively. If x has a conjugate variable ξ, then we get by applying the conjugate relation termwise and then summing the equation

$$\tau(\xi(z-x)^{-1}) = \tau \otimes \tau((z-x)^{-1} \otimes (z-x)^{-1}). \tag{8.16}$$

Conversely if $\xi \in L^2(x)$ satisfies this equation for $|z| > \|x\|$, then ξ is the conjugate variable for x. If x is a self-adjoint random variable affiliated with M and $z \in \mathbb{C}^+$, then $(z-x)^{-1} \in M$, and we can ask for a self-adjoint operator $\xi \in L^2(x)$ such that Equation (8.16) holds. If such a ξ exists, we say that ξ is the conjugate variable for x, thus extending the definition to the unbounded case.

The following proposition is an unpublished result by Belinschi and Bercovici.

Proposition 18. *Consider $x = x^* \in M$ and assume that $\Phi^*(x) < \infty$. Then the distribution μ_x is absolutely continuous with respect to Lebesgue measure, and the density p is in $L^3(\mathbb{R})$; moreover, we have*

$$\Phi^*(x) = \frac{4}{3}\pi^2 \int p^3(u) \, du.$$

Proof: Again, we will only provide formal arguments. The main deficiency of the following is that we have to invoke unbounded operators, and the statements we are going to use are only established for bounded operators in our presentation. However, this can be made rigorous by working with operators affiliated with M and by extending the previous theorem to the unbounded setting.

Let t be a Cauchy-distributed random variable which is free from x. (Note that t is an unbounded operator!) Consider for $\varepsilon > 0$ the random variable $x_\varepsilon := x + \varepsilon t$. It can be shown that adding a free variable cannot increase the free Fisher information, since one gets the conjugate variable of x_ε by conditioning the conjugate variable of x onto the L^2-space generated by x_ε. See Exercise 7 below for the argument in the bounded case. For this to make sense in the unbounded case, we use resolvents as above (Remark 17) to say what a conjugate variable is. Hence, $\Phi^*(x_\varepsilon) \leq \Phi^*(x)$ for all $\varepsilon > 0$. But, for any $\varepsilon > 0$, the distribution of x_ε is the free convolution of μ_x with a scaled Cauchy distribution. By Remark 3.34, we have $G_{x_\varepsilon}(z) = G_x(z + i\varepsilon)$, and, hence, by the Stieltjes inversion formula, the distribution of x_ε has a density p_ε which is given by

$$p_\varepsilon(u) = -\frac{1}{\pi}\mathrm{Im}G_x(u + i\varepsilon) = \frac{1}{\pi}\int_{\mathbb{R}} \frac{\varepsilon}{(u - v)^2 + \varepsilon^2} d\mu_x(v).$$

Since this density is always in $L^3(\mathbb{R})$, we know by (the unbounded version of) the previous theorem that

$$\Phi^*(x_\varepsilon) = \int p_\varepsilon(u)^3\, du.$$

So we get

$$\sup_{\varepsilon>0} \frac{1}{\pi^3} \int |\mathrm{Im}G_x(u + i\varepsilon)|^3\, du = \sup_{\varepsilon>0} \Phi^*(x_\varepsilon) \leq \Phi^*(x).$$

This implies (e.g. see [109]) that G_x belongs to the Hardy space $H^3(\mathbb{C}^+)$, and thus μ_x is absolutely continuous and its density is in $L^3(\mathbb{R})$. □

Some important properties of the free Fisher information are collected in the following theorem. For the proof, we refer to Voiculescu's original paper [187].

Theorem 19. *The free Fisher information Φ^* has the following properties (where all appearing variables are self-adjoint and live in a tracial W^*-probability space).*

1) Φ^* is superadditive:

$$\Phi^*(x_1, \ldots, x_n, y_1, \ldots, y_m) \geq \Phi^*(x_1, \ldots, x_n) + \Phi^*(y_1, \ldots, y_m). \tag{8.17}$$

2) *We have the* free Cramér Rao inequality:

$$\Phi^*(x_1,\dots,x_n) \geq \frac{n^2}{\tau(x_1^2)+\cdots+\tau(x_n^2)}. \tag{8.18}$$

3) *We have the* free Stam inequality. *If* $\{x_1,\dots,x_n\}$ *and* $\{y_1,\dots,y_n\}$ *are free, then we have*

$$\frac{1}{\Phi^*(x_1+y_1,\dots,x_n+y_n)} \geq \frac{1}{\Phi^*(x_1,\dots,x_n)} + \frac{1}{\Phi^*(y_1,\dots,y_n)}. \tag{8.19}$$

(This is true even if some of Φ^* *are* $+\infty$*.)*

4) Φ^* *is lower semicontinuous. If, for each* $i = 1,\dots,n$, $x_i^{(k)}$ *converges to* x_i *in the weak operator topology as* $k \to \infty$*, then we have*

$$\liminf_{k\to\infty} \Phi^*(x_1^{(k)},\dots,x_n^{(k)}) \geq \Phi^*(x_1,\dots,x_n). \tag{8.20}$$

Of course, we expect that additivity of the free Fisher information corresponds to the freeness of the variables. We will investigate this more closely in the next section.

8.4 Additivity of Φ^* and freeness

Since cumulants are better suited than moments to deal with freeness, we will first rewrite the conjugate relations into cumulant form.

Theorem 20. *Let* (M,τ) *be a tracial* W^**-probability space and* $x_i = x_i^* \in M$ *for* $i = 1,\dots,n$*. Consider* $\xi_1,\dots,\xi_n \in L^2(M)$*. The following statements are equivalent:*

(i) $\xi_1,\dots,\xi_n$ *satisfy the conjugate relations* (8.12).

(ii) *We have for all* $m \geq 1$ *and* $1 \leq i, i(1),\dots,i(m) \leq n$ *that*

$$\begin{aligned} \kappa_1(\xi_i) &= 0 \\ \kappa_2(\xi_i, x_{i(1)}) &= \delta_{i\,i(1)} \\ \kappa_{m+1}(\xi_i, x_{i(1)},\dots,x_{i(m)}) &= 0 \qquad\qquad (m \geq 2). \end{aligned}$$

Remark 21. Note that up to now we considered only cumulants where all arguments are elements of the algebra M; here, we have the situation where one argument is from L^2 and all the other arguments are from $L^\infty = M$. This is well defined by approximation using the normality of the trace and poses no problems, since multiplying an element from L^2 with an operator from L^∞ gives again an element from L^2; or one can work directly with the inner product on L^2. Cumulants with more than two arguments from L^2 would be problematic. Moreover, one can apply our result, Equation (2.19), when the entries of our cumulant are products, again provided that there are at most two elements from L^2.

Exercise 6. Prove Theorem 20.

Exercise 7. Prove the claim following Theorem 12 that if x_1 and x_2 are free and x_1 has a conjugate variable ξ, then ξ satisfies the conjugate relations for $x_1 + x_2$.

We can now prove the easy direction of the relation between free Fisher information and freeness. This result is due to Voiculescu [187]; our proof using cumulants is from [139].

Theorem 22. *Let (M, τ) be a tracial W^*-probability space, and consider $x_i = x_i^* \in M$ $(i = 1, \ldots, n)$ and $y_j = y_j^* \in M$ $(j = 1, \ldots, m)$. If $\{x_1, \ldots, x_n\}$ and $\{y_1, \ldots, y_m\}$ are free, then we have*

$$\Phi^*(x_1, \ldots, x_n, y_1, \ldots, y_m) = \Phi^*(x_1, \ldots, x_n) + \Phi^*(y_1, \ldots, y_m).$$

(This is true even if some of Φ^* are $+\infty$.)

Proof: If $\Phi^*(x_1, \ldots, x_n) = \infty$ or if $\Phi^*(y_1, \ldots, y_m) = \infty$, then the statement is clear, by the superadditivity of Φ^* from Theorem 19.

So assume $\Phi^*(x_1, \ldots, x_n) < \infty$ and $\Phi^*(y_1, \ldots, y_m) < \infty$. This means that we have a conjugate system $\xi_1, \ldots, \xi_n \in L^2(x_1, \ldots, x_n)$ for $x_1, \ldots, x_n$ and a conjugate system $\eta_1, \ldots, \eta_m \in L^2(y_1, \ldots, y_m)$ for $y_1, \ldots, y_m$. We claim now that $\xi_1, \ldots, \xi_n, \eta_1, \ldots, \eta_m$ is a conjugate system for $x_1, \ldots, x_n, y_1, \ldots, y_m$. It is clear that we have $\xi_1, \ldots, \xi_n, \eta_1, \ldots, \eta_m \in L^2(x_1, \ldots, x_n, y_1, \ldots, y_m)$; it only remains to check the conjugate relations. We do this in terms of cumulants, verifying the relations (*ii*) using Theorem 20. The relations involving only x's and ξ's or only y's and η's are satisfied because of the conjugate relations for either x/ξ or y/η. Because of $\xi_i \in L^2(x_1, \ldots, x_n)$ and $\eta_j \in L^2(y_1, \ldots, y_m)$ and the fact that $\{x_1, \ldots, x_n\}$ and $\{y_1, \ldots, y_m\}$ are free, we have furthermore the vanishing (see Remark 21) of all cumulants with mixed arguments from $\{x_1, \ldots, x_n, \xi_1, \ldots, \xi_n\}$ and $\{y_1, \ldots, y_m, \eta_1, \ldots, \eta_m\}$. But this gives then all the conjugate relations. □

The less straightforward implication, namely, that additivity of the free Fisher information implies freeness, relies on the following relation for commutators between variables and their conjugate variables. This, as well as the consequence for free Fisher information, was proved by Voiculescu in [189], whereas our proofs use again adaptations of ideas from [139].

Theorem 23. *Let (M, τ) be a tracial W^*-probability space and $x_i = x_i^* \in M$ for $i = 1, \ldots, n$. Let $\xi_1, \ldots, \xi_n \in L^2(x_1, \ldots, x_n)$ be a conjugate system for $x_1, \ldots, x_n$. Then we have*

$$\sum_{i=1}^{n}[x_i, \xi_i] = 0$$

(where $[a, b] = ab - ba$ denotes the commutator of a and b).

Proof: Let us put

$$c := \sum_{i=1}^{n}[x_i, \xi_i] \in L^2(x_1, \ldots, x_n).$$

Then it suffices to show

$$\tau(c x_{i(1)} \cdots x_{i(m)}) = 0 \qquad \text{for all } m \geq 0 \text{ and all } 1 \leq i(1), \ldots, i(m) \leq n.$$

In terms of cumulants, this is equivalent to

$$\kappa_{m+1}(c, x_{i(1)}, \ldots, x_{i(m)}) = 0 \quad \text{for all } m \geq 0 \text{ and all } 1 \leq i(1), \ldots, i(m) \leq n.$$

By using the formula for cumulants with products as entries, Theorem 2.13, we get

$$\begin{aligned}
&\kappa_{m+1}(c, x_{i(1)}, \ldots, x_{i(m)}) \\
&= \sum_{i=1}^{m} \big(\kappa_{m+1}(x_i \xi_i, x_{i(1)}, \ldots, x_{i(m)}) - \kappa_{m+1}(\xi_i x_i, x_{i(1)}, \ldots, x_{i(m)})\big) \\
&= \sum_{i=1}^{m} \big(\kappa_2(\xi_i, x_{i(1)})\kappa_m(x_i, x_{i(2)}, \ldots, x_{i(m)}) - \kappa_2(\xi_i, x_{i(m)})\kappa_m(x_i, x_{i(1)}, \ldots, x_{i(m-1)})\big) \\
&= \kappa_m(x_{i(1)}, x_{i(2)}, \ldots, x_{i(m)}) - \kappa_m(x_{i(m)}, x_{i(1)}, \ldots, x_{i(m-1)}) \\
&= 0,
\end{aligned}$$

because, in the case of the first sum, the only partition, π, that satisfies the two conditions that ξ_i is in a block of size two and $\pi \vee \{(1, 2), (3), \cdots, (m+2)\} = 1_{m+2}$ is $\pi = \{(1, 4, 5, \ldots, m+2), (2, 3)\}$ and, in the case of the second sum, the only partition, σ, that satisfies the two conditions that ξ_i is in a block of size two and $\sigma \vee \{(1, 2), (3), \cdots, (m+2)\} = 1_{m+2}$ is $\sigma = \{(1, m+2), (2, 3, 4, \ldots, m+1)\}$. The last equality follows from the fact that τ is a trace; see Exercise 2.8. □

Theorem 24. *Let (M, τ) be a tracial W^*-probability space and $x_i = x_i^* \in M$ for $i = 1, \ldots, n$ and $y_j = y_j^* \in M$ for $j = 1, \ldots, m$. Assume that*

$$\Phi^*(x_1, \ldots, x_n, y_1, \ldots, y_m) = \Phi^*(x_1, \ldots, x_n) + \Phi^*(y_1, \ldots, y_m) < \infty.$$

Then, $\{x_1, \ldots, x_n\}$ and $\{y_1, \ldots, y_m\}$ are free.

Proof: Let $\xi_1, \ldots, \xi_n, \eta_1, \ldots, \eta_m \in L^2(x_1, \ldots, x_n, y_1, \ldots, y_m)$ be the conjugate system for $x_1, \ldots, x_n, y_1, \ldots, y_m$. Since this means in particular that $\xi_1, \ldots, \xi_n$ satisfy the conjugate relations for $x_1, \ldots, x_n$, we know that $P\xi_1, \ldots, P\xi_n$ is the conjugate system for $x_1, \ldots, x_n$, where P is the orthogonal projection onto $L^2(x_1, \ldots, x_n)$. In the same way, $Q\eta_1, \ldots, Q\eta_m$ is the conjugate system for $y_1, \ldots, y_m$, where Q is the orthogonal projection onto $L^2(y_1, \ldots, y_m)$. But then we have

$$\begin{aligned}\sum_{i=1}^{n} \|\xi_i\|_2^2 + \sum_{j=1}^{m} \|\eta_j\|_2^2 &= \Phi^*(x_1, \ldots, x_n, y_1, \ldots, y_m)\\ &= \Phi^*(x_1, \ldots, x_n) + \Phi^*(y_1, \ldots, y_m)\\ &= \sum_{i=1}^{n} \|P\xi_i\|_2^2 + \sum_{j=1}^{m} \|Q\eta_j\|_2^2.\end{aligned}$$

However, this means that the projection P has no effect on the ξ_i and the projection Q has no effect on η_j; hence, the additivity of the Fisher information is saying that $\xi_1, \ldots, \xi_n$ is already the conjugate system for $x_1, \ldots, x_n$ and $\eta_1, \ldots, \eta_m$ is already the conjugate system for $y_1, \ldots, y_m$. By Theorem 23, this implies that

$$\sum_{i=1}^{n} [x_i, \xi_i] = 0 \qquad \text{and} \qquad \sum_{j=1}^{m} [y_j, \eta_j] = 0.$$

In order to prove the asserted freeness, we have to check that all mixed cumulants in $\{x_1, \ldots, x_n\}$ and $\{y_1, \ldots, y_m\}$ vanish. In this situation, a mixed cumulant means there is at least one x_i and at least one y_j. Moreover, because we are working with a tracial state, it suffices to show $\kappa_{r+2}(x_i, z_1, \ldots, z_r, y_j) = 0$ for all $r \geq 0$; $i = 1, \ldots, n$; $j = 1, \ldots, m$; and $z_1, \ldots, z_r \in \{x_1, \ldots, x_n, y_1, \ldots, y_m\}$. Consider such a situation. Then we have

$$\begin{aligned}0 &= \kappa_{r+3}\Big(\sum_{k=1}^{n} [x_k, \xi_k], x_i, z_1, \ldots, z_r, y_j\Big)\\ &= \sum_{k=1}^{n} \underbrace{\kappa_{r+3}(x_k\xi_k, x_i, z_1, \ldots, z_r, y_j)}_{\kappa_2(\xi_k, x_i)\cdot\kappa_{r+2}(x_k, z_1, \ldots, z_r, y_j)} - \sum_{k=1}^{n} \underbrace{\kappa_{r+3}(\xi_k x_k, x_i, z_1, \ldots, z_r, y_j)}_{\kappa_2(\xi_k, y_j)\cdot\kappa_{r+2}(x_k, x_i, z_1, \ldots, z_r)}\\ &= \kappa_{r+2}(x_i, z_1, \ldots, z_r, y_j),\end{aligned}$$

because, by the conjugate relations, $\kappa_2(\xi_k, x_i) = \delta_{ki}$ and $\kappa_2(\xi_k, y_j) = 0$ for all $k = 1, \ldots, n$ and all $j = 1, \ldots, m$. $\square$

8.5 The non-microstates free entropy χ^*

By analogy with the classical situation, we would expect that the free Fisher information of $x_1, \dots, x_n$ is the derivative of the free entropy for a Brownian motion starting in $x_1, \dots, x_n$. Reversing this, the free entropy should be the integral over free Fisher information along Brownian motions. Since we cannot prove this at the moment for the microstates free entropy χ (which we defined in the last chapter), we use this idea to define another version of free entropy, which we denote by χ^*. Of course, we hope that at some point in the not-too-distant future, we will be able to show that $\chi = \chi^*$.

Definition 25. Let (M, τ) be a tracial W^*-probability space. For random variables $x_i = x_i^* \in M$ $(i = 1, \dots, n)$, the *non-microstates free entropy* is defined by

$$\chi^*(x_1, \dots, x_n) := \frac{1}{2} \int_0^\infty \left(\frac{n}{1+t} - \Phi^*(x_1 + \sqrt{t} s_1, \dots, x_n + \sqrt{t} s_n) \right) dt + \frac{n}{2} \log(2\pi e), \tag{8.21}$$

where $s_1, \dots, s_n$ are free semi-circular random variables which are free from $\{x_1, \dots, x_n\}$.

One can now rewrite the properties of Φ^* into properties of χ^*. In the next theorem, we collect the most important ones. The proofs are mostly straightforward (given the properties of Φ^*), and we refer again to Voiculescu's original papers [187, 189].

Theorem 26. *The non-microstates free entropy has the following properties (where all variables which appear are self-adjoint and are in a tracial W^*-probability space).*

1) *For $n = 1$, we have $\chi^*(x) = \chi(x)$.*
2) *We have the upper bound*

$$\chi^*(x_1, \dots, x_n) \le \frac{n}{2} \log(2\pi n^{-1} C^2), \tag{8.22}$$

where $C^2 = \tau(x_1^2 + \cdots + x_n^2)$.

3) *χ^* is subadditive:*

$$\chi^*(x_1, \dots, x_n, y_1, \dots, y_m) \le \chi^*(x_1, \dots, x_n) + \chi^*(y_1, \dots, y_m). \tag{8.23}$$

4) *If $\{x_1, \dots, x_n\}$ and $\{y_1, \dots, y_m\}$ are free, then*

$$\chi^*(x_1, \dots, x_n, y_1, \dots, y_m) = \chi^*(x_1, \dots, x_n) + \chi^*(y_1, \dots, y_m). \tag{8.24}$$

5) *On the other hand, if*

$$\chi^*(x_1, \dots, x_n, y_1, \dots, y_m) = \chi^*(x_1, \dots, x_n) + \chi^*(y_1, \dots, y_m) > -\infty$$

then $\{x_1, \dots, x_n\}$ and $\{y_1, \dots, y_m\}$ are free.

6) χ^* is upper semicontinuous. If, for each $i = 1, \ldots, n$, $x_i^{(k)}$ converges for $k \to \infty$ in the weak operator topology to x_i, then we have

$$\limsup_{k\to\infty} \chi^*(x_1^{(k)}, \ldots, x_n^{(k)}) \leq \chi^*(x_1, \ldots, x_n). \tag{8.25}$$

7) We have the following log-*Sobolev inequality. If $\Phi^*(x_1, \ldots, x_n) < \infty$, then*

$$\chi^*(x_1, \ldots, x_n) \geq \frac{n}{2} \log\left(\frac{2\pi n e}{\Phi^*(x_1, \ldots, x_n)}\right). \tag{8.26}$$

In particular:

$$\Phi^*(x_1, \ldots, x_n) < \infty \implies \chi^*(x_1, \ldots, x_n) > -\infty. \tag{8.27}$$

Though we do not know at the moment whether $\chi = \chi^*$ in general, we have at least one half of this by the following deep result of Biane, Capitaine, and Guionnet [40].

Theorem 27. *Let (M, τ) be a tracial W^*-probability space and $x_i = x_i^* \in M$ for $i = 1, \ldots, n$. Then we have*

$$\chi(x_1, \ldots, x_n) \leq \chi^*(x_1, \ldots, x_n). \tag{8.28}$$

8.6 Operator algebraic applications of free Fisher information

Assume that $\Phi^*(x_1, \ldots, x_n) < \infty$. Then, by (8.27), we have that $\chi^*(x_1, \ldots, x_n) > -\infty$. If we believe that $\chi^* = \chi$, then by our results from the last chapter, this would imply certain properties of the von Neumann algebra generated by $x_1, \ldots, x_n$. In particular, $\mathrm{vN}(x_1, \ldots, x_n)$ would not have property Γ. (Note that the inequality $\chi \leq \chi^*$ from Theorem 27 goes in the wrong direction to obtain this conclusion.)

We will now show directly the absence of property Γ from the assumption $\Phi^*(x_1, \ldots, x_n) < \infty$. This result is due to Dabrowski, and we will follow quite closely his arguments from [63].

In the following, we will always work in a tracial W^*-probability space (M, τ) and consider $x_i = x_i^* \in M$ for $i = 1, \ldots, n$. We assume that $\Phi^*(x_1, \ldots, x_n) < \infty$ and denote by $\xi_1, \ldots, \xi_n$ the conjugate system for $x_1, \ldots, x_n$. Recall also from Theorem 13 that finite Fisher information excludes algebraic relations among $x_1, \ldots, x_n$; hence, ∂_i is defined as an unbounded operator on $\mathbb{C}\langle x_1, \ldots, x_n\rangle$. In particular, if $P \in \mathbb{C}\langle X_1, \ldots, X_n\rangle$ and $p = P(x_1, \ldots, x_n)$, then $\partial_i p$ is the same as $(\partial_i P)(x_1, \ldots, x_n)$.

The crucial technical calculations are contained in the following lemma.

Lemma 28. *Assume that $\Phi^*(x_1,\ldots,x_n) < \infty$. Then we have for all $p \in \mathbb{C}\langle x_1,\ldots,x_n\rangle$*

$$(n-1)\|[p,1\otimes 1]\|_2^2 = \sum_{i=1}^n \langle [p,x_i],[p,\xi_i]\rangle + 2\mathrm{Re}\Big(\sum_{i=1}^n \langle \partial_i p,[1\otimes 1,[p,x_i]]\rangle\Big) \tag{8.29}$$

(note that $[p,1\otimes 1]$ should here be understood as module operations, i.e. we have $[p,1\otimes 1] = p\otimes 1 - 1\otimes p$).

Proof: We write, for arbitrary $j \in \{1,\ldots,n\}$,

$$\begin{aligned}\|[p,1\otimes 1]\|_2^2 &= \langle [p,1\otimes 1],[p,1\otimes 1]\rangle\\ &= \langle \partial_j[p,x_j],[p,1\otimes 1]\rangle - \langle [\partial_j p,x_j],[p,1\otimes 1]\rangle.\end{aligned}$$

We rewrite the first term, by using (8.6), as

$$\begin{aligned}\langle \partial_j[p,x_j],[p,1\otimes 1]\rangle &= \langle [p,x_j],\partial_j^*[p,1\otimes 1]\rangle\\ &= \langle [p,x_j],\partial_j^*(p\otimes 1 - 1\otimes p)\rangle\\ &= \langle [p,x_j], p\xi_j - id\otimes\tau(\partial_j p) - \xi_j p + \tau\otimes id(\partial_j p)\rangle\\ &= \big\langle [p,x_j],[p,\xi_j]\big\rangle + \big\langle \big[1\otimes 1,[p,x_j]\big],\partial_j p\big\rangle,\end{aligned}$$

and the second term as

$$\langle [\partial_j p,x_j],[p,1\otimes 1]\rangle = \big\langle \partial_j p,\big[p,[1\otimes 1,x_j]\big]\big\rangle - \big\langle \partial_j p,\big[1\otimes 1,[p,x_j]\big]\big\rangle.$$

The first term of the latter is

$$\begin{aligned}\big\langle \partial_j p,\big[p,[1\otimes 1,x_j]\big]\big\rangle &= \big\langle \partial_j p,(1\otimes x_j)[p,1\otimes 1] - [p,1\otimes 1](x_j\otimes 1)\big\rangle\\ &= \langle 1\otimes x_j\cdot\partial_j p - \partial_j p\cdot x_j\otimes 1,[p,1\otimes 1]\rangle.\end{aligned}$$

Note that summing the last expression over j yields, by Lemma 4,

$$\begin{aligned}\sum_{j=1}^n \langle 1\otimes x_j\cdot\partial_j p - \partial_j p\cdot x_j\otimes 1,[p,1\otimes 1]\rangle &= \langle -(p\otimes 1 - 1\otimes p),[p,1\otimes 1]\rangle\\ &= -\langle [p,1\otimes 1],[p,1\otimes 1]\rangle.\end{aligned}$$

Summing all our equations over j gives Equation (8.29). □

Corollary 29. *Assume that* $\Phi^*(x_1,\dots,x_n) < \infty$. *Then we have for all* $t \in \mathrm{vN}(x_1,\dots,x_n)$

$$(n-1)\|t-\tau(t)\|_2^2 \le \frac{1}{2}\sum_{i=1}^n \Big\{\langle [t,x_i],[t,\xi_i]\rangle + 4\|[t,x_i]\|_2\cdot\|\xi_i\|_2\cdot\|t\|\Big\}.$$

Proof: It suffices to prove the statement for $t = p \in \mathbb{C}\langle x_1,\dots,x_n\rangle$. First note that

$$\|[p,1\otimes 1]\|_2^2 = \langle p\otimes 1 - 1\otimes p, p\otimes 1 - 1\otimes p\rangle = 2(\tau(p^*p) - |\tau(p)|^2) = 2\|p-\tau(p)\|_2^2.$$

Thus, (8.29) gives

$$(n-1)\|p-\tau(p)\|_2^2 = \frac{1}{2}\sum_{i=1}^n \langle [p,x_i],[p,\xi_i]\rangle + \mathrm{Re}\Big(\sum_{i=1}^n \langle \partial_i p, [1\otimes 1,[p,x_i]]\rangle\Big).$$

We write the second summand as

$$\begin{aligned}\langle \partial_i p,[1\otimes 1,[p,x_i]]\rangle &= \langle \partial_i p, [p,x_i]\otimes 1 - 1\otimes [p,x_i]\rangle\\ &= \langle id\otimes\tau(\partial_i p),[p,x_i]\rangle - \langle \tau\otimes id(\partial_i p),[p,x_i]\rangle;\end{aligned}$$

hence, we can estimate its real part by

$$\mathrm{Re}\langle \partial_i p,[1\otimes 1,[p,x_i]]\rangle \le 2\|(id\otimes\tau)\partial_i p\|_2\cdot\|[p,x_i]\|_2 + 2\|(\tau\otimes id)\partial_i p\|_2\cdot\|[p,x_i]\|_2$$

which, by Equation (8.10), gives the assertion. □

Recall from Definition 7.5 that a von Neumann algebra has property Γ if it has a non-trivial central sequence.

Theorem 30. *Let* $n \ge 2$ *and* $\Phi^*(x_1,\dots,x_n) < \infty$. *Then* $\mathrm{vN}(x_1,\dots,x_n)$ *does not have property* Γ *(and hence is a factor).*

Proof: Let $(t_k)_{k\in\mathbb{N}}$ be a central sequence in $\mathrm{vN}(x_1,\dots,x_n)$. (Recall that central sequences are, by definition, bounded in operator norm.) This means in particular that $[t_k,x_i]$ converges, for $k\to\infty$, in $L^2(M)$ to 0, for all $i=1,\dots,n$. But then, by Corollary 29, we also have $\|t_k-\tau(t_k)\|_2 \to 0$, which means that our central sequence is trivial. Thus, there exists no non-trivial central sequence. □

8.7 Absence of atoms for self-adjoint polynomials

In Theorem 13 we have seen that finite Fisher information (i.e. the existence of a conjugate system) implies that the variables are algebraically free. This means that for non-trivial $P \in \mathbb{C}\langle X_1,\dots,X_n\rangle$, the operator $p := P(x_1,\dots,x_n)$ cannot be zero. The ideas from the proof of this statement can actually be refined in order to prove a much deeper statement, namely, the absence of atoms for the distribution μ_p for any such self-adjoint polynomial. Note that atoms at position t in the distribution of μ_p correspond to the existence of a non-trivial eigenspace of p for the eigenvalue t. By replacing our polynomial by $p - t1$, we shift the atom to 0, and thus asking the question whether non-trivial polynomials can have non-trivial kernels. This can be rephrased in a more algebraic language in the form $pw = 0$ where w is the orthogonal projection onto this kernel. Whereas p is a polynomial, the projection w will in general just be an element in the von Neumann algebra. Hence, the question of atoms is, at least for self-adjoint polynomials, the same as the question of zero divisors in the following sense.

Definition 31. A *zero divisor* w for $0 \neq p \in \mathbb{C}\langle x_1, \ldots, x_n \rangle$ is a non-trivial element $0 \neq w \in \mathrm{vN}(x_1, \ldots, x_n)$ such that $pw = 0$.

Theorem 32. *Let (M, τ) be a tracial W^*-probability space and $x_i = x_i^* \in M$ for $i = 1, \ldots, n$. Assume that $\Phi^*(x_1, \ldots, x_n) < \infty$. Then for any non-trivial $p \in \mathbb{C}\langle x_1, \ldots, x_n \rangle$, there exists no zero divisor.*

Proof: The rough idea of the proof follows the same line as the proof of Theorem 13; namely, assume that we have a zero divisor for some polynomial, and then one shows that by differentiating this statement, one also has a zero divisor for a polynomial of lesser degree. Thus, one can reduce the general case to the (non-trivial) degree 0 case, where obviously no zero divisors exist.

More precisely, assume that we have $pw = 0$ for non-trivial $p \in \mathbb{C}\langle x_1, \ldots, x_n \rangle$ and $w \in \mathrm{vN}(x_1, \ldots, x_n)$. Furthermore, we can assume that both p and w are self-adjoint (otherwise, consider $p^* p w w^* = 0$). Then $pw = 0$ implies also $wp = 0$. We will now consider the equation $wpw = 0$ and take the derivative ∂_i of this. Of course, we have now the problem that w is not necessarily in the domain $D(\partial_i)$ of our derivative. However, by approximating w by polynomials and controlling norms via Dabrowski's inequality from Theorem 10, one can show that the following formal arguments can be justified rigorously.

From $wpw = 0$, we get

$$0 = \partial_i(wpw) = \partial_i w \cdot 1 \otimes pw + w \otimes 1 \cdot \partial_i p \cdot 1 \otimes w + wp \otimes 1 \cdot \partial_i w.$$

Because of $pw = 0$ and $wp = 0$, the first and the third term vanish, and we are left with $w \otimes 1 \cdot \partial_i p \cdot 1 \otimes w = 0$. Again we apply $\tau \otimes id$ to this, in order to get an equation in the algebra instead of the tensor product; we get

$$\underbrace{[(\tau \otimes id)(w \otimes 1 \cdot \partial_i p)]}_{=:q} w = 0.$$

Hence, we have $qw = 0$ and q is a polynomial of smaller degree. However, this q is in general not self-adjoint, and thus the other equation $wq = 0$ is now not a consequence. But since we are in a tracial setting, basic theory of equivalence of projections for von Neumann algebras shows that we have a non-trivial $v \in \mathrm{vN}(x_1, \ldots, x_n)$ such that $vq = 0$. Indeed, the projections onto $\ker(q)$ and $\ker(q^*)$ are equivalent. Since $qw = 0$, we have $\ker(q) \neq \{0\}$ and thus $\ker(q^*) \neq \{0\}$. This means that $\operatorname{ran}(q)$ is not dense and hence there is $v \neq 0$ with $vq = 0$. Then we can continue with $vqw = 0$ in the same way as above and get a further reduction of our polynomial. Of course, we have to avoid that taking the derivative gives a trivial polynomial, but since the above works for all ∂_i with $i = 1, \dots, n$, we have enough flexibility to avoid this.

For the details of the proof, we refer to the original work [121]. □

The condition $\Phi^*(x_1, \dots, x_n) < \infty$ is not the weakest possible; in [51] it was shown that the conclusion of Theorem 32 still holds under the assumption of maximal free entropy dimension.

8.8 Additional exercises

Exercise 8.

(*i*) Let $s_1, \dots, s_n$ be n free semi-circular elements and $\partial_1, \dots, \partial_n$ the corresponding non-commutative derivatives. Show that one has

$$\partial_i^*(1 \otimes 1) = s_i \qquad \text{for all } i = 1, \dots, n.$$

(*ii*) Show that the conclusion from (*i*) actually characterizes a family of n free semi-circulars. Equivalently, let $\xi_1, \dots, \xi_n$ be the conjugate system for self-adjoint variables $x_1, \dots, x_n$ in some tracial W^*-probability space. Assume that $\xi_i = x_i$ for all $i = 1, \dots, n$. Show that $x_1, \dots, x_n$ are n free semi-circular variables.

Exercise 9. Let $s_1, \dots, s_n$ be n free semi-circular elements. Fix a natural number m, and let $f : \{1, \dots, n\}^m \to \mathbb{C}$ be any function that "vanishes on the diagonals", i.e. $f(i_1, \dots, i_m) = 0$ whenever there are $k \neq l$ such that $i_k = i_l$. Put

$$p := \sum_{i_1, \dots, i_m = 1}^{n} f(i_1, \dots, i_m) s_{i_1} \cdots s_{i_m} \in \mathbb{C}\langle s_1, \dots, s_n \rangle.$$

Calculate $\sum_{i=1}^n \partial_i^* \partial_i p$.

Notation 33. *In the following $(C_n)_{n \in \mathbb{N}_0}$ and $(U_n)_{n \in \mathbb{N}_0}$ will be the* Chebyshev polynomials of the first and second kind, respectively *(rescaled to the interval $[-2, 2]$), i.e. the sequence of polynomials $C_n, U_n \in \mathbb{C}\langle X \rangle$ which are defined recursively by*

$$C_0(X) = 2, \qquad C_1(X) = X, \qquad C_{n+1}(X) = XC_n(X) - C_{n-1}(X) \quad (n \geq 1)$$

and

$$U_0(X) = 1, \qquad U_1(X) = X, \qquad U_{n+1}(X) = XU_n(X) - U_{n-1}(X) \quad (n \geq 1).$$

These polynomials already appeared in Chapter 5. See, in particular, Exercise 5.12.

Exercise 10. Let $\partial : \mathbb{C}\langle X\rangle \to \mathbb{C}\langle X\rangle \otimes \mathbb{C}\langle X\rangle$ be the non-commutative derivative with respect to X. Show that

$$\partial U_n(X) = \sum_{k=1}^{n} U_{k-1}(X) \otimes U_{n-k}(X) \qquad \text{for all } n \in \mathbb{N}.$$

Exercise 11. Let s be a semi-circular variable of variance 1. Let ∂ be the non-commutative derivative with respect to s, considered as an unbounded operator on L^2.

(i) Show that the $(U_n)_{n\in\mathbb{N}_0}$ are the orthogonal polynomials for the semi-circle distribution, i.e. that

$$\tau(U_m(s)U_n(s)) = \delta_{m,n}.$$

(ii) Show that

$$\partial^*(U_n(s) \otimes U_m(s)) = U_{n+m+1}(s).$$

(iii) Show that for any $p \in \mathbb{C}\langle s\rangle$, we have

$$\|\partial^*(p \otimes 1)\|_2 = \|p\|_2 \qquad \text{and} \qquad \|(id \otimes \tau)\partial p\|_2 \leq \|p\|_2.$$

(Note that the latter is in this case a stronger version of Theorem 10.)

(iv) The statement in (iii) shows that $(id \otimes \tau) \circ \partial$ is a bounded operator with respect to $\|\cdot\|_2$. Show that this is not true for ∂, by proving that $\|U_n(s)\|_2 = 1$ and $\|\partial U_n(s)\|_2 = \sqrt{n}$.

Exercise 12.

(i) Show that we have for all $n, m \geq 0$

$$C_n U_m = \begin{cases} U_{n+m} + U_{m-n}, & n \leq m \\ U_{n+m}, & n = m+1 \\ U_{n+m} - U_{n-m-2}, & n \geq m+2 \end{cases}.$$

(ii) Let (M, τ) be a tracial W^*-probability space and $x = x^* \in M$. Put $\alpha_n := \tau(U_{n-1}(x))$. Assume that

$$\xi := \sum_{n=1}^{\infty} \alpha_n C_n(x) \in L^2(M, \tau).$$

Show that ξ is the conjugate variable for x.

Exercise 13. For $P = (P_1, \dots, P_n)$ with $P_1, \dots, P_n \in \mathbb{C}\langle X_1, \dots, X_n\rangle$, we define the non-commutative Jacobian

$$\mathcal{J}P = (\partial_j P_i)_{i,j=1}^n \in M_n(\mathbb{C}\langle X_1, \dots, X_n\rangle^{\otimes 2}).$$

If $Q = (Q_1, \dots, Q_n)$ with $Q_1, \dots, Q_n \in \mathbb{C}\langle X_1, \dots, X_n\rangle$, then we define

$$P \circ Q = (P_1 \circ Q, \dots, P_n \circ Q)$$

and $P_i \circ Q \in \mathbb{C}\langle X_1, \dots, X_n\rangle$ by

$$P_i \circ Q(X_1, \dots, X_n) := P_i(Q_1(X_1, \dots, X_n), \dots, Q_n(X_1, \dots, X_n)).$$

Express $\mathcal{J}(P \circ Q)$ in terms of $\mathcal{J}P$ and $\mathcal{J}Q$.

Exercise 14. Let (M, τ) be a tracial W^*-probability space and $x_i = x_i^* \in M$ for $i = 1, \dots, n$. Assume $\Phi^*(x_1, \dots, x_n) < \infty$.

(i) Show that we have for $\lambda > 0$

$$\Phi^*(\lambda x_1, \dots, \lambda x_n) = \frac{1}{\lambda^2} \Phi^*(x_1, \dots, x_n).$$

(ii) Let now $A = (a_{ij})_{i,j=1}^n \in M_n(\mathbb{R})$ be a real invertible $n \times n$ matrix, and put

$$y_i := \sum_{j=1}^n a_{ij} x_j .$$

Determine the relation between a conjugate system for $x_1, \dots, x_n$ and a conjugate system for $y_1, \dots, y_n$. Conclude from this the following.

- If A is orthogonal, then we have

$$\Phi^*(x_1, \dots, x_n) = \Phi^*(y_1, \dots, y_n).$$

- For general A, we have

$$\frac{1}{\|A\|^2} \Phi^*(y_1, \dots, y_n) \le \Phi^*(x_1, \dots, x_n) \le \|A\|^2 \Phi^*(y_1, \dots, y_n).$$

Chapter 9
Operator-Valued Free Probability Theory and Block Random Matrices

Gaussian random matrices fit quite well into the framework of free probability theory, asymptotically they are semi-circular elements, and they have also nice freeness properties with other (e.g. non-random) matrices. Gaussian random matrices are used as input in many basic models in many different mathematical, physical, or engineering areas. Free probability theory provides then useful tools for the calculation of the asymptotic eigenvalue distribution for such models. However, in many situations, Gaussian random matrices are only the first approximation to the considered phenomena, and one would also like to consider more general kinds of such random matrices. Such generalizations often do not fit into the framework of our usual free probability theory. However, there exists an extension, operator-valued free probability theory, which still shares the basic properties of free probability but is much more powerful because of its wider domain of applicability. In this chapter, we will first motivate the operator-valued version of a semi-circular element and then present the general operator-valued theory. Here we will mainly work on a formal level; the analytic description of the theory, as well as its powerful consequences, will be dealt with in the following chapter.

9.1 Gaussian block random matrices

Consider $A_N = (a_{ij})_{i,j=1}^N$. Our usual assumptions for a Gaussian random matrix are that the entries a_{ij} are, apart from the symmetry condition $a_{ij} = a_{ji}^*$, independent and identically distributed with a centred normal distribution. There are many ways to relax these conditions, for example, one might consider noncentred normal distributions, relax the identical distribution by allowing a dependency of the variance on the entry, or even give up the independence by allowing correlations between the entries. One possibility for such correlations would be *block matrices*, where our random matrix is build up as a $d \times d$ matrix out of blocks, where each block is an ordinary Gaussian random matrix, but we allow that the blocks might repeat. For example, for $d = 3$, we might consider a block matrix

J.A. Mingo, R. Speicher, *Free Probability and Random Matrices*,
Fields Institute Monographs 35, DOI 10.1007/978-1-4939-6942-5_9

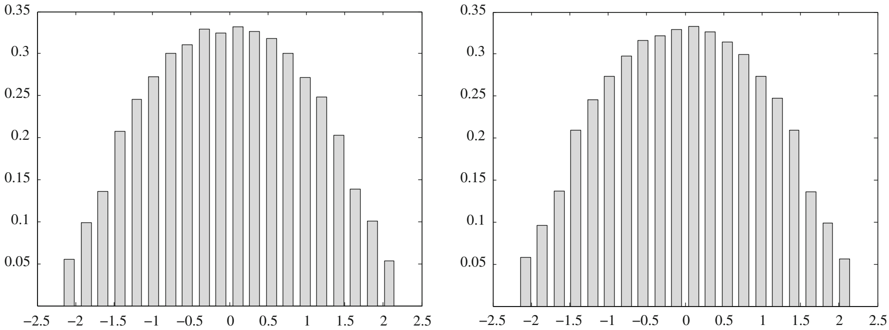


Fig. 9.1 Histogram of the dN eigenvalues of a random matrix X_N, for $N = 1000$, for two different realizations

$$X_N = \frac{1}{\sqrt{3}} \begin{pmatrix} A_N & B_N & C_N \\ B_N & A_N & B_N \\ C_N & B_N & A_N \end{pmatrix}, \tag{9.1}$$

where A_N, B_N, C_N are independent self-adjoint Gaussian $N \times N$-random matrices. As usual we are interested in the asymptotic eigenvalue distribution of X_N as $N \to \infty$.

As in Chapter 5 we can look at numerical simulations for the eigenvalue distribution of such matrices. In Fig. 9.1 there are two realizations of the random matrix above for $N = 1000$. This suggests that again we have almost sure convergence to a deterministic limit distribution. One sees, however, that this limiting distribution is not a semi-circle.

In this example, we have of course the following description of the limiting distribution. Because the joint distribution of $\{A_N, B_N, C_N\}$ converges to that of $\{s_1, s_2, s_3\}$, where $\{s_1, s_2, s_3\}$ are free standard semi-circular elements, the limit eigenvalue distribution we seek is the same as the distribution μ_X of

$$X = \frac{1}{\sqrt{3}} \begin{pmatrix} s_1 & s_2 & s_3 \\ s_2 & s_1 & s_2 \\ s_3 & s_2 & s_1 \end{pmatrix} \tag{9.2}$$

with respect to $\mathrm{tr}_3 \otimes \varphi$ (where φ is the state acting on s_1, s_2, s_3). Actually, because we have the almost sure convergence of A_N, B_N, C_N (with respect to tr_N) to s_1, s_2, s_3, this implies that the empirical eigenvalue distribution of X_N converges almost surely to μ_X. Thus, free probability yields directly the almost sure existence of a limiting eigenvalue distribution of X_N. However, the main problem, namely, the concrete determination of this limit μ_X, cannot be achieved within usual free probability theory. Matrices of semi-circular elements do in general not behave nicely with respect to $\mathrm{tr_d} \otimes \varphi$. However, there exists a generalization, *operator-valued free probability theory*, which is tailor-made to deal with such matrices.

In order to see what goes wrong on the usual level and what can be saved on an "operator-valued" level, we will now try to calculate the moments of X in our usual combinatorial way. To construct our first example, we shall need the idea of a circular family of operators, generalizing the idea of a semi-circular family given in Definition 2.6

Definition 1. Let $\{c_1, \dots, c_n\}$ be operators in $(\mathcal{A}, \varphi)$. If $\{\mathrm{Re}(c_1), \mathrm{Im}(c_1), \dots, \mathrm{Re}(c_n), \mathrm{Im}(c_n)\}$ is a semi-circular family, we say that $\{c_1, \dots, c_n\}$ is a *circular family*. We are allowing the possibility that some of $\mathrm{Re}(c_i)$ or $\mathrm{Im}(c_i)$ is 0. So a semi-circular family is a circular family.

Exercise 1. Using the notation of Section 6.8, show that for $\{c_1, \dots, c_n\}$ to be a circular family, it is necessary and sufficient that for every $i_1, \dots, i_m \in [n]$ and every $\epsilon_1, \dots, \epsilon_m \in \{-1, 1\}$ we have

$$\varphi(c_{i_1}^{(\epsilon_1)} \cdots c_{i_m}^{(\epsilon_m)}) = \sum_{\pi \in NC_2(m)} \kappa_\pi(c_{i_1}^{(\epsilon_1)}, \dots, c_{i_m}^{(\epsilon_m)}).$$

Let us consider the more general situation where X is a $d \times d$ matrix $X = (s_{ij})_{i,j=1}^d$, where $\{s_{ij}\}$ is a circular family with a covariance function σ, i.e.

$$\varphi(s_{ij} s_{kl}) = \sigma(i, j; k, l). \tag{9.3}$$

The covariance function σ can here be prescribed quite arbitrarily, only subject to some symmetry conditions in order to ensure that X is self-adjoint. Thus, we allow arbitrary correlations between different entries, but also that the variance of the s_{ij} depends on (i, j). Note that we do not necessarily ask that all entries are semi-circular. Off-diagonal elements can also be circular elements, as long as we have $s_{ij}^* = s_{ji}$.

By Exercise 1, we have

$$\begin{aligned} \mathrm{tr}_d \otimes \varphi(X^m) &= \frac{1}{d} \sum_{i(1),\dots,i(m)=1}^{d} \varphi\big[s_{i_1 i_2} \cdots s_{i_m i_1}\big] \\ &= \frac{1}{d} \sum_{\pi \in NC_2(m)} \sum_{i(1),\dots,i(m)=1}^{d} \prod_{(p,q)\in\pi} \sigma\big(i_p, i_{p+1}; i_q, i_{q+1}\big). \end{aligned}$$

We can write this in the form

$$\mathrm{tr}_d \otimes \varphi(X^m) = \sum_{\pi \in NC_2(m)} \mathcal{K}_\pi,$$

where

$$\mathcal{K}_\pi := \frac{1}{d} \sum_{i_1,\dots,i_m=1}^{d} \prod_{(p,q)\in\pi} \sigma\big(i_p, i_{p+1}; i_q, i_{q+1}\big).$$

So the result looks very similar to our usual description of semi-circular elements, in terms of a sum over non-crossing pairings. However, the problem here is that the $\mathcal{K}_\pi$ are not multiplicative with respect to the block decomposition of π, and thus they do not qualify to be considered as cumulants. Even worse, there does not exist a straightforward recursive way of expressing $\mathcal{K}_\pi$ in terms of "smaller" $\mathcal{K}_\sigma$. Thus, we are outside the realm of the usual recursive techniques of free probability theory.

However, one can save most of those techniques by going to an "operator-valued" level. The main point of such an operator-valued approach is to write $\mathcal{K}_\pi$ as the trace of a $d \times d$-matrix κ_π, and then realize that κ_π has the usual nice recursive structure.

Namely, let us define the matrix $\kappa_\pi = ([\kappa_\pi]_{ij})_{i,j=1}^{d}$ by

$$[\kappa_\pi]_{ij} := \sum_{i_1\dots,i_m,i_{m+1}=1}^{d} \delta_{i i_1}\delta_{j i_{m+1}} \prod_{(p,q)\in\pi} \sigma\big(i_p, i_{p+1}; i_q, i_{q+1}\big).$$

Then clearly we have $\mathcal{K}_\pi = \mathrm{tr}_d(\kappa_\pi)$. Furthermore, the value of κ_π can be determined by an iterated application of the *covariance mapping*

$$\eta : M_d(\mathbb{C}) \to M_d(\mathbb{C}) \qquad \text{given by} \qquad \eta(B) := id \otimes \varphi[XBX],$$

i.e. for $B = (b_{ij}) \in M_d(\mathbb{C})$, we have $\eta(B) = ([\eta(B)]_{ij}) \in M_d(\mathbb{C})$ with

$$[\eta(B)]_{ij} = \sum_{k,l=1}^{d} \sigma(i,k;l,j) b_{kl}.$$

The main observation is now that the value of κ_π is given by an iterated application of this mapping η according to the nesting of the blocks of π. If one identifies a non-crossing pairing with an arrangement of brackets, then the way that η has to be iterated is quite obvious. Let us clarify these remarks with an example.

Consider the non-crossing pairing

$$\pi = \{(1,4),(2,3),(5,6)\} \in NC_2(6).$$

The corresponding κ_π is given by

$$[\kappa_\pi]_{ij} = \sum_{i_2,i_3,i_4,i_5,i_6=1}^{d} \sigma(i,i_2;i_4,i_5)\cdot\sigma(i_2,i_3;i_3,i_4)\cdot\sigma(i_5,i_6;i_6,j).$$

We can then sum over the index i_3 (corresponding to the block $(2,3)$ of π) without interfering with the other blocks, giving

$$
\begin{aligned}
[\kappa_\pi]_{ij} &= \sum_{i_2,i_4,i_5,i_6=1}^{d} \sigma(i,i_2;i_4,i_5)\cdot\sigma(i_5,i_6;i_6,j)\cdot\sum_{i_3=1}^{d}\sigma(i_2,i_3;i_3,i_4)\\
&= \sum_{i_2,i_4,i_5,i_6=1}^{d} \sigma(i,i_2;i_4,i_5)\cdot\sigma(i_5,i_6;i_6,j)\cdot[\eta(1)]_{i_2i_4}.
\end{aligned}
$$

Effectively we have removed the block $(2,3)$ of π and replaced it by the matrix $\eta(1)$.

Now we can do the summation over $i(2)$ and $i(4)$ without interfering with the other blocks, thus yielding

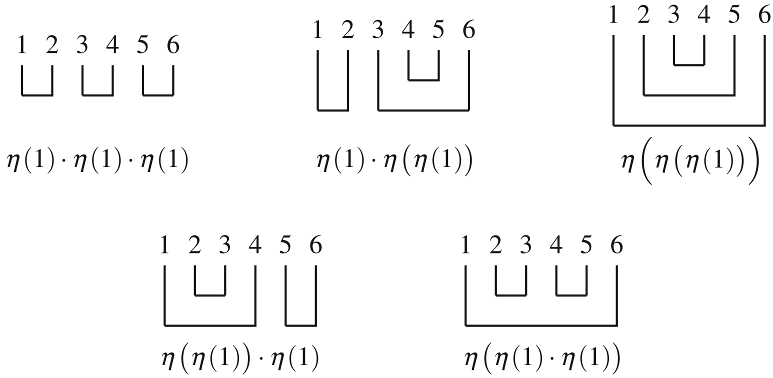
$$
\begin{aligned}
[\kappa_\pi]_{ij} &= \sum_{i_5,i_6=1}^{d} \sigma(i_5,i_6;i_6,j)\cdot\sum_{i_2,i_4=1}^{d}\sigma(i,i_2;i_4,i_5)\cdot[\eta(1)]_{i_2i_4}\\
&= \sum_{i_5,i_6=1}^{d} \sigma(i_5,i_6;i_6,j)\cdot\big[\eta\big(\eta(1)\big)\big]_{ii_5}.
\end{aligned}
$$

We have now removed the block $(1,4)$ of π, and the effect of this was that we had to apply η to whatever was embraced by this block (in our case, $\eta(1)$).

Finally, we can do the summation over i_5 and i_6 corresponding to the last block $(5,6)$ of π; this results in

$$
\begin{aligned}
[\kappa_\pi]_{i,j} &= \sum_{i_5=1}^{d}\big[\eta\big(\eta(1)\big)\big]_{ii_5}\cdot\sum_{i_6=1}^{d}\sigma(i_5,i_6;i_6,j)\\
&= \sum_{i_5=1}^{d}\big[\eta\big(\eta(1)\big)\big]_{ii_5}\cdot[\eta(1)]_{i_5j}\\
&= \big[\eta\big(\eta(1)\big)\cdot\eta(1)\big]_{ij}.
\end{aligned}
$$

Thus, we finally have $\kappa_\pi = \eta\big(\eta(1)\big)\cdot\eta(1)$, which corresponds to the bracket expression $(X(XX)X)(XX)$. In the same way, every non-crossing pairing results in an iterated application of the mapping η. For the five non-crossing pairings of six elements, one gets the following results:



Thus, for $m = 6$, we get for $\mathrm{tr}_d \otimes \varphi(X^6)$ the expression

$$\mathrm{tr}_d \Big\{ \eta(1)\cdot\eta(1)\cdot\eta(1) + \eta(1)\cdot\eta\big(\eta(1)\big) + \\ + \eta\big(\eta(1)\big)\cdot\eta(1) + \eta\big(\eta(1)\cdot\eta(1)\big) + \eta\big(\eta\big(\eta(1)\big)\big)\Big\}.$$

Let us summarize our calculations for general moments. We have

$$\mathrm{tr}_d \otimes \varphi(X^m) = \mathrm{tr}_d \Big\{ \sum_{\pi \in NC_2(m)} \kappa_\pi \Big\},$$

where each κ_π is a $d \times d$ matrix, determined in a recursive way as above, by an iterated application of the mapping η. If we remove tr_d from this equation, then we get formally the equation for a semi-circular distribution. Define

$$E := id \otimes \varphi : M_d(\mathcal{C}) \to M_d(\mathbb{C}),$$

and then we have that the operator-valued moments of X satisfy

$$E(X^m) = \sum_{\pi \in NC_2(m)} \kappa_\pi. \tag{9.4}$$

An element X whose operator-valued moments $E(X^m)$ are calculated in such a way is called an *operator-valued semi-circular element* (because only pairings are needed).

One can now repeat essentially all combinatorial arguments from the scalar situation in this case. One only has to take care that the nesting of the blocks of π is respected. Let us try this for the reformulation of the relation (9.4) in terms of formal power series. We are using the usual argument by doing the summation over all $\pi \in NC_2(m)$ by collecting terms according to the block containing the first element 1. If π is a non-crossing pairing of m elements and $(1, r)$ is the block of π containing 1, then the remaining blocks of π must fall into two classes, those making up a non-crossing pairing of the numbers $2, 3, \ldots, r-1$ and those making up a non-crossing pairing of the numbers $r+1, r+2, \ldots, m$. Let us call the former pairing π_1 and the latter π_2, so that we can write $\pi = (1, r) \cup \pi_1 \cup \pi_2$. Then the description above of κ_π shows that $\kappa_\pi = \eta(\kappa_{\pi_1}) \cdot \kappa_{\pi_2}$. This results in the following recurrence relation for the operator-valued moments:

$$E[X^m] = \sum_{k=0}^{m-2} \eta\big(E[X^k]\big) \cdot E[X^{m-k-2}].$$

If we go over to the corresponding generating power series,

$$M(z) = \sum_{m=0}^{\infty} E[X^m] z^m,$$

then this yields the relation $M(z) = 1 + z^2 \eta\big(M(z)\big) \cdot M(z)$.

Note that $m(z) := \mathrm{tr}_d(M(z))$ is the generating power series of the moments $\mathrm{tr}_d \otimes \varphi(X^m)$, in which we are ultimately interested. Thus, it is preferable to go over from $M(z)$ to the corresponding operator-valued Cauchy transform $G(z) := z^{-1} M(1/z)$. For this the equation above takes on the form

$$zG(z) = 1 + \eta(G(z)) \cdot G(z). \tag{9.5}$$

Furthermore, we have for the Cauchy transform g of the limiting eigenvalue distribution μ_X of our block matrices X_N that

$$g(z) = z^{-1} m(1/z) = \mathrm{tr}_d \left(z^{-1} M(1/z)\right) = \mathrm{tr}_d(G(z)).$$

Since the number of non-crossing pairings of $2k$ elements is given by the Catalan number C_k, for which one has $C_k \leq 4^k$, we can estimate the (operator) norm of the matrix $E(X^{2k})$ by

$$\|E(X^{2k})\| \leq \|\eta\|^k \cdot \#(NC_2(2k)) \leq \|\eta\|^k \cdot 2^{2k}.$$

Applying tr_d, this yields that the support of the limiting eigenvalue distribution of X_N is contained in the interval $[-2\|\eta\|^{1/2}, +2\|\eta\|^{1/2}]$. Since all odd moments are zero, the measure is symmetric. Furthermore, the estimate above on the operator-valued moments $E(X^m)$ shows that

$$G(z) = \sum_{k=0}^{\infty} \frac{E(X^{2k})}{z^{2k+1}}$$

is a power series expansion in $1/z$ of $G(z)$, which converges in a neighbourhood of ∞. Since on bounded sets, $\{B \in M_d(\mathbb{C}) \mid \|B\| \leq K\}$ for some $K > 0$, the mapping

$$B \mapsto z^{-1}1 + z^{-1}\eta(B) \cdot B$$

is a contraction for $|z|$ sufficiently large, $G(z)$ is, for large z, uniquely determined as the solution of the equation (9.5).

If we write G as $G(z) = E\big((z - X)^{-1}\big)$, then this shows that it is not only a formal power series but actually an analytic ($M_d(\mathbb{C})$-valued) function on the whole upper complex half-plane. Analytic continuation shows then the validity of (9.5) for all z in the upper half-plane.

Let us summarize our findings in the following theorem, which was proved in [147].

Theorem 2. *Fix $d \in \mathbb{N}$. Consider, for each $N \in \mathbb{N}$, block matrices*

$$X_N = \begin{pmatrix} A^{(11)} & \dots & A^{(1d)} \\ \vdots & \ddots & \vdots \\ A^{(d1)} & \dots & A^{(dd)} \end{pmatrix} \tag{9.6}$$

where, for each $i, j = 1, \dots, d$, the blocks $A^{(ij)} = \big(a_{rp}^{(ij)}\big)_{r,p=1}^{N}$ are Gaussian $N \times N$ random matrices such that the collection of all entries

$$\{a_{rp}^{(ij)} \mid i, j = 1, \dots, d;\ r, p = 1, \dots, N\}$$

of the matrix X_N forms a Gaussian family which is determined by

$$a_{rp}^{(ij)} = \overline{a_{pr}^{(ji)}} \qquad \text{for all } i, j = 1, \dots, d;\ r, p = 1, \dots, N$$

and the prescription of mean zero and covariance

$$E[a_{rp}^{(ij)} a_{qs}^{(kl)}] = \frac{1}{n}\delta_{rs}\delta_{pq} \cdot \sigma(i, j; k, l), \tag{9.7}$$

where $n := dN$.

Then, for $N \to \infty$, the $n \times n$ matrix X_N has a limiting eigenvalue distribution whose Cauchy transform g is determined by $g(z) = \mathrm{tr}_d(G(z))$, where G is an $M_d(\mathbb{C})$-valued analytic function on the upper complex half-plane, which is uniquely determined by the requirement that for $z \in \mathbb{C}^+$

$$\lim_{|z|\to\infty} zG(z) = 1, \tag{9.8}$$

(where 1 is the identity of $M_d(\mathbb{C})$) and that for all $z \in \mathbb{C}^+$, G satisfies the matrix equation (9.5).

Note also that in [94], it was shown that there exists exactly one solution of the fixed point equation (9.5) with a certain positivity property.

There exists a vast literature on dealing with such or similar generalizations of Gaussian random matrices. Most of them deal with the situation where the entries are still independent, but not identically distributed; usually, such matrices are referred to as *band matrices*. The basic insight that such questions can be treated within the framework of operator-valued free probability theory is due to Shlyakhtenko [155]. A very extensive treatment of band matrices (not using the language of free probability, but the quite related Wigner-type moment method) was given by Anderson and Zeitouni [6].

Example 3. Let us now reconsider the limit (9.2) of our motivating band matrix (9.1). Since there are some symmetries in the block pattern, the corresponding G will also have some additional structure. To work this out, let us examine η more carefully. If $B \in M_3(\mathbb{C})$, $B = (b_{ij})_{ij}$, then

$$\eta(B) = \frac{1}{3}\begin{pmatrix} b_{11}+b_{22}+b_{33} & b_{12}+b_{21}+b_{23} & b_{13}+b_{31}+b_{22} \\ b_{21}+b_{12}+b_{32} & b_{11}+b_{22}+b_{33}+b_{13}+b_{31} & b_{12}+b_{23}+b_{32} \\ b_{13}+b_{31}+b_{22} & b_{23}+b_{32}+b_{21} & b_{11}+b_{22}+b_{33} \end{pmatrix}.$$

We shall see later on that it is important to find the smallest unital subalgebra $\mathcal{C}$ of $M_3(\mathbb{C})$ that is invariant under η. We have

$$\eta(1) = \begin{pmatrix} 1 & 0 & \frac{1}{3} \\ 0 & 1 & 0 \\ \frac{1}{3} & 0 & 1 \end{pmatrix} = 1 + \frac{1}{3}H, \qquad \text{where } H = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix},$$

$$\eta(H) = \frac{1}{3}\begin{pmatrix} 0 & 0 & 2 \\ 0 & 2 & 0 \\ 2 & 0 & 0 \end{pmatrix} = \frac{2}{3}H + \frac{2}{3}E, \qquad \text{where } E = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

and

$$\eta(E) = \frac{1}{3}\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} = \frac{1}{3}1 + \frac{1}{3}H.$$

Now $HE = EH = 0$ and $H^2 = 1 - E$, so $\mathcal{C}$, the span of $\{1, H, E\}$, is a three-dimensional commutative subalgebra invariant under η. Let us show that if G satisfies $zG(z) = 1 + \eta(G(z))G(z)$ and is analytic, then $G(z) \in \mathcal{C}$ for all $z \in \mathbb{C}^+$.

Let $\Phi : M_3(\mathbb{C}) \to M_3(\mathbb{C})$ be given by $\Phi(B) = z^{-1}(1 + \eta(B)B)$. One easily checks that

$$\|\Phi(B)\| \le |z|^{-1}(1 + \|\eta\|\|B\|^2)$$

and

$$\|\Phi(B_1) - \Phi(B_2)\| \le |z|^{-1}\|\eta\|(\|B_1\| + \|B_2\|)\|B_1 - B_2\|.$$

Here $\|\eta\|$ is the norm of η as a map from $M_3(\mathbb{C})$ to $M_3(\mathbb{C})$. Since η is completely positive, we have $\|\eta\| = \|\eta(1)\|$. In this particular example, $\|\eta\| = 4/3$.

Now let $\mathcal{D}_\epsilon = \{B \in M_3(\mathbb{C}) \mid \|B\| < \epsilon\}$. If the pair $z \in \mathbb{C}^+$ and $\epsilon > 0$ simultaneously satisfies

$$1 + \|\eta\|\epsilon^2 < |z|\epsilon \qquad \text{and} \qquad 2\epsilon\|\eta\| < |z|,$$

then $\Phi(\mathcal{D}_\epsilon) \subseteq \mathcal{D}_\epsilon$ and $\|\Phi(B_1) - \Phi(B_2)\| \le c\|B_1 - B_2\|$ for $B_1, B_2 \in \mathcal{D}_\epsilon$ and $c = 2\epsilon|z|^{-1}\|\eta\| < 1$. So when $|z|$ is sufficiently large, both conditions are satisfied and Φ has a unique fixed point in $\mathcal{D}_\epsilon$. If we choose $B \in \mathcal{D}_\epsilon \cap \mathcal{C}$, then all iterates of Φ applied to B will remain in $\mathcal{C}$, and so the unique fixed point will be in $\mathcal{D}_\epsilon \cap \mathcal{C}$.

Since $M_3(\mathbb{C})$ is finite-dimensional, there are a finite number of linear functionals, $\{\varphi_i\}_i$, on $M_3(\mathbb{C})$ (6 in our particular example) such that $\mathcal{C} = \cap_i \ker(\varphi_i)$. Also for each i, $\varphi_i \circ G$ is analytic so it is identically 0 on $\mathbb{C}^+$ if it vanishes on a non-empty open subset of $\mathbb{C}^+$. We have seen above that $G(z) \in \mathcal{C}$ provided $|z|$ is sufficiently large; thus $G(z) \in \mathcal{C}$ for all $z \in \mathbb{C}^+$.

Hence, G and $\eta(G)$ must be of the form

$$G = \begin{pmatrix} f & 0 & h \\ 0 & e & 0 \\ h & 0 & f \end{pmatrix}, \quad \eta(G) = \frac{1}{3}\begin{pmatrix} 2f+e & 0 & e+2h \\ 0 & 2f+e+2h & 0 \\ e+2h & 0 & 2f+e \end{pmatrix}.$$

So Equation (9.5) gives the following system of equations:

$$\begin{aligned} zf &= 1 + \frac{e\,(f+h) + 2\,(f^2+h^2)}{3}, \\ ze &= 1 + \frac{e\,(e+2\,(f+h))}{3}, \\ zh &= \frac{4fh + e\,(f+h)}{3}. \end{aligned} \tag{9.9}$$

This system of equations can be solved numerically for z close to the real axis; then

$$g(z) = \mathrm{tr}_3\big(G(z)\big) = (2f(z)+e(z))/3, \quad \frac{d\mu(t)}{dt} = -\frac{1}{\pi}\lim_{s\to 0} \mathrm{Im}g\,(t+is) \tag{9.10}$$

gives the sought eigenvalue distribution. In Fig. 9.2 we compare this numerical solution (solid curve) with the histogram for the X_N from Fig. 9.1, with blocks of size 1000×1000.

9.2 General theory of operator-valued free probability

Not only semi-circular elements can be lifted to an operator-valued level, but such a generalization exists for the whole theory. The foundation for this was laid by Voiculescu in [184]; Speicher showed in [163] that the combinatorial description of

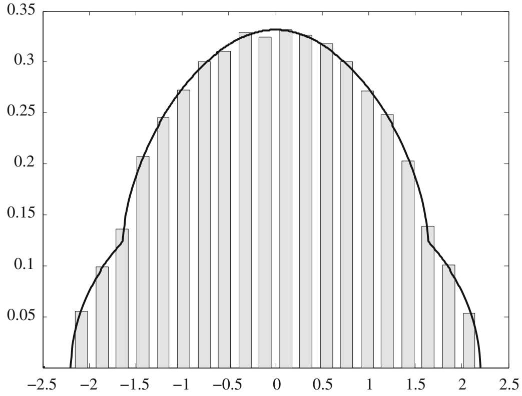


Fig. 9.2 Comparison of the histogram of eigenvalues of X_N, from Fig. 9.1, with the numerical solution according to (9.9) and (9.10)

free probability resting on the notion of free cumulants extends also to the operator-valued case. We want to give here a short survey of some definitions and results.

Definition 4. Let $\mathcal{A}$ be a unital algebra and consider a unital subalgebra $\mathcal{B} \subset \mathcal{A}$. A linear map $E : \mathcal{A} \to \mathcal{B}$ is a *conditional expectation* if

$$E(b) = b \qquad \forall b \in \mathcal{B} \tag{9.11}$$

and

$$E(b_1 a b_2) = b_1 E(a) b_2 \qquad \forall a \in \mathcal{A}, \quad \forall b_1, b_2 \in \mathcal{B}. \tag{9.12}$$

An *operator-valued probability space* $(\mathcal{A}, E, \mathcal{B})$ consists of $\mathcal{B} \subset \mathcal{A}$ and a conditional expectation $E : \mathcal{A} \to \mathcal{B}$.

The *operator-valued distribution* of a random variable $x \in \mathcal{A}$ is given by all *operator-valued moments* $E(xb_1xb_2 \cdots b_{n-1}x) \in \mathcal{B}$ ($n \in \mathbb{N}, b_1, \ldots, b_{n-1} \in \mathcal{B}$).

Since, by the bimodule property (9.12),

$$E(b_0 x b_1 x b_2 \cdots b_{n-1} x b_n) = b_0 \cdot E(x b_1 x b_2 \cdots b_{n-1} x) \cdot b_n,$$

there is no need to include b_0 and b_n in the operator-valued distribution of x.

Definition 5. Consider an operator-valued probability space $(\mathcal{A}, E, \mathcal{B})$ and a family $(\mathcal{A}_i)_{i \in I}$ of subalgebras with $\mathcal{B} \subset \mathcal{A}_i$ for all $i \in I$. The subalgebras $(\mathcal{A}_i)_{i \in I}$ are *free with respect to* E or *free with amalgamation over* $\mathcal{B}$ if $E(a_1 \cdots a_n) = 0$ whenever $a_i \in \mathcal{A}_{j_i}$, $j_1 \neq j_2 \neq \cdots \neq j_n$, and $E(a_i) = 0$ for all $i = 1, \ldots, n$. Random variables in $\mathcal{A}$ or subsets of $\mathcal{A}$ are free with amalgamation over $\mathcal{B}$ if the algebras generated by $\mathcal{B}$ and the variables or the algebras generated by $\mathcal{B}$ and the subsets, respectively, are so.

Note that the subalgebra generated by $\mathcal{B}$ and some variable x is not just the linear span of monomials of the form bx^n, but, because elements from $\mathcal{B}$ and our variable x do not commute in general, we must also consider general monomials of the form $b_0xb_1x\cdots b_nxb_{n+1}$.

If $\mathcal{B} = \mathcal{A}$, then any two subalgebras of $\mathcal{A}$ are free with amalgamation over $\mathcal{B}$; so the claim of freeness with amalgamation gets weaker as the subalgebra gets larger until the subalgebra is the whole algebra at which point the claim is empty.

Operator-valued freeness works mostly like ordinary freeness, one only has to take care of the order of the variables; in all expressions, they have to appear in their original order!

Example 6. 1) If x and $\{y_1, y_2\}$ are free, then one has as in the scalar case

$$E(y_1xy_2) = E\big(y_1E(x)y_2\big); \tag{9.13}$$

and more general, for $b_1, b_2 \in \mathcal{B}$,

$$E(y_1b_1xb_2y_2) = E\big(y_1b_1E(x)b_2y_2\big). \tag{9.14}$$

In the scalar case (where $\mathcal{B}$ would just be $\mathbb{C}$ and $E = \varphi : \mathcal{A} \to \mathbb{C}$ a unital linear functional), we write of course $\varphi\big(y_1\varphi(x)y_2\big)$ in the factorized form $\varphi(y_1y_2)\varphi(x)$. In the operator-valued case, this is not possible; we have to leave the $E(x)$ at its position between y_1 and y_2.

2) If $\{x_1, x_2\}$ and $\{y_1, y_2\}$ are free over $\mathcal{B}$, then one has the operator-valued version of (1.14),

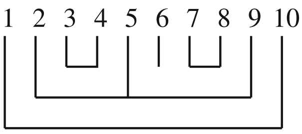
$$\begin{aligned} E(x_1y_1x_2y_2) = {} & E\big(x_1E(y_1)x_2\big)\cdot E(y_2) + E(x_1)\cdot E\big(y_1E(x_2)y_2\big) \\ & -E(x_1)E(y_1)E(x_2)E(y_2). \end{aligned} \tag{9.15}$$

Definition 7. Consider an operator-valued probability space $(\mathcal{A}, E, \mathcal{B})$. We define the corresponding *(operator-valued) free cumulants* $(\kappa_n^{\mathcal{B}})_{n\in\mathbb{N}}$, $\kappa_n^{\mathcal{B}} : \mathcal{A}^n \to \mathcal{B}$, by the moment-cumulant formula

$$E(a_1\cdots a_n) = \sum_{\pi\in NC(n)} \kappa_\pi^{\mathcal{B}}(a_1, \ldots, a_n), \tag{9.16}$$

where arguments of $\kappa_\pi^{\mathcal{B}}$ are distributed according to the blocks of π, but the cumulants are nested inside each other according to the nesting of the blocks of π.

Example 8. Consider the non-crossing partition



$$\pi = \big\{(1,10),(2,5,9),(3,4),(6),(7,8)\big\} \in NC(10).$$

The corresponding free cumulant $\kappa_\pi^{\mathcal{B}}$ is given by

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

Remark 9. Let us give a more formal definition of the operator-valued free cumulants in the following.

1) First note that the bimodule property (9.12) for E implies for $\kappa^{\mathcal{B}}$ the property

$$\kappa_n^{\mathcal{B}}(b_0 a_1, b_1 a_2, \dots, b_n a_n b_{n+1}) = b_0 \kappa_n^{\mathcal{B}}(a_1 b_1, a_2 b_2, \dots, a_n) b_{n+1}$$

for all $a_1, \dots, a_n \in \mathcal{A}$ and $b_0, \dots, b_{n+1} \in \mathcal{B}$. This can also stated by saying that $\kappa_n^{\mathcal{B}}$ is actually a map on the $\mathcal{B}$-module tensor product

$$\mathcal{A}^{\otimes_{\mathcal{B}} n} = \mathcal{A} \otimes_{\mathcal{B}} \mathcal{A} \otimes_{\mathcal{B}} \cdots \otimes_{\mathcal{B}} \mathcal{A}.$$

2) Let now any sequence $\{T_n\}_n$ of $\mathcal{B}$-bimodule maps: $T_n : \mathcal{A}^{\otimes_{\mathcal{B}} n} \to \mathcal{B}$ be given. Instead of $T_n(x_1 \otimes_{\mathcal{B}} \cdots \otimes_{\mathcal{B}} x_n)$, we shall write $T_n(x_1, \dots, x_n)$. Then there exists a unique extension of T, indexed by non-crossing partitions, so that for every $\pi \in NC(n)$, we have a map $T_\pi : \mathcal{A}^{\otimes_{\mathcal{B}} n} \to \mathcal{B}$ so that the following conditions are satisfied:

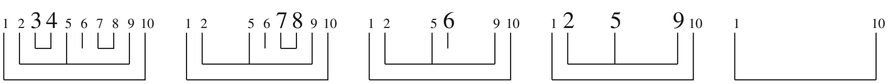
   (i) when $\pi = 1_n$, we have $T_\pi = T_n$;
   (ii) whenever $\pi \in NC(n)$ and $V = \{l+1, \dots, l+k\}$ is an interval in π then

$$\begin{aligned} T_\pi(x_1, \dots, x_n) &= T_{\pi'}(x_1, \dots, x_l T_k(x_{l+1}, \dots, x_{l+k}), x_{l+k+1}, \dots, x_n) \\ &= T_{\pi'}(x_1, \dots, x_l, T_k(x_{l+1}, \dots, x_{l+k}) x_{l+k+1}, \dots, x_n), \end{aligned}$$

   where $\pi' \in NC(n-k)$ is the partition obtained by deleting from π the block V. When $l = 0$, we interpret this property to mean

$$T_\pi(x_1, \dots, x_n) = T_{\pi'}(T_k(x_1, \dots, x_k) x_{k+1}, \dots, x_n).$$

This second property is called the *insertion property*. One should notice that every non-crossing partition can be reduced to a partition with a single block by the process of *interval stripping*. For example, with the partition $\pi = \{(1,10), (2,5,9), (3,4), (6), (7,8)\}$ from above, we strip the interval $(3,4)$ to obtain $\{(1,10), (2,5,9), (6), (7,8)\}$. We strip the interval $(7,8)$ to obtain $\{(1,10), (2,5,9), (6), \}$, then we strip the (one element) interval (6) to obtain $\{(1,10), (2,5,9)\}$, and finally we strip the interval $(2,5,9)$ to obtain the partition with a single block $\{(1,10)\}$.



The insertion property requires that the family $\{T_\pi\}_\pi$ be compatible with interval stripping. Thus, if there is an extension satisfying (*i*) and (*ii*), it must be unique. Moreover, we can compute T_π by stripping intervals, and the outcome is independent of the order in which we strip the intervals.

3) Let us call a family $\{T_\pi\}_\pi$ determined as above *multiplicative*. Then it is quite straightforward to check the following.
   - Let $\{T_\pi\}_\pi$ be a multiplicative family of $\mathcal{B}$-bimodule maps and define a new family by
   $$S_\pi = \sum_{\substack{\sigma \in NC(n) \\ \sigma \le \pi}} T_\sigma \qquad (\pi \in NC(n)). \tag{9.17}$$
   Then the family $\{S_\pi\}_\pi$ is also multiplicative.
   - The relation (9.17) between two multiplicative families is via Möbius inversions also equivalent to
   $$T_\pi = \sum_{\substack{\sigma \in NC(n) \\ \sigma \le \pi}} \mu(\sigma, \pi) S_\sigma \qquad (\pi \in NC(n)), \tag{9.18}$$
   where μ is the Möbius function on non-crossing partitions; see Remark 2.9. Again, multiplicativity of $\{S_\pi\}_\pi$ implies multiplicativity of $\{T_\pi\}_\pi$, if the latter is defined in terms of the former via (9.18).

4) Now we can use the previous to define the free cumulants $\kappa_n^{\mathcal{B}}$. As a starting point, we use the multiplicative family $\{E_\pi\}_\pi$ which is given by the "moment maps"
$$E_n : \mathcal{A}^{\otimes_{\mathcal{B}} n} \to \mathcal{B}, \qquad E_n(a_1, a_2, \dots, a_n) = E(a_1 a_2 \cdots a_n).$$
For $\pi = \{(1,10),(2,5,9),(3,4),(6),(7,8)\} \in NC(10)$ from Example 8, the E_π is, for example, given by
$$E_\pi(a_1, \dots, a_{10}) = E\Big(a_1 \cdot E\big(a_2 \cdot E(a_3 a_4) \cdot a_5 \cdot E(a_6) \cdot E(a_7 a_8) \cdot a_9\big) \cdot a_{10}\Big).$$
Then we define the multiplicative family $\{\kappa_\pi^{\mathcal{B}}\}_\pi$ by
$$\kappa_\pi^{\mathcal{B}} = \sum_{\substack{\sigma \in NC(n) \\ \sigma \le \pi}} \mu(\sigma, \pi) E_\sigma \qquad (\pi \in NC(n)),$$
which is equivalent to (9.16). In particular, this means that the $\kappa_n^{\mathcal{B}}$ are given by
$$\kappa_n^{\mathcal{B}}(a_1, \dots, a_n) = \sum_{\pi \in NC(n)} \mu(\pi, 1_n) E_\pi(a_1, \dots, a_n). \tag{9.19}$$

Definition 10. 1) For $a \in \mathcal{A}$ we define its *(operator-valued) Cauchy transform* $G_a : \mathcal{B} \to \mathcal{B}$ by

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

and its *(operator-valued) R-transform* $R_a : \mathcal{B} \to \mathcal{B}$ by

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

2) We say that $s \in \mathcal{A}$ is *$\mathcal{B}$-valued semi-circular* if $\kappa_n^{\mathcal{B}}(sb_1, sb_2, \ldots, sb_{n-1}, s) = 0$ for all $n \neq 2$, and all $b_1, \ldots, b_{n-1} \in \mathcal{B}$.

If $s \in \mathcal{A}$ is $\mathcal{B}$-valued semi-circular, then by the moment-cumulant formula, we have

$$E(s^n) = \sum_{\pi \in NC_2(n)} \kappa_\pi(s, \ldots, s).$$

This is consistent with (9.4) of our example $\mathcal{A} = M_d(\mathcal{C})$ and $\mathcal{B} = M_d(\mathbb{C})$, where these κ's were defined by iterated applications of $\eta(B) = E(XBX) = \kappa_2^{\mathcal{B}}(XB, X)$.

As in the scalar-valued case, one has the following properties; see [163, 184, 190].

Theorem 11. *1) The relation between the Cauchy and the R-transform is given by*

$$bG(b) = 1 + R(G(b)) \cdot G(b) \qquad \text{or} \qquad G(b) = (b - R(G(b)))^{-1}. \tag{9.20}$$

2) Freeness of x and y over $\mathcal{B}$ is equivalent to the vanishing of mixed $\mathcal{B}$-valued cumulants in x and y. This implies, in particular, the additivity of the R-transform: $R_{x+y}(b) = R_x(b) + R_y(b)$, if x and y are free over $\mathcal{B}$.

3) If x and y are free over $\mathcal{B}$, then we have the subordination property

$$G_{x+y}(b) = G_x\big[b - R_y\big(G_{x+y}(b)\big)\big]. \tag{9.21}$$

4) If s is an operator-valued semi-circular element over $\mathcal{B}$, then $R_s(b) = \eta(b)$, where $\eta : \mathcal{B} \to \mathcal{B}$ is the linear map given by $\eta(b) = E(sbs)$.

Remark 12. 1) As for the moments, one has to allow in the operator-valued cumulants elements from $\mathcal{B}$ to spread everywhere between the arguments. So with $\mathcal{B}$-valued cumulants in random variables $x_1, \ldots, x_r \in \mathcal{A}$, we actually mean all expressions of the form $\kappa_n^{\mathcal{B}}(x_{i_1}b_1, x_{i_2}b_2, \ldots, x_{i_{n-1}}b_{n-1}, x_{i_n})$ ($n \in \mathbb{N}$, $1 \leq i(1), \ldots, i(n) \leq r$, $b_1, \ldots, b_{n-1} \in \mathcal{B}$).

2) One might wonder about the nature of the operator-valued Cauchy and R-transforms. One way to interpret the definitions and the statements is as convergent power series. For this one needs a Banach algebra setting, and then everything can be justified as convergent power series for appropriate b, namely, with $\|b\|$ sufficiently small in the R-transform case and with b invertible and $\|b^{-1}\|$ sufficiently small in the Cauchy transform case. In those domains, they are B-valued analytic functions and such F have a series expansion of the form (say F is analytic in a neighbourhood of $0 \in \mathcal{B}$)

$$F(b) = F(0) + \sum_{k=1}^{\infty} F_k(b, \ldots, b), \tag{9.22}$$

where F_k is a symmetric multilinear function from the k-fold product $\mathcal{B} \times \cdots \times \mathcal{B}$ to $\mathcal{B}$. In the same way as for usual formal power series, one can consider (9.22) as a formal multilinear function series (given by the sequence $(F_k)_k$ of the coefficients of F), with the canonical definitions for sums, products, and compositions of such series. One can then also read Definition 10 and Theorem 11 as statements about such formal multilinear function series. For a more thorough discussion of this point of view (and more results about operator-valued free probability), one should consult the work of Dykema [68].

As illuminated in Section 9.1 for the case of an operator-valued semi-circle, many statements from the scalar-valued version of free probability are still true in the operator-valued case; actually, on a combinatorial (or formal multilinear function series) level, the proofs are essentially the same as in the scalar-valued case, and one only has to take care that one respects the nested structure of the blocks of non-crossing partitions. One can also extend some of the theory to an analytic level. In particular, the operator-valued Cauchy transform is an analytic operator-valued function (in the sense of Fréchet-derivatives) on the *operator upper half-plane* $\mathbb{H}^+(\mathcal{B}) := \{b \in \mathcal{B} \mid \mathrm{Im}(b) > 0 \text{ and invertible}\}$. In the next chapter, we will have something more to say about this, when coming back to the analytic theory of operator-valued convolution.

One should, however, note that the analytic theory of operator-valued free convolution lacks at the moment some of the deeper statements of the scalar-valued theory; developing a reasonable analogue of complex function theory on an operator-valued level, addressed as *free analysis*, is an active area in free probability (and also other areas) at the moment; see, for example, [107, 193–195, 202].

9.3 Relation between scalar-valued and matrix-valued cumulants

Let us now present a relation from [140] between matrix-valued and scalar-valued cumulants, which shows that taking matrices of random variables goes nicely with freeness, at least if we allow for the operator-valued version.

Proposition 13. *Let $(\mathcal{C}, \varphi)$ be a non-commutative probability space and fix $d \in \mathbb{N}$. Then $(\mathcal{A}, E, \mathcal{B})$, with*

$$\mathcal{A} := M_d(\mathcal{C}), \qquad \mathcal{B} := M_d(\mathbb{C}) \subset M_d(\mathcal{C}), \qquad E := id \otimes \varphi : M_d(\mathcal{C}) \to M_d(\mathbb{C}),$$

is an operator-valued probability space. We denote the scalar cumulants with respect to φ by κ and the operator-valued cumulants with respect to E by $\kappa^{\mathcal{B}}$. Consider now $a_{ij}^k \in \mathcal{C}$ $(i, j = 1, \ldots, d;\ k = 1, \ldots, n)$ and put, for each $k = 1, \ldots, n$, $A_k = (a_{ij}^k)_{i,j=1}^d \in M_d(\mathcal{C})$. Then the operator-valued cumulants of the A_k are given in terms of the cumulants of their entries as follows:

$$[\kappa_n^{\mathcal{B}}(A_1, A_2, \ldots, A_n)]_{ij} = \sum_{i_2,\ldots,i_n=1}^{d} \kappa_n\big(a_{i i_2}^1, a_{i_2 i_3}^2, \ldots, a_{i_n j}^n\big). \tag{9.23}$$

Proof: Let us begin by noting that

$$[E(A_1 A_2 \cdots A_n]_{ij} = \sum_{i_2,\ldots,i_n=1}^{d} \varphi\big(a_{i i_2}^1 a_{i_2 i_3}^2 \cdots a_{i_n j}^n\big).$$

Let $\pi \in NC(n)$ be a non-crossing partition; we claim that

$$[E_\pi(A_1, A_2, \ldots, A_n]_{ij} = \sum_{i_2,\ldots,i_n=1}^{d} \varphi_\pi\big(a_{i i_2}^1, a_{i_2 i_3}^2, \ldots, a_{i_n j}^n\big).$$

If π has two blocks: $\pi = \{(1, \ldots, k), (k + 1, \ldots, n)\}$, then this is just matrix multiplication. We then get the general case by using the insertion property and induction. By Möbius inversion, we have

$$\begin{aligned}
[\kappa_n^{\mathcal{B}}(A_1, A_2, \ldots, A_n]_{ij} &= \sum_{\pi \in NC(n)} \mu(\pi, 1_n)[E_\pi(A_1, A_2, \ldots, A_n]_{ij}. \\
&= \sum_{i_2,\ldots,i_n=1}^{d} \sum_{\pi \in NC(n)} \mu(\pi, 1_n)\varphi_\pi\big(a_{i i_2}^1, a_{i_2 i_3}^2, \ldots, a_{i_n j}^n\big) \\
&= \sum_{i_2,\ldots,i_n=1}^{d} \kappa_\pi\big(a_{i i_2}^1, a_{i_2 i_3}^2, \ldots, a_{i_n j}^n\big).
\end{aligned}$$

□

Corollary 14. *If the entries of two matrices are free in $(\mathcal{C}, \varphi)$, then the two matrices themselves are free with respect to $E : M_d(\mathcal{C}) \to M_d(\mathbb{C})$.*

Proof: Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be the subalgebras of $\mathcal{A}$ which are generated by $\mathcal{B}$ and by the respective matrix. Note that the entries of any matrix from $\mathcal{A}_1$ are free from the entries of any matrix from $\mathcal{A}_2$. We have to show that mixed $\mathcal{B}$-valued cumulants in those two algebras vanish. So consider $A_1, \dots, A_n$ with $A_k \in \mathcal{A}_{r(k)}$. We shall show that for all n and all $r(1), \dots, r(n) \in \{1, 2\}$, we have $\kappa_n^{\mathcal{B}}(A_1, \dots, A_n) = 0$ whenever the r's are not all equal. As before we write $A_k = (a_{ij}^k)$. By freeness of the entries, we have $\kappa_n(a_{ii_2}^1, a_{i_2 i_3}^2, \dots, a_{i_n j}^n) = 0$ whenever the r's are not all equal. Then by Theorem 13, the (i, j)-entry of $\kappa_n^{\mathcal{B}}(A_1, \dots, A_n)$ equals 0 and thus $\kappa_n^{\mathcal{B}}(A_1, \dots, A_n) = 0$ as claimed. □

Example 15. If $\{a_1, b_1, c_1, d_1\}$ and $\{a_2, b_2, c_2, d_2\}$ are free in $(\mathcal{C}, \varphi)$, then the proposition above says that

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

are free with amalgamation over $M_2(\mathbb{C})$ in $(M_2(\mathcal{C}), id \otimes \varphi)$. Note that in general they are not free in the scalar-valued non-commutative probability space $(M_2(\mathcal{C}), \mathrm{tr} \otimes \varphi)$. Let us make this distinction clear by looking on a small moment. We have

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

Applying the trace $\psi := \mathrm{tr} \otimes \varphi$, we get in general

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

but under the conditional expectation $E := id \otimes \varphi$, we always have

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

9.4 Moving between different levels

We have seen that in interesting problems, like random matrices with correlation between the entries, the scalar-valued distribution usually has no nice structure. However, often the distribution with respect to an intermediate algebra $\mathcal{B}$ has a nice structure, and thus it makes sense to split the problem into two parts. First, consider the distribution with respect to the intermediate algebra $\mathcal{B}$. Derive all (operator-valued) formulas on this level. Then at the very end, go down to $\mathbb{C}$. This last step usually has to be done numerically. Since our relevant equations (like (9.5)) are not linear, they are not preserved under the application of the mapping $\mathcal{B} \to \mathbb{C}$, meaning that we do not find closed equations on the scalar-valued level. Thus, the first step is nice and gives us some conceptual understanding of the problem, whereas the second step does not give much theoretical insight, but is more of a numerical nature. Clearly, the bigger the last step, i.e. the larger $\mathcal{B}$, the less we win with working on the $\mathcal{B}$-level first. So it is interesting to understand how symmetries of the problem allow us to restrict from $\mathcal{B}$ to some smaller subalgebra $\mathcal{D} \subset \mathcal{B}$. In general, the behaviour of an element as a $\mathcal{B}$-valued random variable might be very different from its behaviour as a $\mathcal{D}$-valued random variable. This is reflected in the fact that in general the expression of the $\mathcal{D}$-valued cumulants of a random variable in terms of its $\mathcal{B}$-valued cumulants is quite complicated. So we can only expect that nice properties with respect to $\mathcal{B}$ pass over to $\mathcal{D}$ if the relation between the corresponding cumulants is easy. The simplest such situation is where the $\mathcal{D}$-valued cumulants are the restriction of the $\mathcal{B}$-valued cumulants. It turns out that it is actually quite easy to decide whether this is the case.

Proposition 16. *Consider unital algebras $\mathbb{C} \subset \mathcal{D} \subset \mathcal{B} \subset \mathcal{A}$ and conditional expectations $E_{\mathcal{B}} : \mathcal{A} \to \mathcal{B}$ and $E_{\mathcal{D}} : \mathcal{A} \to \mathcal{D}$ which are compatible in the sense that $E_{\mathcal{D}} \circ E_{\mathcal{B}} = E_{\mathcal{D}}$. Denote the free cumulants with respect to $E_{\mathcal{B}}$ by $\kappa^{\mathcal{B}}$ and the free cumulants with respect to $E_{\mathcal{D}}$ by $\kappa^{\mathcal{D}}$. Consider now $x \in \mathcal{A}$. Assume that the $\mathcal{B}$-valued cumulants of x satisfy*

$$\kappa_n^{\mathcal{B}}(xd_1, xd_2, \dots, xd_{n-1}, x) \in \mathcal{D} \qquad \forall n \geq 1, \quad \forall d_1, \dots, d_{n-1} \in \mathcal{D}.$$

Then the $\mathcal{D}$-valued cumulants of x are given by the restrictions of the $\mathcal{B}$-valued cumulants: for all $n \geq 1$ and all $d_1, \dots, d_{n-1} \in \mathcal{D}$, we have

$$\kappa_n^{\mathcal{D}}(xd_1, xd_2, \dots, xd_{n-1}, x) = \kappa_n^{\mathcal{B}}(xd_1, xd_2, \dots, xd_{n-1}, x).$$

This statement is from [139]. Its proof is quite straightforward by comparing the corresponding moment-cumulant formulas. We leave it to the reader.

Exercise 2. Prove Proposition 16.

Proposition 16 allows us in particular to check whether a $\mathcal{B}$-valued semi-circular element x is also semi-circular with respect to a smaller $\mathcal{D} \subset \mathcal{B}$. Namely, all $\mathcal{B}$-valued cumulants of x are given by nested iterations of the mapping η. Hence, if η maps $\mathcal{D}$ to $\mathcal{D}$, then this property extends to all $\mathcal{B}$-valued cumulants of x restricted to $\mathcal{D}$.

Corollary 17. *Let $\mathcal{D} \subset \mathcal{B} \subset \mathcal{A}$ be as above. Consider a $\mathcal{B}$-valued semi-circular element x. Let $\eta : \mathcal{B} \to \mathcal{B}$, $\eta(b) = E_{\mathcal{B}}(xbx)$ be the corresponding covariance mapping. If $\eta(\mathcal{D}) \subset \mathcal{D}$, then x is also a $\mathcal{D}$-valued semi-circular element, with covariance mapping given by the restriction of η to $\mathcal{D}$.*

Remark 18. 1) This corollary allows for an easy determination of the smallest canonical subalgebra with respect to which x is still semi-circular. Namely, if x is $\mathcal{B}$-semi-circular with covariance mapping $\eta : \mathcal{B} \to \mathcal{B}$, we let $\mathcal{D}$ be the smallest unital subalgebra of $\mathcal{B}$ which is mapped under η into itself. Note that this $\mathcal{D}$ exists because the intersection of two subalgebras which are invariant under η is again a subalgebra invariant under η. Then x is also semi-circular with respect to this $\mathcal{D}$. Note that the corollary above is not an equivalence, and thus there might be smaller subalgebras than $\mathcal{D}$ with respect to which x is still semi-circular; however, there is no systematic way to detect those.

2) Note also that with some added hypotheses, the above corollary might become an equivalence; for example, in [139] it was shown: Let $(\mathcal{A}, E, \mathcal{B})$ be an operator-valued probability space, such that $\mathcal{A}$ and $\mathcal{B}$ are C^*-algebras. Let $F : \mathcal{B} \to \mathbb{C} =: \mathcal{D} \subset \mathcal{B}$ be a faithful state. Assume that $\tau = F \circ E$ is a faithful trace on $\mathcal{A}$. Let x be a $\mathcal{B}$-valued semi-circular variable in $\mathcal{A}$. Then the distribution of x with respect to τ is the semi-circle law if and only if $E(x^2) \in \mathbb{C}$.

Example 19. Let us see what the statements above tell us about our model case of $d \times d$ self-adjoint matrices with semi-circular entries $X = (s_{ij})_{i,j=1}^d$. In Section 9.1 we have seen that if we allow arbitrary correlations between the entries, then we get a semi-circular distribution with respect to $\mathcal{B} = M_d(\mathbb{C})$. (We calculated this explicitly, but one could also invoke Proposition 13 to get a direct proof of this.) The mapping $\eta : M_d(\mathbb{C}) \to M_d(\mathbb{C})$ was given by

$$[\eta(B)]_{ij} = \sum_{k,l=1}^{d} \sigma(i,k;l,j) b_{kl}.$$

Let us first check in which situations we can expect a scalar-valued semi-circular distribution. This is guaranteed, by the corollary above, if η maps $\mathbb{C}$ to itself, i.e. if $\eta(1)$ is a multiple of the identity matrix. We have

$$[\eta(1)]_{ij} = \sum_{k=1}^{d} \sigma(i,k;k,j).$$

Thus, if $\sum_{k=1}^{d} \sigma(i,k;k,j)$ is zero for $i \neq j$ and otherwise independent from i, then X is semi-circular. The simplest situation where this happens is if all s_{ij}, $1 \leq i \leq j \leq d$, are free and have the same variance.

Let us now consider the more special band matrix situation where s_{ij}, $1 \leq i \leq j \leq d$ are free, but not necessarily of the same variance, i.e. we assume that for $i \leq j, k \leq l$, we have

$$\sigma(i,j;k,l) = \begin{cases} \sigma_{ij}, & \text{if } i = k, j = l \\ 0, & \text{otherwise} \end{cases}. \tag{9.24}$$

Note that this also means that $\sigma(i,k;k,i) = \sigma_{ik}$, because we have $s_{ki} = s_{ik}$. Then

$$[\eta(1)]_{ij} = \delta_{ij} \sum_{k=1}^{d} \sigma_{ik}.$$

We see that in order to get a semi-circular distribution, we do not need the same variance everywhere, but that it suffices to have the same sum over the variances in each row of the matrix.

However, if this sum condition is not satisfied, then we do not have a semi-circular distribution. Still, having all entries free gives more structure than just semi-circularity with respect to $M_d(\mathbb{C})$. Namely, we see that with the covariance (9.24), our η maps diagonal matrices into diagonal matrices. Thus, we can pass from $M_d(\mathbb{C})$ over to the subalgebra $\mathcal{D} \subset M_d(\mathbb{C})$ of diagonal matrices and get that for such situations X is $\mathcal{D}$-semi-circular. The conditional expectation $E_{\mathcal{D}} : \mathcal{A} \to \mathcal{D}$ in this case is of course given by

$$\begin{pmatrix} a_{11} & \dots & a_{1d} \\ \vdots & \ddots & \vdots \\ a_{d1} & \dots & a_{dd} \end{pmatrix} \mapsto \begin{pmatrix} \varphi(a_{11}) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \varphi(a_{dd}) \end{pmatrix}.$$

Even if we do not have free entries, we might still have some symmetries in the correlations between the entries which let us pass to some subalgebra of $M_d(\mathbb{C})$. As pointed out in Remark 18, we should look for the smallest subalgebra which is invariant under η. This was exactly what we did implicitly in our Example 3. There we observed that η maps the subalgebra

$$\mathcal{C} := \left\{ \begin{pmatrix} f & 0 & h \\ 0 & e & 0 \\ h & 0 & f \end{pmatrix} \mid e, f, h \in \mathbb{C} \right\}$$

into itself. (And we actually saw in Example 3 that $\mathcal{C}$ is the smallest such subalgebra, because it is generated from the unit by iterated application of η.) Thus, the X from this example, (9.2), is not only $M_3(\mathbb{C})$-semi-circular but actually also $\mathcal{C}$-semi-circular. In our calculations in Example 3, this was implicitly taken into account, because there we restricted our Cauchy transform G to values in $\mathcal{C}$, i.e. effectively we solved the equation (9.5) for an operator-valued semi-circular element not in $M_3(\mathbb{C})$, but in $\mathcal{C}$.

9.5 A non-self-adjoint example

In order to treat a more complicated example, let us look at a non-self-adjoint situation as it often shows up in applications (e.g. in wireless communication; see [174]). Consider the $d \times d$ matrix $H = B + C$ where $B \in M_d(\mathbb{C})$ is a deterministic matrix and $C = (c_{ij})_{i,j=1}^{d}$ has as entries $*$-free circular elements

c_{ij} $(i, j = 1, \ldots, d)$, without any symmetry conditions, however with varying variance, i.e. $\varphi(c_{ij} c_{ij}^*) = \sigma_{ij}$. What we want to calculate is the distribution of HH^*.

Such an H might arise as the limit of block matrices in Gaussian random matrices, where we also allow a non-zero mean for the Gaussian entries. The means are separated off in the matrix B. We refer to [174] for more information on the use of such non-mean zero Gaussian random matrices (as Ricean model) and why one is interested in the eigenvalue distribution of HH^*.

One can reduce this to a problem involving self-adjoint matrices by observing that HH^* has the same distribution as the square of

$$T := \begin{pmatrix} 0 & H \\ H^* & 0 \end{pmatrix} = \begin{pmatrix} 0 & B \\ B^* & 0 \end{pmatrix} + \begin{pmatrix} 0 & C \\ C^* & 0 \end{pmatrix}.$$

Let us use the notations

$$\hat{B} := \begin{pmatrix} 0 & B \\ B^* & 0 \end{pmatrix} \qquad \text{and} \qquad \hat{C} := \begin{pmatrix} 0 & C \\ C^* & 0 \end{pmatrix}.$$

The matrix $\hat{C}$ is a $2d \times 2d$ self-adjoint matrix with *-free circular entries, thus of the type we considered in Section 9.1. Hence, by the remarks in Example 19, we know that it is a $\mathcal{D}_{2d}$-valued semi-circular element, where $\mathcal{D}_{2d} \subset M_{2d}(\mathbb{C})$ is the subalgebra of diagonal matrices; one checks easily that the covariance function $\eta : \mathcal{D}_{2d} \to \mathcal{D}_{2d}$ is given by

$$\eta \begin{pmatrix} D_1 & 0 \\ 0 & D_2 \end{pmatrix} = \begin{pmatrix} \eta_1(D_2) & 0 \\ 0 & \eta_2(D_1) \end{pmatrix}, \tag{9.25}$$

where $\eta_1 : \mathcal{D}_d \to \mathcal{D}_d$ and $\eta_2 : \mathcal{D}_d \to \mathcal{D}_d$ are given by

$$\eta_1(D_2) = id \otimes \varphi[CD_2C^*]$$
$$\eta_2(D_1) = id \otimes \varphi[C^*D_1C].$$

Furthermore, by using Propositions 13 and 16, one can easily see that $\hat{B}$ and $\hat{C}$ are free over $\mathcal{D}_{2d}$.

Let G_T and G_{T^2} be the $\mathcal{D}_{2d}$-valued Cauchy transform of T and T^2, respectively. We write the latter as

$$G_{T^2}(z) = \begin{pmatrix} G_1(z) & 0 \\ 0 & G_2(z) \end{pmatrix},$$

where G_1 and G_2 are $\mathcal{D}_d$-valued. Note that one also has the general relation $G_T(z) = zG_{T^2}(z^2)$.

By using the general subordination relation (9.21) and the fact that $\hat{C}$ is semi-circular with covariance map η given by (9.25), we can now derive the following equation for G_{T^2}:

$$\begin{aligned} zG_{T^2}(z^2) = G_T(z) &= G_{\hat{B}}\left[z - R_{\hat{C}}(G_T(z))\right] \\ &= E_{\mathcal{D}_{2d}}\left[\left(z - z\eta\begin{pmatrix} G_1(z^2) & 0 \\ 0 & G_2(z^2)\end{pmatrix} - \begin{pmatrix} 0 & B \\ B^* & 0\end{pmatrix}\right)^{-1}\right] \\ &= E_{\mathcal{D}_{2d}}\left[\begin{pmatrix} z - z\eta_1(G_2(z^2)) & -B \\ -B^* & z - z\eta_2(G_1(z^2))\end{pmatrix}^{-1}\right]. \end{aligned}$$

By using the well-known Schur complement formula for the inverse of 2×2 block matrices (see also next chapter for more on this), this yields finally

$$zG_1(z) = E_{\mathcal{D}_d}\left[\left(1 - \eta_1(G_2(z)) + B\frac{1}{z - z\eta_2(G_1(z))}B^*\right)^{-1}\right]$$

and

$$zG_2(z) = E_{\mathcal{D}_d}\left[\left(1 - \eta_2(G_1(z)) + B^*\frac{1}{z - z\eta_1(G_2(z))}B\right)^{-1}\right].$$

These equations have actually been derived in [90] as the fixed point equations for a so-called *deterministic equivalent* of the square of a random matrix with noncentred, independent Gaussians with non-constant variance as entries. Thus, our calculations show that going over to such a deterministic equivalent consists in replacing the original random matrix by our matrix T. We will come back to this notion of "deterministic equivalent" in the next chapter.

Chapter 10
Deterministic Equivalents, Polynomials in Free Variables, and Analytic Theory of Operator-Valued Convolution

The notion of a "deterministic equivalent" for random matrices, which can be found in the engineering literature, is a non-rigorous concept which amounts to replacing a random matrix model of finite size (which is usually unsolvable) by another problem which is solvable, in such a way that, for large N, the distributions of both problems are close to each other. Motivated by our example in the last chapter, we will in this chapter propose a rigorous definition for this concept, which relies on asymptotic freeness results. This "free deterministic equivalent" was introduced by Speicher and Vargas in [166].

This will then lead directly to the problem of calculating the distribution of self-adjoint polynomials in free variables. We will see that, in contrast to the corresponding classical problem on the distribution of polynomials in independent random variables, there exists a general algorithm to deal with such polynomials in free variables. The main idea will be to relate such a polynomial with an operator-valued *linear* polynomial and then use operator-valued convolution to deal with the latter. The successful implementation of this program is due to Belinschi, Mai, and Speicher [23]; see also [12].

10.1 The general concept of a free deterministic equivalent

Voiculescu's asymptotic freeness results on random matrices state that if we consider tuples of independent random matrix ensembles, such as Gaussian, Wigner, or Haar unitaries, their collective behaviour in the large N limit is almost surely that of a corresponding collection of free (semi-)circular and Haar unitary operators. Moreover, if we consider these random ensembles along with deterministic ensembles, having a given asymptotic distribution (with respect to the normalized trace), then, almost surely, the corresponding limiting operators also become free from the random elements. This means of course that if we consider a function in our matrices, then this will, for large N, be approximated by the same function in our limiting operators. We will in the following only consider functions which are

J.A. Mingo, R. Speicher, *Free Probability and Random Matrices*,
Fields Institute Monographs 35, DOI 10.1007/978-1-4939-6942-5_10

given by polynomials. Furthermore, all our polynomials should be self-adjoint (in the sense that if we plug in self-adjoint matrices, we will get as output self-adjoint matrices), so that the eigenvalue distribution of those polynomials can be recovered by calculating traces of powers.

To be more specific, let us consider a collection of independent random and deterministic $N \times N$ matrices:

$$\mathbf{X}_N = \left\{X_1^{(N)}, \ldots, X_{i_1}^{(N)}\right\} : \text{independent self-adjoint Gaussian matrices,}$$
$$\mathbf{Y}_N = \left\{Y_1^{(N)}, \ldots, Y_{i_2}^{(N)}\right\} : \text{independent non-self-adjoint Gaussian matrices,}$$
$$\mathbf{U}_N = \left\{U_1^{(N)}, \ldots, U_{i_3}^{(N)}\right\} : \text{independent Haar distributed unitary matrices,}$$
$$\mathbf{D}_N = \left\{D_1^{(N)}, \ldots, D_{i_4}^{(N)}\right\} : \text{deterministic matrices,}$$

and a self-adjoint polynomial P in non-commuting variables (and their adjoints); we evaluate this polynomial in our matrices

$$P\big(X_1^{(N)}, \ldots, X_{i_1}^{(N)}, Y_1^{(N)}, \ldots, Y_{i_2}^{(N)}, U_1^{(N)}, \ldots, U_{i_3}^{(N)}, D_1^{(N)}, \ldots, D_{i_4}^{(N)}\big) =: P_N.$$

Relying on asymptotic freeness results, we can then compute the asymptotic eigenvalue distribution of P_N by going over the limit. We know that we can find collections $\mathbf{S}, \mathbf{C}, \mathbf{U}, \mathbf{D}$ of operators in a non-commutative probability space $(\mathcal{A}, \varphi)$,

$$\mathbf{S} = \{s_1, \ldots, s_{i_1}\} : \text{free semi-circular elements,}$$
$$\mathbf{C} = \{c_1, \ldots, c_{i_2}\} : *\text{-free circular elements,}$$
$$\mathbf{U} = \{u_1, \ldots, u_{i_3}\} : *\text{-free Haar unitaries,}$$
$$\mathbf{D} = \{d_1, \ldots, d_{i_4}\} : \text{abstract elements,}$$

such that $\mathbf{S}, \mathbf{C}, \mathbf{U}, \mathbf{D}$ are $*$-free and the joint distribution of $d_1, \ldots, d_{i_4}$ is given by the asymptotic joint distribution of $D_1^{(N)}, \ldots, D_{i_4}^{(N)}$. Then, almost surely, the asymptotic distribution of P_N is that of $P\left(s_1, \ldots, s_{i_1}, c_1, \ldots, c_{i_2}, u_1, \ldots, u_{i_3}, d_1, \ldots, d_{i_4}\right) =: p_\infty$, in the sense that, for all k, we have almost surely

$$\lim_{N\to\infty} \operatorname{tr}(P_N^k) = \varphi(p_\infty^k).$$

In this way, we can reduce the problem of the asymptotic distribution of P_N to the study of the distribution of p_∞.

A common obstacle of this procedure is that our deterministic matrices may not have an asymptotic joint distribution. It is then natural to consider, for a fixed N, the corresponding "free model" $P\big(s_1,\ldots,s_{i_1},c_1,\ldots,c_{i_2},u_1,\ldots,u_{i_3},d_1^{(N)},\ldots,d_{i_4}^{(N)}\big)$ $=: p_N^\square$, where, just as before, the random matrices are replaced by the corresponding free operators in some space $(\mathcal{A}_N,\varphi_N)$, but now we let the distribution of $d_1^{(N)},\ldots,d_{i_4}^{(N)}$ be exactly the same as the one of $D_1^{(N)},\ldots,D_{i_4}^{(N)}$ with respect to tr. The free model $p_N^\square$ will be called the *free deterministic equivalent* for P_N. This was introduced and investigated in [166, 175].

(In case one wonders about the notation, $p_N^\square$: the symbol $\square$ is according to [30] the generic qualifier for denoting the free version of some classical object or operation.)

The difference between the distribution of $p_N^\square$ and the (almost sure or expected) distribution of P_N is given by the deviation from freeness of $\mathbf{X}_N,\mathbf{Y}_N,\mathbf{U}_N,\mathbf{D}_N$, the deviation of $\mathbf{X}_N,\mathbf{Y}_N$ from being free (semi)-circular systems, and the deviation of $\mathbf{U}_N$ from a free system of Haar unitaries. Of course, for large N these deviations get smaller, and thus the distribution of $p_N^\square$ becomes a better approximation for the distribution of P_N.

Let us denote by G_N the Cauchy transform of P_N and by $G_N^\square$ the Cauchy transform of the free deterministic equivalent $p_N^\square$. Then, the usual asymptotic freeness estimates show that moments of P_N are, for large N, with very high probability close to corresponding moments of $p_N^\square$ (where the estimates involve also the operator norms of the deterministic matrices). This means that for $N\to\infty$, the difference between the Cauchy transforms G_N and $G_N^\square$ goes almost surely to zero, even if there do not exist individual limits for both Cauchy transforms.

In the engineering literature, there exists also a version of the notion of a deterministic equivalent (apparently going back to Girko [78], see also [90]). This deterministic equivalent consists in replacing the Cauchy transform G_N of the considered random matrix model (for which no analytic solution exists) by a function $\hat{G}_N$ which is defined as the solution of a specified system of equations. The specific form of those equations is determined in an ad hoc way, depending on the considered problem, by making approximations for the equations of G_N, such that one gets a closed system of equations. In many examples of deterministic equivalents (e.g. see [62, Chapter 6]), it turns out that actually the Cauchy transform of our free deterministic equivalent is the solution to those modified equations, i.e. that $\hat{G}_N = G_N^\square$. We saw one concrete example of this in Section 9.5 of the last chapter.

Our definition of a deterministic equivalent gives a more conceptual approach and shows clearly how this notion relates with free probability theory. In some sense, this indicates that the only meaningful way to get a closed system of equations when dealing with random matrices is to replace the random matrices by free variables.

Deterministic equivalents are thus polynomials in free variables, and it remains to develop tools to deal with such polynomials in an effective way. It turns out that operator-valued free probability theory provides such tools. We will elaborate on this in the remaining sections of this chapter.

10.2 A motivating example: reduction to multiplicative convolution

In the following, we want to see how problems about polynomials in free variables can be treated by means of operator-valued free probability. The main idea in this context is that complicated polynomials can be transformed into simpler ones by going to matrices (and thus go from scalar-valued to operator-valued free probability). Since the only polynomials which we can effectively deal with are sums and products (corresponding to additive and multiplicative convolution, respectively), we should aim to transform general polynomials into sums or products.

In this section, we will treat one special example from [25] to get an idea how this can be achieved. In this case, we will transform our problem into a product of two free operator-valued matrices.

Let a_1, a_2, b_1, b_2 be self-adjoint random variables in a non-commutative probability space $(\mathcal{C}, \varphi)$, such that $\{a_1, a_2\}$ and $\{b_1, b_2\}$ are free and consider the polynomial $p = a_1 b_1 a_1 + a_2 b_2 a_2$. This p is self-adjoint and its distribution, i.e. the collection of its moments, is determined by the joint distribution of $\{a_1, a_2\}$, the joint distribution of $\{b_1, b_2\}$, and the freeness between $\{a_1, a_2\}$ and $\{b_1, b_2\}$. However, there is no direct way of calculating this distribution.

We observe now that the distribution μ_p of p is the same (modulo a Dirac mass at zero) as the distribution of the element

$$\begin{pmatrix} a_1 b_1 a_1 + a_2 b_2 a_2 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} a_1 & a_2 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} b_1 & 0 \\ 0 & b_2 \end{pmatrix} \begin{pmatrix} a_1 & 0 \\ a_2 & 0 \end{pmatrix}, \tag{10.1}$$

in the non-commutative probability space $(M_2(\mathcal{C}), \mathrm{tr}_2 \otimes \varphi)$. But this element has the same moments as

$$\begin{pmatrix} a_1 & 0 \\ a_2 & 0 \end{pmatrix} \begin{pmatrix} a_1 & a_2 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} b_1 & 0 \\ 0 & b_2 \end{pmatrix} = \begin{pmatrix} a_1^2 & a_1 a_2 \\ a_2 a_1 & a_2^2 \end{pmatrix} \begin{pmatrix} b_1 & 0 \\ 0 & b_2 \end{pmatrix} =: AB. \tag{10.2}$$

So, with μ_{AB} denoting the distribution of AB with respect to $\mathrm{tr}_2 \otimes \varphi$, we have

$$\mu_{AB} = \frac{1}{2}\mu_p + \frac{1}{2}\delta_0.$$

Since A and B are not free with respect to $\mathrm{tr}_2 \otimes \varphi$, we cannot use scalar-valued multiplicative free convolution to calculate the distribution of AB. However, with $E : M_2(\mathcal{C}) \to M_2(\mathbb{C})$ denoting the conditional expectation onto deterministic 2×2 matrices, we have that the scalar-valued distribution μ_{AB} is given by taking the trace tr_2 of the operator-valued distribution of AB with respect to E. But on this operator-valued level, the matrices A and B are, by Corollary 9.14, free with amalgamation over $M_2(\mathbb{C})$. Furthermore, the $M_2(\mathbb{C})$-valued distribution of A is determined by the joint distribution of a_1 and a_2, and the $M_2(\mathbb{C})$-valued distribution of B is determined by the joint distribution of b_1 and b_2. Hence, the scalar-valued distribution μ_p will be given by first calculating the $M_2(\mathbb{C})$-valued free multiplicative convolution of A

and B to obtain the $M_2(\mathbb{C})$-valued distribution of AB and then getting from this the (scalar-valued) distribution μ_{AB} by taking the trace over $M_2(\mathbb{C})$. Thus, we have rewritten our original problem as a problem on the product of two free operator-valued variables.

10.3 The general case: reduction to operator-valued additive convolution via the linearization trick

Let us now be more ambitious and look at an arbitrary self-adjoint polynomial $P \in \mathbb{C}\langle X_1, \dots, X_n\rangle$, evaluated as $p = P(x_1, \dots, x_n) \in \mathcal{A}$ in free variables $x_1, \dots, x_n \in \mathcal{A}$. In the last section, we replaced our original variable by a matrix which has (up to some atoms), with respect to $\mathrm{tr} \otimes \varphi$, the same distribution and which is actually a product of matrices in the single operators. It is quite unlikely that we can do the same in general. However, if we do not insist on using the trace as our state on matrices but allow, for example, the evaluation at the $(1, 1)$ entry, then we gain much flexibility and can indeed find an equivalent matrix which splits even into a sum of matrices of the individual variables. What we essentially need for this is, given the polynomial P, to construct in a systematic way a matrix, such that the entries of this matrix are polynomials of degree 0 or 1 in our variables and such that the inverse of this matrix has as $(1, 1)$ entry $(z - P)^{-1}$. Let us ignore for the moment the degree condition on the entries and just concentrate on the invertibility questions. The relevant tool in this context is the following well-known result about Schur complements.

Proposition 1. *Let $\mathcal{A}$ be a complex and unital algebra and let elements $a, b, c, d \in \mathcal{A}$ be given. We assume that d is invertible in $\mathcal{A}$. Then the following statements are equivalent:*

(i) The matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is invertible in $M_2(\mathbb{C}) \otimes \mathcal{A}$.

(ii) The Schur complement *$a - bd^{-1}c$ is invertible in $\mathcal{A}$.*

If the equivalent conditions (i) and (ii) are satisfied, we have the relation

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 \\ -d^{-1}c & 1 \end{pmatrix} \begin{pmatrix} (a - bd^{-1}c)^{-1} & 0 \\ 0 & d^{-1} \end{pmatrix} \begin{pmatrix} 1 & -bd^{-1} \\ 0 & 1 \end{pmatrix}. \tag{10.3}$$

In particular, the $(1, 1)$ entry of the inverse is given by $(a - bd^{-1}c)^{-1}$:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \begin{pmatrix} (a - bd^{-1}c)^{-1} & * \\ * & * \end{pmatrix}.$$

Proof: A direct calculation shows that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & bd^{-1} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a - bd^{-1}c & 0 \\ 0 & d \end{pmatrix} \begin{pmatrix} 1 & 0 \\ d^{-1}c & 1 \end{pmatrix} \tag{10.4}$$

holds. Since the first and third matrix are both invertible in $M_2(\mathbb{C}) \otimes \mathcal{A}$,

$$\begin{pmatrix} 1 & bd^{-1} \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & -bd^{-1} \\ 0 & 1 \end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix} 1 & 0 \\ d^{-1}c & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 \\ -d^{-1}c & 1 \end{pmatrix},$$

the stated equivalence of (*i*) and (*ii*), as well as formula (10.3), follows from (10.4). □

What we now need, given our operator $p = P(x_1, \dots, x_n)$, is to find a block matrix such that the $(1,1)$ entry of the inverse of this block matrix corresponds to the resolvent $(z-p)^{-1}$ and that furthermore all the entries of this block matrix have at most degree 1 in our variables. More precisely, we are looking for an operator

$$\hat{p} = b_0 \otimes 1 + b_1 \otimes x_1 + \cdots + b_n \otimes x_n \in M_N(\mathbb{C}) \otimes \mathcal{A}$$

for some matrices $b_0, \dots, b_n \in M_N(\mathbb{C})$ of dimension N, such that $z-p$ is invertible in $\mathcal{A}$ if and only if $\Lambda(z) - \hat{p}$ is invertible in $M_N(\mathbb{C}) \otimes \mathcal{A}$. Hereby, we put

$$\Lambda(z) = \begin{pmatrix} z & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix} \qquad \text{for all } z \in \mathbb{C}. \tag{10.5}$$

As we will see in the following, the linearization in terms of the dimension $N \in \mathbb{N}$ and the matrices $b_0, \dots, b_n \in M_N(\mathbb{C})$ usually depends only on the given polynomial $P \in \mathbb{C}\langle X_1, \dots, X_n\rangle$ and not on the special choice of elements $x_1, \dots, x_n \in \mathcal{A}$.

The first famous linearization trick in the context of operator algebras and random matrices goes back to Haagerup and Thorbjørnsen [88, 89] and turned out to be a powerful tool in many different respects. However, there was the disadvantage that, even if we start from a self-adjoint polynomial P, in general, we will not end up with a linearization $\hat{p}$, which is self-adjoint as well. Then, in [5], Anderson presented a new version of this linearization procedure, which preserved self-adjointness.

One should note, however, that the idea of linearizing polynomial (or actually rational, see Section 10.6)) problems by going to matrices is actually much older and is known under different names in different communities like "Higman's trick" [98] or "linearization by enlargement" in non-commutative ring theory [56], "recognizable power series" in automata theory and formal languages [154], or "descriptor realization" in control theory [93]. For a survey on linearization, non-commutative system realization, and its use in free probability, see [95].

Here is now our precise definition of linearization.

Definition 2. Let $P \in \mathbb{C}\langle X_1, \dots, X_n\rangle$ be given. A matrix

$$\hat{P} := \begin{pmatrix} 0 & U \\ V & Q \end{pmatrix} \in M_N(\mathbb{C}) \otimes \mathbb{C}\langle X_1, \dots, X_n\rangle,$$

where

- $N \in \mathbb{N}$ is an integer,
- $Q \in M_{N-1}(\mathbb{C}) \otimes \mathbb{C}\langle X_1, \dots, X_n\rangle$ is invertible
- and U is a row vector and V is a column vector, both of size $N-1$ with entries in $\mathbb{C}\langle X_1, \dots, X_n\rangle$,

is called a *linearization of* P, if the following conditions are satisfied:

(i) There are matrices $b_0, \dots, b_n \in M_N(\mathbb{C})$, such that

$$\hat{P} = b_0 \otimes 1 + b_1 \otimes X_1 + \cdots + b_n \otimes X_n,$$

i.e. the polynomial entries in Q, U, and V all have degree ≤ 1.

(ii) It holds true that $P = -UQ^{-1}V$.

Applying the Schur complement, Proposition 1, to this situation yields then the following:

Corollary 3. *Let $\mathcal{A}$ be a unital algebra and let elements $x_1, \dots, x_n \in \mathcal{A}$ be given. Assume $P \in \mathbb{C}\langle X_1, \dots, X_n\rangle$ has a linearization*

$$\hat{P} = b_0 \otimes 1 + b_1 \otimes X_1 + \cdots + b_n \otimes X_n \in M_N(\mathbb{C}) \otimes \mathbb{C}\langle X_1, \dots, X_n\rangle$$

with matrices $b_0, \dots, b_n \in M_N(\mathbb{C})$. Then the following conditions are equivalent for any complex number $z \in \mathbb{C}$:

(i) The operator $z - p$ with $p := P(x_1, \dots, x_n)$ is invertible in $\mathcal{A}$.

(ii) The operator $\Lambda(z) - \hat{p}$ with $\Lambda(z)$ defined as in (10.5) and

$$\hat{p} := b_0 \otimes 1 + b_1 \otimes x_1 + \cdots + b_n \otimes x_n \in M_N(\mathbb{C}) \otimes \mathcal{A}$$

is invertible in $M_N(\mathbb{C}) \otimes \mathcal{A}$.

Moreover, if (i) and (ii) are fulfilled for some $z \in \mathbb{C}$, we have that

$$\left[(\Lambda(z) - \hat{p})^{-1}\right]_{1,1} = (z-p)^{-1}.$$

Proof: By the definition of a linearization, Definition 2, we have a block decomposition of the form

$$\hat{p} := \begin{pmatrix} 0 & u \\ v & q \end{pmatrix} \in M_N(\mathbb{C}) \otimes \mathcal{A}$$

where $u = U(x_1, \dots, x_n)$, $v = V(x_1, \dots, x_n)$ and $q = Q(x_1, \dots, x_n)$. Furthermore, we know that $q \in M_{N-1}(\mathbb{C}) \otimes \mathcal{A}$ is invertible and $p = -uq^{-1}v$ holds. This implies

$$\Lambda(z) - \hat{p} = \begin{pmatrix} z & -u \\ -v & -q \end{pmatrix},$$

and the statements follow from Proposition 1. □

Now, it only remains to ensure the existence of linearizations of this kind.

Proposition 4. *Any polynomial $P \in \mathbb{C}\langle X_1, \dots, X_n\rangle$ admits a linearization $\hat{P}$ in the sense of Definition 2. If P is self-adjoint, then the linearization can be chosen to be self-adjoint.*

The proof follows by combining the following simple observations:

Exercise 1.

(i) Show that $X_j \in \mathbb{C}\langle X_1, \dots, X_n\rangle$ has a linearization

$$\hat{X}_j = \begin{pmatrix} 0 & X_j \\ 1 & -1 \end{pmatrix} \in M_2(\mathbb{C}) \otimes \mathbb{C}\langle X_1, \dots, X_n\rangle.$$

(This statement looks simplistic taken for itself, but it will be useful when combined with the third part.)

(ii) A monomial of the form $P := X_{i_1} X_{i_2} \cdots X_{i_k} \in \mathbb{C}\langle X_1, \dots, X_n\rangle$ for $k \geq 2$, $i_1, \dots, i_k \in \{1, \dots, n\}$ has a linearization

$$\hat{P} = \begin{pmatrix} & & & X_{i_1} \\ & & X_{i_2} & -1 \\ & \iddots & \iddots & \\ X_{i_k} & -1 & & \end{pmatrix} \in M_k(\mathbb{C}) \otimes \mathbb{C}\langle X_1, \dots, X_n\rangle.$$

(iii) If the polynomials $P_1, \dots, P_k \in \mathbb{C}\langle X_1, \dots, X_n\rangle$ have linearizations

$$\hat{P}_j = \begin{pmatrix} 0 & U_j \\ V_j & Q_j \end{pmatrix} \in M_{N_j}(\mathbb{C}) \otimes \mathbb{C}\langle X_1, \dots, X_n\rangle$$

for $j = 1, \dots, n$, then their sum $P := P_1 + \cdots + P_k$ has the linearization

$$\hat{P} = \begin{pmatrix} 0 & U_1 & \dots & U_k \\ V_1 & Q_1 & & \\ \vdots & & \ddots & \\ V_k & & & Q_k \end{pmatrix} \in M_N(\mathbb{C}) \otimes \mathbb{C}\langle X_1, \dots, X_n\rangle$$

with $N := (N_1 + \cdots + N_k) - k + 1$.

(iv) If

$$\begin{pmatrix} 0 & U \\ V & Q \end{pmatrix} \in M_N(\mathbb{C}) \otimes \mathbb{C}\langle X_1, \dots, X_n\rangle$$

is a linearization of P, then

$$\begin{pmatrix} 0 & U & V^* \\ U^* & 0 & Q^* \\ V & Q & 0 \end{pmatrix} \in M_{2N-1}(\mathbb{C}) \otimes \mathbb{C}\langle X_1, \ldots, X_n \rangle$$

is a linearization of $P + P^*$.

10.4 Analytic theory of operator-valued convolutions

In the last two sections, we indicated how problems in free variables can be transformed into operator-valued simpler problems. In particular, the distribution of a self-adjoint polynomial $p = P(x_1, \ldots, x_n)$ in free variables $x_1, \ldots, x_n$ can be deduced from the operator-valued distribution of a corresponding linearization

$$\hat{p} := b_0 \otimes 1 + b_1 \otimes x_1 + \cdots + b_n \otimes x_n \in M_N(\mathbb{C}) \otimes \mathcal{A}.$$

Note that for this linearization, the freeness of the variables plays no role. Where it becomes crucial is the observation that the freeness of $x_1, \ldots, x_n$ implies, by Corollary 9.14, the freeness over $M_N(\mathbb{C})$ of $b_1 \otimes x_1, \ldots, b_n \otimes x_n$. (Note that there is no classical counter part of this for the case of independent variables.) Hence, the distribution of $\hat{p}$ is given by the operator-valued free additive convolution of the distributions of $b_1 \otimes x_1, \ldots, b_n \otimes x_n$. Furthermore, since the distribution of x_i determines also the $M_N(\mathbb{C})$-valued distribution of $b_i \otimes x_i$, we have finally reduced the determination of the distribution of $P(x_1, \ldots, x_n)$ to a problem involving operator-valued additive free convolution. As pointed out in Section 9.2, we can in principle deal with such a convolution.

However, in the last chapter we treated the relevant tools, in particular the operator-valued R-transform, only as formal power series, and it is not clear how one should be able to derive explicit solutions from such formal equations. But worse, even if the operator-valued Cauchy and R-transforms are established as analytic objects, it is not clear how to solve operator-valued equations like the one in Theorem 9.11. There are rarely any non-trivial operator-valued examples where an explicit solution can be written down; and also numerical methods for such equations are problematic – a main obstacle being that those equations usually have many solutions, and it is a priori not clear how to isolate the one with the right positivity properties. As we have already noticed in the scalar-valued case, it is the subordination formulation of those convolutions which comes to the rescue. From an analytic and also a numerical point of view, the subordination function is a much nicer object than the R-transform.

So, in order to make good use of our linearization algorithm, we need also a well-developed subordination theory of operator-valued free convolution. Such a theory exists and we will present in the following the relevant statements. For proofs and more details, we refer to the original papers [23, 25].

10.4.1 General notations

A *C^*-operator-valued probability space* $(\mathcal{M}, E, \mathcal{B})$ is an operator-valued probability space, where $\mathcal{M}$ is a C^*-algebra, $\mathcal{B}$ is a C^*-subalgebra of $\mathcal{M}$, and E is completely positive. In such a setting, we use for $x \in \mathcal{M}$ the notation $x > 0$ for the situation where $x \geq 0$ and x is invertible; note that this is equivalent to the fact that there exists a real $\varepsilon > 0$ such that $x \geq \varepsilon 1$. Any element $x \in \mathcal{M}$ can be uniquely written as $x = \mathrm{Re}(x) + i\,\mathrm{Im}(x)$, where $\mathrm{Re}(x) = (x + x^*)/2$ and $\mathrm{Im}(x) = (x - x^*)/(2i)$ are self-adjoint. We call $\mathrm{Re}(x)$ and $\mathrm{Im}(x)$ the real and imaginary part of x.

The appropriate domain for the operator-valued Cauchy transform G_x for a self-adjoint element $x = x^*$ is the *operator upper half-plane*

$$\mathbb{H}^+(\mathcal{B}) := \{b \in \mathcal{B} : \mathrm{Im}(b) > 0\}.$$

Elements in this open set are all invertible, and $\mathbb{H}^+(\mathcal{B})$ is invariant under conjugation by invertible elements in $\mathcal{B}$, i.e. if $b \in \mathbb{H}^+(\mathcal{B})$ and $c \in GL(\mathcal{B})$ is invertible, then $cbc^* \in \mathbb{H}^+(\mathcal{B})$.

We shall use the following analytic mappings, all defined on $\mathbb{H}^+(\mathcal{B})$; all transforms have a natural Schwarz-type analytic extension to the lower half-plane given by $f(b^*) = f(b)^*$; in all formulas below, $x = x^*$ is fixed in $\mathcal{M}$:

- the moment generating function:

$$\Psi_x(b) = E\left[(1 - bx)^{-1} - 1\right] = E\left[(b^{-1} - x)^{-1}\right] b^{-1} - 1 = G_x(b^{-1})b^{-1} - 1; \tag{10.6}$$

- the reciprocal Cauchy transform:

$$F_x(b) = E\left[(b - x)^{-1}\right]^{-1} = G_x(b)^{-1}; \tag{10.7}$$

- the *eta transform*:

$$\eta_x(b) = \Psi_x(b)(1 + \Psi_x(b))^{-1} = 1 - bF_x(b^{-1}); \tag{10.8}$$

- the *h transform*:

$$h_x(b) = E\left[(b - x)^{-1}\right]^{-1} - b = F_x(b) - b. \tag{10.9}$$

10.4.2 Operator-valued additive convolution

Here is now the main theorem from [23] on operator-valued free additive convolution.

Theorem 5. *Assume that $(\mathcal{M}, E, \mathcal{B})$ is a C^*-operator-valued probability space and $x, y \in \mathcal{M}$ are two self-adjoint operator-valued random variables which are* *free over $\mathcal{B}$. Then there exists a unique pair of Fréchet (and thus also Gateaux) analytic maps $\omega_1, \omega_2 \colon \mathbb{H}^+(\mathcal{B}) \to \mathbb{H}^+(\mathcal{B})$ so that*

(*i*) $\operatorname{Im}(\omega_j(b)) \geq \operatorname{Im}(b)$ *for all* $b \in \mathbb{H}^+(\mathcal{B})$*,* $j \in \{1, 2\}$*;*
(*ii*) $F_x(\omega_1(b)) + b = F_y(\omega_2(b)) + b = \omega_1(b) + \omega_2(b)$ *for all* $b \in \mathbb{H}^+(\mathcal{B})$;
(*iii*) $G_x(\omega_1(b)) = G_y(\omega_2(b)) = G_{x+y}(b)$ *for all* $b \in \mathbb{H}^+(\mathcal{B})$.

Moreover, if $b \in \mathbb{H}^+(\mathcal{B})$, then $\omega_1(b)$ is the unique fixed point of the map

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

and

$$\omega_1(b) = \lim_{n\to\infty} f_b^{\circ n}(w) \qquad \text{for any } w \in \mathbb{H}^+(\mathcal{B}),$$

where $f_b^{\circ n}$ denotes the n-fold composition of f_b with itself. Similar statements hold for ω_2, with f_b replaced by $w \mapsto h_x(h_y(w) + b) + b$.

10.4.3 Operator-valued multiplicative convolution

There is also an analogous theorem for treating the operator-valued multiplicative free convolution, see [25].

Theorem 6. *Let $(\mathcal{M}, E, \mathcal{B})$ be a W^*-operator-valued probability space; i.e. $\mathcal{M}$ is a von Neumann algebra and $\mathcal{B}$ a von Neumann subalgebra. Let $x > 0$, $y = y^* \in \mathcal{M}$ be two random variables with invertible expectations, free over $\mathcal{B}$. There exists a Fréchet holomorphic map $\omega_2 \colon \{b \in \mathcal{B} \colon \operatorname{Im}(bx) > 0\} \to \mathbb{H}^+(\mathcal{B})$, such that*

(*i*) $\eta_y(\omega_2(b)) = \eta_{xy}(b)$, $\operatorname{Im}(bx) > 0$*;*
(*ii*) *$\omega_2(b)$ and $b^{-1}\omega_2(b)$ are analytic around zero;*
(*iii*) *for any $b \in \mathcal{B}$ so that $\operatorname{Im}(bx) > 0$, the map $g_b \colon \mathbb{H}^+(\mathcal{B}) \to \mathbb{H}^+(\mathcal{B})$, $g_b(w) = bh_x(h_y(w)b)$ is well defined and analytic, and for any fixed $w \in \mathbb{H}^+(\mathcal{B})$,*

$$\omega_2(b) = \lim_{n\to\infty} g_b^{\circ n}(w),$$

in the weak operator topology.

Moreover, if one defines $\omega_1(b) := h_y(\omega_2(b))b$, then

$$\eta_{xy}(b) = \omega_2(b)\eta_x(\omega_1(b))\omega_2(b)^{-1}, \quad \operatorname{Im}(bx) > 0.$$

10.5 Numerical example

Let us present a numerical example for the calculation of self-adjoint polynomials in free variables. We consider the polynomial $p = P(x, y) = xy + yx + x^2$ in the free variables x and y. This p has a linearization

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

which means that the Cauchy transform of p can be recovered from the operator-valued Cauchy transform of $\hat{p}$, namely, we have

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

But this $\hat{p}$ can now be written as

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

and hence is the sum of two self-adjoint variables $\tilde{X}$ and $\tilde{Y}$, which are free over $M_3(\mathbb{C})$. So we can use the subordination result from Theorem 5 in order to calculate the Cauchy transform G_p of p:

$$\begin{pmatrix} G_p(z) & * \\ * & * \end{pmatrix} = G_{\hat{p}}(b) = G_{\tilde{X}+\tilde{Y}}(b) = G_{\tilde{X}}(\omega_1(b)),$$

where $\omega_1(b)$ is determined by the fixed point equation from Theorem 5.

There are no explicit solutions of those fixed point equations in $M_3(\mathbb{C})$, but a numerical implementation relying on iterations is straightforward. One point to note is that b as defined above is not in the open set $\mathbb{H}^+(M_3(\mathbb{C}))$, but lies on its boundary. Thus, in order to be in the frame as needed in Theorem 5, one has to move inside the upper half-plane, by replacing

$$b = \begin{pmatrix} z & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad \text{by} \qquad \begin{pmatrix} z & 0 & 0 \\ 0 & i\varepsilon & 0 \\ 0 & 0 & i\varepsilon \end{pmatrix}$$

and send $\varepsilon > 0$ to zero at the end.

Figure 10.1 shows the agreement between the achieved theoretic result and the histogram of the eigenvalues of a corresponding random matrix model.

10.6 The case of rational functions

As we mentioned before, the linearization procedure works as well in the case of non-commutative rational functions. Here is an example of such a case.

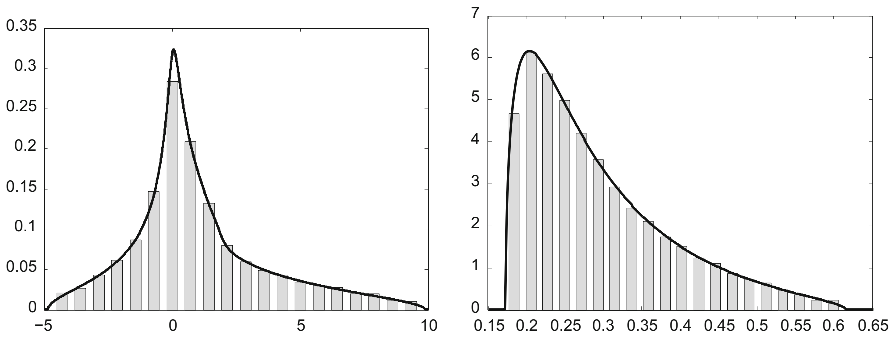


Fig. 10.1 Plots of the distribution of $p(x, y) = xy + yx + x^2$ (left) for free x, y, where x is semi-circular and y Marchenko-Pastur, and of the rational function $r(x_1, x_2)$ (right) for free semi-circular elements x_1 and x_2; in both cases the theoretical limit curve is compared with the histogram of the eigenvalues of a corresponding random matrix model

Consider the following self-adjoint rational function

$$r(x_1, x_2) = (4-x_1)^{-1} + (4-x_1)^{-1} x_2 \left((4 - x_1) - x_2(4 - x_1)^{-1} x_2\right)^{-1} x_2 (4-x_1)^{-1}$$

in two free variables x_1 and x_2. The fact that we can write it as

$$r(x_1, x_2) = \begin{pmatrix} \frac{1}{2} & 0 \end{pmatrix} \begin{pmatrix} 1 - \frac{1}{4}x_1 & -\frac{1}{4}x_2 \\ -\frac{1}{4}x_2 & 1 - \frac{1}{4}x_1 \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{2} \\ 0 \end{pmatrix}$$

gives us immediately a self-adjoint linearization of the form

$$\hat{r}(x_1, x_2) = \begin{pmatrix} 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & -1 + \frac{1}{4}x_1 & \frac{1}{4}x_2 \\ 0 & \frac{1}{4}x_2 & -1 + \frac{1}{4}x_1 \end{pmatrix}$$

$$= \begin{pmatrix} 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & -1 + \frac{1}{4}x_1 & 0 \\ 0 & 0 & -1 + \frac{1}{4}x_1 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & \frac{1}{4}x_2 \\ 0 & \frac{1}{4}x_2 & 0 \end{pmatrix}.$$

So again, we can write the linearization as the sum of two $M_3(\mathbb{C})$-free variables, and we can invoke Theorem 5 for the calculation of its operator-valued Cauchy transform. In Fig. 10.1, we compare the histogram of eigenvalues of $r(X_1, X_2)$ for one realization of independent Gaussian random matrices X_1, X_2 of size 1000×1000 with the distribution of $r(x_1, x_2)$ for free semi-circular elements x_1, x_2, calculated according to this algorithm.

Other examples for the use of operator-valued free probability methods can be found in [12].

10.7 Additional exercise

Exercise 2. Consider the C^*-algebra $M_n(\mathbb{C})$ of $n \times n$ matrices over $\mathbb{C}$. By definition, we have

$$\mathbb{H}^+(M_n(\mathbb{C})) := \{B \in M_n(\mathbb{C}) \mid \exists \varepsilon > 0 : \mathrm{Im}(B) \geq \varepsilon 1\},$$

where $\mathrm{Im}(B) := (B - B^*)/(2i)$.

(i) In the case $n = 2$, show that in fact

$$\mathbb{H}^+(M_2(\mathbb{C})) := \left\{ \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \middle| \ \mathrm{Im}(b_{11}) > 0, \mathrm{Im}(b_{11})\mathrm{Im}(b_{22}) > \frac{1}{4}|b_{12} - \overline{b_{21}}|^2 \right\}.$$

(ii) For general $n \in \mathbb{N}$, prove: if a matrix $B \in M_n(\mathbb{C})$ belongs to $\mathbb{H}^+(M_n(\mathbb{C}))$, then all eigenvalues of B lie in the complex upper half-plane $\mathbb{C}^+$. Is the converse also true?

Chapter 11
Brown Measure

The Brown measure is a generalization of the eigenvalue distribution for a general (not necessarily normal) operator in a finite von Neumann algebra (i.e. a von Neumann algebra which possesses a trace). It was introduced by Larry Brown in [46], but fell into obscurity soon after. It was revived by Haagerup and Larsen [85] and played an important role in Haagerup's investigations around the invariant subspace problem [87]. By using a "hermitization" idea, one can actually calculate the Brown measure by $M_2(\mathbb{C})$-valued free probability tools. This leads to an extension of the algorithm from the last chapter to the calculation of arbitrary polynomials in free variables. For generic non-self-adjoint random matrix models, their asymptotic complex eigenvalue distribution is expected to converge to the Brown measure of the ($*$-distribution) limit operator. However, because the Brown measure is not continuous with respect to convergence in $*$-moments, this is an open problem in the general case.

11.1 Brown measure for normal operators

Let (M, τ) be a W^*-probability space and consider an operator $a \in M$. The relevant information about a is contained in its $*$-distribution which is by definition the collection of all $*$-moments of a with respect to τ. In the case of self-adjoint or normal a, we can identify this distribution with an analytic object, a probability measure μ_a on the spectrum of a. Let us first recall these facts.

If $a = a^*$ is self-adjoint, there exists a uniquely determined probability measure μ_a on $\mathbb{R}$ such that for all $n \in \mathbb{N}$

$$\tau(a^n) = \int_{\mathbb{R}} t^n \, d\mu_a(t)$$

and the support of μ_a is the spectrum of a; see also the discussion after equation (2.2) in Chapter 2.

J.A. Mingo, R. Speicher, *Free Probability and Random Matrices*,
Fields Institute Monographs 35, DOI 10.1007/978-1-4939-6942-5_11

More general, if $a \in M$ is *normal* (i.e. $aa^* = a^*a$), then the spectral theorem provides us with a projection-valued spectral measure E_a, and the Brown measure is just the spectral measure $\mu_a = \tau \circ E_a$. Note that in the normal case μ_a may not be determined by the moments of a. Indeed, if $a = u$ is a Haar unitary, then the moments of u are the same as the moments of the zero operator. Of course, their $*$-moments are different. For a normal operator a, its spectral measure μ_a is uniquely determined by

$$\tau(a^n a^{*m}) = \int_{\mathbb{C}} z^n \bar{z}^m \, d\mu_a(z) \tag{11.1}$$

for all $m, n \in \mathbb{N}$. The support of μ_a is again the spectrum of a.

We will now try to assign to any operator $a \in M$ a probability measure μ_a on its spectrum, which contains relevant information about the $*$-distribution of a. This μ_a will be called the *Brown measure* of a. One should note that for non-normal operators, there are many more $*$-moments of a than those appearing in (11.1). There is no possibility to capture all the $*$-moments of a by the $*$-moments of a probability measure. Hence, we will necessarily loose some information about the $*$-distribution of a when we go over to the Brown measure of a. It will also turn out that we need our state τ to be a trace in order to define μ_a. Hence, in the following, we will only work in tracial W^*-probability spaces (M, τ). Recall that this means that τ is a faithful and normal trace. Von Neumann algebras which admit such faithful and normal traces are usually addressed as *finite* von Neumann algebras. If M is a finite factor, then a tracial state $\tau : M \to \mathbb{C}$ is unique on M and is automatically normal and faithful.

11.2 Brown measure for matrices

In the finite-dimensional case $M = M_n(\mathbb{C})$, the Brown measure μ_T for a normal matrix $T \in M_n(\mathbb{C})$, determined by (11.1), really is the eigenvalue distribution of the matrix. It is clear that in the case of matrices, we can extend this definition to the general, non-normal case. For a general matrix $T \in M_n(\mathbb{C})$, the spectrum $\sigma(T)$ is given by the roots of the characteristic polynomial

$$P(\lambda) = \det(\lambda I - T) = (\lambda - \lambda_1) \cdots (\lambda - \lambda_n),$$

where $\lambda_1, \ldots, \lambda_n$ are the roots repeated according to algebraic multiplicity. In this case, we have as eigenvalue distribution (and thus as Brown measure)

$$\mu_T = \frac{1}{n}(\delta_{\lambda_1} + \cdots + \delta_{\lambda_n}).$$

We want to extend this definition of μ_T to an infinite-dimensional situation. Since the characteristic polynomial does not make sense in such a situation, we have to find an analytic way of determining the roots of $P(\lambda)$ which survives also in an infinite-dimensional setting.

Consider

$$\log |P(\lambda)| = \log |\det(\lambda I - T)| = \sum_{i=1}^{n} \log |\lambda - \lambda_i|.$$

We claim that the function $\lambda \mapsto \log |\lambda|$ is harmonic in $\mathbb{C}\backslash\{0\}$ and that in general it has Laplacian

$$\nabla^2 \log |\lambda| = 2\pi \delta_0 \tag{11.2}$$

in the distributional sense. Here, the Laplacian is given by

$$\nabla^2 = \frac{\partial^2}{\partial \lambda_{\mathrm{r}}^2} + \frac{\partial^2}{\partial \lambda_{\mathrm{i}}^2},$$

where λ_{r} and λ_{i} are the real and imaginary part of $\lambda \in \mathbb{C}$. (Note that we use the symbol ∇^2 for the Laplacian, since we reserve the symbol Δ for the Fuglede-Kadison determinant of the next section.)

Let us prove this claim on the behaviour of $\log |\lambda|$. For $\lambda \neq 0$, we write ∇^2 in terms of polar coordinates

$$\nabla^2 = \frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \theta^2}$$

and have

$$\nabla^2 \log |\lambda| = \left(\frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r}\right) \log r = -\frac{1}{r^2} + \frac{1}{r^2} = 0.$$

Ignoring the singularity at 0, we can write formally

$$\begin{aligned}
\int_{B(0,r)} \nabla^2 \log |\lambda| d\lambda_{\mathrm{r}} d\lambda_{\mathrm{i}} &= \int_{B(0,r)} \operatorname{div}(\operatorname{grad} \log |\lambda|) d\lambda_{\mathrm{r}} d\lambda_{\mathrm{i}} \\
&= \int_{\partial B(0,r)} \operatorname{grad} \log |\lambda| \cdot \mathbf{n} dA \\
&= \int_{\partial B(0,r)} \frac{\mathbf{n}}{r} \cdot \mathbf{n} dA \\
&= \frac{1}{r} \cdot 2\pi r \\
&= 2\pi.
\end{aligned}$$

That is,

$$\int_{B(0,r)} \nabla^2 \log|\lambda| d\lambda_{\mathrm{r}} d\lambda_{\mathrm{i}} = 2\pi,$$

independent of $r > 0$. Hence, $\nabla^2 \log|\lambda|$ must be $2\pi\delta_0$.

Exercise 1. By integrating against a test function show rigorously that $\nabla^2 \log|\lambda| = 2\pi\delta_0$ as distributions.

Given the fact (11.2), we can now rewrite the eigenvalue distribution μ_T in the form

$$\mu_T = \frac{1}{n}(\delta_{\lambda_1} + \cdots + \delta_{\lambda_n}) = \frac{1}{2\pi n}\nabla^2 \sum_{i=1}^{n} \log|\lambda - \lambda_i| = \frac{1}{2\pi n}\nabla^2 \log|\det(T - \lambda I)|. \tag{11.3}$$

As there exists a version of the determinant in an infinite-dimensional setting, we can use this formula to generalize the definition of μ_T.

11.3 Fuglede-Kadison determinant in finite von Neumann algebras

In order to use (11.3) in infinite dimensions, we need a generalization of the determinant. Such a generalization was provided by Fuglede and Kadison [75] in 1952 for operators in a finite factor M; the case of a general finite von Neumann algebra is an straightforward extension.

Definition 1. Let (M, τ) be a tracial W^*-probability space and consider $a \in M$. Its *Fuglede-Kadison determinant* $\Delta(a)$ is defined as follows. If a is invertible, one can put

$$\Delta(a) = \exp[\tau(\log|a|)] \in (0, \infty),$$

where $|a| = (a^*a)^{1/2}$. More generally, we define

$$\Delta(a) = \lim_{\varepsilon \searrow 0} \exp\left[\tau\left(\log(a^*a + \varepsilon)^{1/2}\right)\right] \in [0, \infty).$$

By functional calculus and the monotone convergence theorem, the limit always exists.

This determinant Δ has the following properties:

- $\Delta(ab) = \Delta(a)\Delta(b)$ for all $a, b \in M$.
- $\Delta(a) = \Delta(a^*) = \Delta(|a|)$ for all $a \in M$.
- $\Delta(u) = 1$ when u is unitary.
- $\Delta(\lambda a) = |\lambda|\Delta(a)$ for all $\lambda \in \mathbb{C}$ and $a \in M$.
- $a \mapsto \Delta(a)$ is upper semicontinuous in norm-topology and in $\|\cdot\|_p$-norm for all $p > 0$.

Let us check what this definition gives in the case of matrices, $M = M_n(\mathbb{C})$, $\tau = \mathrm{tr}$. If T is invertible, then we can write

$$|T| = U \begin{pmatrix} t_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & t_n \end{pmatrix} U^*,$$

with $t_i > 0$. Then we have

$$\log |T| = U \begin{pmatrix} \log t_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \log t_n \end{pmatrix} U^*$$

and

$$\Delta(T) = \exp\left(\frac{1}{n}(\log t_1 + \cdots + \log t_n)\right) = \sqrt[n]{t_1 \cdots t_n} = \sqrt[n]{\det |T|} = \sqrt[n]{|\det T|}. \tag{11.4}$$

Note that $\det |T| = |\det T|$, because we have the polar decomposition $T = V|T|$, where V is unitary and hence $|\det V| = 1$.

Thus, we have in finite dimensions

$$\mu_T = \frac{1}{2\pi n}\nabla^2 \log |\det(T - \lambda I)| = \frac{1}{2\pi}\nabla^2(\log \Delta(T - \lambda I)).$$

So we are facing the question whether it is possible to make sense out of

$$\frac{1}{2\pi}\nabla^2(\log \Delta(a - \lambda)) \tag{11.5}$$

for operators a in general finite von Neumann algebras, where Δ denotes the Fuglede-Kadison determinant. (Here and in the following, we will write $a - \lambda$ for $a - \lambda 1$.)

11.4 Subharmonic functions and their Riesz measures

Definition 2. A function $f : \mathbb{R}^2 \to [-\infty, \infty)$ is called *subharmonic* if

(i) f is upper semicontinuous, i.e.

$$f(z) \geq \limsup_{n\to\infty} f(z_n), \qquad \text{whenever} \qquad z_n \to z;$$

(ii) f satisfies the submean inequality: for every circle the value of f at the centre is less or equal to the mean value of f over the circle, i.e.

$$f(z) \le \frac{1}{2\pi}\int_0^{2\pi} f(z+re^{i\theta})d\theta;$$

(*iii*) f is not constantly equal to $-\infty$.

If f is subharmonic then f is Borel measurable, $f(z) > -\infty$ almost everywhere with respect to Lebesgue measure and $f \in L^1_{loc}(\mathbb{R}^2)$. One has the following classical theorem for subharmonic functions; e.g. see [13, 92].

Theorem 3. *If f is subharmonic on $\mathbb{R}^2 \equiv \mathbb{C}$, then $\nabla^2 f$ exists in the distributional sense, and it is a positive Radon measure ν_f; i.e. ν_f is uniquely determined by*

$$\frac{1}{2\pi}\int_{\mathbb{R}^2} f(\lambda)\cdot\nabla^2\varphi(\lambda)d\lambda_{\mathrm{r}}d\lambda_{\mathrm{i}} = \int_{\mathbb{C}}\varphi(z)d\nu_f(z) \qquad \textit{for all } \varphi \in C_c^\infty(\mathbb{R}^2).$$

If ν_f has compact support, then

$$f(\lambda) = \frac{1}{2\pi}\int_{\mathbb{C}} \log|\lambda - z|d\nu_f(z) + h(\lambda),$$

where h is a harmonic function on $\mathbb{C}$.

Definition 4. The measure $\nu_f = \nabla^2 f$ is called the *Riesz measure* of the subharmonic function f.

11.5 Definition of the Brown measure

If we apply this construction to our question about (11.5), we get the construction of the Brown measure as follows. This was defined by L. Brown in [46] (for the case of factors); for more information, see also [85].

Theorem 5. *Let (M,τ) be a tracial W^*-probability space. Then we have:*

(*i*) *The function $\lambda \mapsto \log \Delta(a-\lambda)$ is subharmonic.*
(*ii*) *The corresponding Riesz measure*

$$\mu_a := \frac{1}{2\pi}\nabla^2 \log\Delta(a-\lambda) \tag{11.6}$$

is a probability measure on $\mathbb{C}$ with support contained in the spectrum of a.
(*iii*) *Moreover, one has for all $\lambda \in \mathbb{C}$*

$$\int_{\mathbb{C}} \log|\lambda - z|d\mu_a(z) = \log\Delta(a-\lambda) \tag{11.7}$$

and this characterizes μ_a among all probability measures on $\mathbb{C}$.

Definition 6. The measure μ_a from Theorem 5 is called the *Brown measure* of a.

Proof [Sketch of Proof of Theorem 5(i)]: Suppose $a \in M$. We want to show that $f(\lambda) := \log \Delta(a-\lambda)$ is subharmonic. We have

$$\Delta(a) = \lim_{\varepsilon \searrow 0} \exp\big[\tau\big(\log(a^*a+\varepsilon)^{1/2}\big)\big].$$

Thus

$$\log \Delta(a) = \frac{1}{2}\lim_{\varepsilon \searrow 0} \tau(\log(a^*a+\varepsilon)),$$

as a decreasing limit as $\varepsilon \searrow 0$. So, with the notations

$$a_\lambda := a-\lambda, \qquad f_\varepsilon(\lambda) := \frac{1}{2}\tau(\log(a_\lambda^* a_\lambda + \varepsilon)),$$

we have

$$f(\lambda) = \lim_{\varepsilon \searrow 0} f_\varepsilon(\lambda).$$

For $\varepsilon > 0$, the function f_ε is a C^2-function, and therefore f_ε being subharmonic is equivalent to $\nabla^2 f_\varepsilon \geq 0$ as a function. But $\nabla^2 f_\varepsilon$ can be computed explicitly:

$$\nabla^2 f_\varepsilon(\lambda) = 2\varepsilon\tau\big((a_\lambda a_\lambda^* + \varepsilon)^{-1}(a_\lambda^* a_\lambda + \varepsilon)^{-1}\big). \tag{11.8}$$

Since we have for general positive operators x and y that $\tau(xy) = \tau(x^{1/2}yx^{1/2}) \geq 0$, we see that $\nabla^2 f_\varepsilon(\lambda) \geq 0$ for all $\lambda \in \mathbb{C}$ and thus f_ε is subharmonic.

The fact that $f_\varepsilon \searrow f$ implies then that f is upper semicontinuous and satisfies the submean inequality. Furthermore, if $\lambda \notin \sigma(a)$, then $a-\lambda$ is invertible; hence, $\Delta(a-\lambda) > 0$, and thus $f(\lambda) \neq -\infty$. Hence, f is subharmonic. □

Exercise 2. We want to prove here (11.8). We consider $f_\varepsilon(\lambda)$ as a function in λ and $\bar\lambda$; hence, the Laplacian is given by (where as usual $\lambda = \lambda_r + i\lambda_i$)

$$\nabla^2 = \frac{\partial^2}{\partial\lambda_r^2} + \frac{\partial^2}{\partial\lambda_i^2} = 4\frac{\partial^2}{\partial\bar\lambda\partial\lambda}$$

where

$$\frac{\partial}{\partial\lambda} = \frac{1}{2}\left(\frac{\partial}{\partial\lambda_r} - i\frac{\partial}{\partial\lambda_i}\right), \qquad \frac{\partial}{\partial\bar\lambda} = \frac{1}{2}\left(\frac{\partial}{\partial\lambda_r} + i\frac{\partial}{\partial\lambda_i}\right).$$

(*i*) Show that we have for each $n \in \mathbb{N}$ (by relying heavily on the fact that τ is a trace)

$$\frac{\partial}{\partial \lambda}\tau[(a_\lambda^* a_\lambda)^n] = -n\tau[(a_\lambda^* a_\lambda)^{n-1} a_\lambda^*]$$

and

$$\frac{\partial}{\partial \bar{\lambda}}\tau[(a_\lambda^* a_\lambda)^n a_\lambda^*] = -\sum_{j=0}^{n} \tau[(a_\lambda a_\lambda^*)^j (a_\lambda^* a_\lambda)^{n-j}].$$

(*ii*) Prove (11.8) by using the power series expansion of

$$\log(a_\lambda^* a_\lambda + \varepsilon) = \log \varepsilon + \log\left(1 + \frac{a_\lambda^* a_\lambda}{\varepsilon}\right).$$

In the case of a normal operator, the Brown measure is just the spectral measure $\tau \circ E_a$, where E_a is the projection-valued spectral measure according to the spectral theorem. In that case, μ_a is determined by the equality of the $*$-moments of μ_a and of a, i.e. by

$$\int_{\mathbb{C}} z^n \bar{z}^m d\mu_a(z) = \tau(a^n a^{*m}) \qquad \text{if } a \text{ is normal}$$

for all $m, n \in \mathbb{N}$. If a is not normal, then this equality does not hold anymore. Only the equality of the moments is always true, i.e. for all $n \in \mathbb{N}$

$$\int_{\mathbb{C}} z^n d\mu_a(z) = \tau(a^n) \qquad \text{and} \qquad \int_{\mathbb{C}} \bar{z}^n d\mu_a(z) = \tau(a^{*n}).$$

One should note, however, that the Brown measure of a is in general actually determined by the $*$-moments of a. This is the case, since τ is faithful and the Brown measure depends only on τ restricted to the von Neumann algebra generated by a; the latter is uniquely determined by the $*$-moments of a; see also Chapter 6, Theorem 6.2.

What one can say in general about the relation between the $*$-moments of μ_a and of a is the following *generalized Weyl Inequality* of Brown [46]. For any $a \in M$ and $0 < p < \infty$, we have

$$\int_{\mathbb{C}} |z|^p d\mu_a(z) \le \|a\|_p^p = \tau(|a|^p).$$

This was strengthened by Haagerup and Schultz [87] in the following way: If M_{inv} denotes the invertible elements in M, then we actually have for all $a \in M$ and every $p > 0$ that

$$\int_{\mathbb{C}} |z|^p d\mu_a(z) = \inf_{b \in M_{inv}} \|bab^{-1}\|_p^p.$$

Note here that because of $\Delta(bab^{-1}) = \Delta(a)$, we have $\mu_{bab^{-1}} = \mu_a$ for $b \in M_{inv}$.

Exercise 3. Let (M, τ) be a tracial W^*-probability space and $a \in M$. Let $p(z)$ be a polynomial in the variable z (not involving $\bar{z}$), hence $p(a) \in M$. Show that the Brown measure of $p(a)$ is the push-forward of the Brown measure of a, i.e. $\mu_{p(a)} = p_*(\mu_a)$, where the push-forward $p_*(\nu)$ of a measure ν is defined by $p_*(\nu)(E) = \nu(p^{-1}(E))$ for any measurable set E.

The calculation of the Brown measure of concrete non-normal operators is usually quite hard, and there are not too many situations where one has explicit solutions. We will in the following present some of the main concrete results.

11.6 Brown measure of R-diagonal operators

R-diagonal operators were introduced by Nica and Speicher [136]. They provide a class of, in general non-normal, operators which are usually accessible to concrete calculations. In particular, one is able to determine their Brown measure quite explicitly.

R-diagonal operators can be considered in general $*$-probability spaces, but we will restrict here to the tracial W^*-probability space situation; only there the notion of Brown measure makes sense.

Definition 7. An operator a in a tracial W^*-probability space (M, τ) is called *R-diagonal* if its only non-vanishing $*$-cumulants (i.e. cumulants where each argument is either a or a^*) are alternating, i.e. of the form $\kappa_{2n}(a, a^*, a, a^*, \ldots, a, a^*) = \kappa_{2n}(a^*, a, a^*, a \ldots, a^*, a)$ for some $n \in \mathbb{N}$.

Main examples for R-diagonal operators are Haar unitaries and Voiculescu's circular operator. With the exception of multiples of Haar unitaries, R-diagonal operators are not normal. One main characterization [136] of R-diagonal operators is the following: a is R-diagonal if and only if a has the same $*$-distribution as up where u is a Haar unitary, $p \geq 0$, and u and p are $*$-free. If $\ker(a) = \{0\}$, then this can be refined to the characterization that R-diagonal operators have a polar decomposition of the form $a = u|a|$, where u is Haar unitary and $|a|$ is $*$-free from u.

The Brown measure of R-diagonal operators was calculated by Haagerup and Larsen [85]. The following theorem contains their main statements on this.

Theorem 8. *Let (M, τ) be a tracial W^*-probability space and $a \in M$ be R-diagonal. Assume that* $\ker(a) = \{0\}$ *and that a^*a is not a constant operator. Then we have the following:*

(*i*) *The support of the Brown measure μ_a is given by*

$$supp(\mu_a) = \{z \in \mathbb{C} \mid \|a^{-1}\|_2^{-1} \leq |z| \leq \|a\|_2\}, \tag{11.9}$$

where we put $\|a^{-1}\|_2^{-1} = 0$ if $a^{-1} \notin L^2(M, \tau)$.

(*ii*) μ_a *is invariant under rotations about* $0 \in \mathbb{C}$.
(*iii*) *For* $0 < t < 1$, *we have*

$$\mu_a(B(0,r)) = t \qquad \textit{for} \qquad r = \frac{1}{\sqrt{S_{a^*a}(t-1)}}, \tag{11.10}$$

where S_{a^*a} *is the* S*-transform of the operator* a^*a *and* $B(0,r)$ *is the open disk with radius* r.
(*iv*) *The conditions (i), (ii), and (iii) determine* μ_a *uniquely.*
(*v*) *The spectrum of an* R*-diagonal operator* a *coincides with supp*(μ_a) *unless* $a^{-1} \in L^2(M,\tau)\backslash M$ *in which case supp*(μ_a) *is the annulus* (11.9)*, while the spectrum of* a *is the full closed disk with radius* $\|a\|_2$.

For the third part, one has to note that

$$t \mapsto \frac{1}{\sqrt{S_{a^*a}(t-1)}}$$

maps $(0,1)$ onto $(\|a^{-1}\|_2^{-1}, \|a\|_2)$.

11.6.1 A little about the proof

We give some key ideas of the proof from [85]; for another proof, see [158].

Consider $\lambda \in \mathbb{C}$ and put $\alpha := |\lambda|$. A key point is to find a relation between $\mu_{|a|}$ and $\mu_{|a-\lambda|}$. For a probability measure σ, we denote its symmetrized version by $\tilde{\sigma}$, i.e. for any measurable set E, we have $\tilde{\sigma}(E) = (\sigma(E) + \sigma(-E))/2$. Then one has the relation

$$\tilde{\mu}_{|a-\lambda|} = \tilde{\mu}_{|a|} \boxplus \frac{1}{2}(\delta_\alpha + \delta_{-\alpha}), \tag{11.11}$$

or in terms of the R-transforms:

$$R_{\tilde{\mu}_{|a-\lambda|}}(z) = R_{\tilde{\mu}_{|a|}}(z) + \frac{\sqrt{1+4\alpha^2 z^2} - 1}{2z}.$$

Hence, $\mu_{|a|}$ determines $\mu_{|a-\lambda|}$, which determines

$$\int_{\mathbb{C}} \log|\lambda - z| d\mu_a(z) = \log \Delta(a-\lambda) = \log \Delta(|a-\lambda|) = \int_0^\infty \log(t) d\mu_{|a-\lambda|}(t).$$

Exercise 4. Prove (11.11) by showing that if a is R-diagonal then the matrices

$$\begin{pmatrix} 0 & a \\ a^* & 0 \end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix} 0 & \lambda \\ \bar{\lambda} & 0 \end{pmatrix}$$

are free in the $(M_2(\mathbb{C}) \otimes M, \mathrm{tr} \otimes \tau)$.

11.6.2 Example: circular operator

Let us consider, as a concrete example, the circular operator $c = (s_1 + i s_2)/\sqrt{2}$, where s_1 and s_2 are free standard semi-circular elements.

The distribution of c^*c is free Poisson with rate 1, given by the density $\sqrt{4-t}/2\pi t$ on $[0,4]$, and thus the distribution $\mu_{|c|}$ of the absolute value $|c|$ is the quarter-circular distribution with density $\sqrt{4-t^2}/\pi$ on $[0,2]$. We have $\|c\|_2 = 1$ and $\|c^{-1}\|_2 = \infty$, and hence the support of the Brown measure of c is the closed unit disk, $\mathrm{supp}(\mu_c) = \overline{B(0,1)}$. This coincides with the spectrum of c.

In order to apply Theorem 8, we need to calculate the S-transform of c^*c. We have $R_{c^*c}(z) = 1/(1-z)$, and thus $S_{c^*c}(z) = 1/(1+z)$ (because $z \mapsto zR(z)$ and $w \mapsto wS(w)$ are inverses of each other; see [137, Remark 16.18] and also the discussion around [137, Eq. (16.8)]). So, for $0 < t < 1$, we have $S_{c^*c}(t-1) = 1/t$. Thus, $\mu_c(B(0,\sqrt{t})) = t$, or, for $0 < r < 1$, $\mu_c(B(0,r)) = r^2$. Together with the rotation invariance, this shows that μ_c is the uniform measure on the unit disk $\overline{B(0,1)}$.

11.6.3 The circular law

The circular law is the non-self-adjoint version of Wigner's semi-circle law. Consider an $N \times N$ matrix where all entries are independent and identically distributed. If the distribution of the entries is Gaussian, then this ensemble is also called *Ginibre ensemble*. It is very easy to check that the $*$-moments of the Ginibre random matrices converge to the corresponding $*$-moments of the circular operator. So it is quite plausible to expect that the Brown measure (i.e. the eigenvalue distribution) of the Ginibre random matrices converges to the Brown measure of the circular operator, i.e. to the uniform distribution on the disk. This statement is known as the *circular law*. However, one has to note that the above is not a proof for the circular law, because the Brown measure is not continuous with respect to our notion of convergence in $*$-distribution. One can construct easily examples where this fails.

Exercise 5. Consider the sequence $(T_N)_{N\geq 2}$ of nilpotent $N \times N$ matrices

$$T_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad T_3 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad T_4 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad T_5 = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \cdots$$

Show that,

- ◦ with respect to tr, T_N converges in $*$-moments to a Haar unitary element,
- ◦ the Brown measure of a Haar unitary element is the uniform distribution on the circle of radius 1,
- ◦ but the asymptotic eigenvalue distribution of T_N is given by δ_0.

However, for nice random matrix ensembles, the philosophy of convergence of the eigenvalue distribution to the Brown measure of the limit operator seems to be correct. For the Ginibre ensemble, one can write down quite explicitly its eigenvalue distribution, and then it is easy to check the convergence to the circular law. If the distribution of the entries is not Gaussian, then one still has convergence to the circular law under very general assumptions (only second moment of the distribution has to exist), but proving this in full generality has only been achieved recently. For a survey on this, see [42, 171].

11.6.4 The single ring theorem

There are also canonical random matrix models for R-diagonal operators. If one considers on (non-self-adjoint) $N \times N$ matrices a density of the form

$$P_N(A) = \text{const} \cdot e^{-\frac{N}{2}\mathrm{Tr}(f(A^*A))},$$

then one can check, under suitable assumptions on the function f, that the $*$-distribution of the corresponding random matrix A converges to an R-diagonal operator (whose concrete form is of course determined in terms of f). So again one expects that the eigenvalue distribution of those random matrices converges to the Brown measure of the limit R-diagonal operator, whose form is given in Theorem 8. (In particular, this limiting eigenvalue distribution lives on an, possibly degenerate, annulus, i.e. a single ring, even if f has several minima.) This has been proved recently by Guionnet, Krishnapur, and Zeitouni [82].

11.7 Brown measure of elliptic operators

An *elliptic operator* is of the form $a = \alpha s_1 + i\beta s_2$, where $\alpha, \beta > 0$ and s_1 and s_2 are free standard semi-circular operators. An elliptic operator is not R-diagonal, unless $\alpha = \beta$ (in which case it is a circular operator). The following theorem was proved by Larsen [116] and by Biane and Lehner [38].

Theorem 9. *Consider the elliptic operator*

$$a = (\cos\theta)s_1 + i(\sin\theta)s_2, \qquad 0 < \theta < \frac{\pi}{2}.$$

Put $\gamma := \cos(2\theta)$ *and* $\lambda = \lambda_{\mathrm{r}} + i\lambda_{\mathrm{i}}$. *Then the spectrum of* a *is the ellipse*

$$\sigma(a) = \left\{\lambda \in \mathbb{C} \mid \frac{\lambda_{\mathrm{r}}^2}{(1+\gamma)^2} + \frac{\lambda_{\mathrm{i}}^2}{(1-\gamma)^2} \le 1\right\},$$

and the Brown measure μ_a *is the measure with constant density on* $\sigma(a)$*:*

$$d\mu_a(\lambda) = \frac{1}{\pi(1-\gamma^2)} 1_{\sigma(a)}(\lambda) d\lambda_{\mathrm{r}} d\lambda_{\mathrm{i}}.$$

11.8 Brown measure for unbounded operators

The Brown measure can also be extended to unbounded operators which are affiliated to a tracial W^*-probability space; for the notion of "affiliated operators", see our discussion before Definition 8.15 in Chapter 8. This extension of the Brown measure was done by Haagerup and Schultz in [86].

Δ and μ_a can be defined for unbounded a provided $\int_1^\infty \log(t) d\mu_{|a|}(t) < \infty$, in which case

$$\Delta(a) = \exp\left(\int_0^\infty \log(t) d\mu_{|a|}(t)\right) \in [0, \infty),$$

and the Brown measure μ_a is still determined by (11.7).

Example 10. Let c_1 and c_2 be two $*$-free circular elements and consider $a := c_1 c_2^{-1}$. If c_1, c_2 live in the tracial W^*-probability space (M, τ), then $a \in L^p(M, \tau)$ for $0 < p < 1$. In this case, $\Delta(a - \lambda)$ and μ_a are well defined. In order to calculate μ_a, one has to extend the class of R-diagonal operators and the formulas for their Brown measure to unbounded operators. This was done in [86]. Since the product of an R-diagonal element with a $*$ free element is R-diagonal, too, we have that a is R-diagonal. So to use (the unbounded version of) Theorem 8, we need to calculate the S-transform of a^*a. Since with c_2, also its inverse c_2^{-1} is R-diagonal, we have $S_{|a|^2} = S_{|c_1|^2} S_{|c_2^{-1}|^2}$. The S-transform of the first factor is $S_{|c_1|^2}(z) = 1/(1+z)$; compare Section 11.6.2. Furthermore, the S-transforms of x and x^{-1} are, for positive x, in general related by $S_x(z) = 1/S_{x^{-1}}(-1-z)$. Since $|c_2^{-1}|^2 = |c_2^*|^{-2}$ and since c_2^* has the same distribution as c_2, we have that $S_{|c_2^{-1}|^2} = S_{|c_2|^{-2}}$ and thus

$$S_{|c_2^{-1}|^2}(z) = S_{|c_2|^{-2}} = \frac{1}{S_{|c_2|^2}(-1-z)} = \frac{1}{\frac{1}{1-1-z}} = -z.$$

This gives then $S_{|a|^2}(z) = -z/(1+z)$, for $-1 < z < 0$, or $S_{|a|^2}(t-1) = (1-t)/t$ for $0 < t < 1$. So our main formula (11.10) from Theorem 8 gives $\mu_a(B(0, \sqrt{t/(1-t)})) = t$ or $\mu_a(B(0,r)) = r^2/(1+r^2)$. We have $\|a\|_2 = \infty = \|a^{-1}\|_2$, and thus $\operatorname{supp}(\mu_a) = \mathbb{C}$. The above formula for the measure of balls gives then the density

$$d\mu_a(\lambda) = \frac{1}{\pi}\frac{1}{(1+|\lambda|^2)^2} d\lambda_{\mathrm{r}} d\lambda_{\mathrm{i}}. \tag{11.12}$$

For more details and, in particular, proofs of the above used facts about R-diagonal elements and the relation between S_x and $S_{x^{-1}}$, one should see the original paper of Haagerup and Schultz [86].

11.9 Hermitization method: using operator-valued free probability for calculating the Brown measure

Note that formula (11.7) for determining the Brown measure can also be written as

$$\int_{\mathbb{C}} \log|\lambda - z| d\mu_a(z) = \log \Delta(a-\lambda) = \log \Delta(|a-\lambda|) = \int_0^\infty \log(t) d\mu_{|a-\lambda|}(t). \tag{11.13}$$

This tells us that we can understand the Brown measure of a non-normal operator a if we understand the distributions of all Hermitian operators $|a-\lambda|$ for all $\lambda \in \mathbb{C}$ sufficiently well. In the random matrix literature, this idea goes back at least to Girko [77] and is usually addressed as *hermitization method*. A contact of this idea with the world of free probability was made on a formal level in the works of Janik, Nowak, Papp, and Zahed [103] and of Feinberg and Zee [71]. In [24], it was shown that operator-valued free probability is the right frame to deal with this rigorously. (Examples for explicit operator-valued calculations were also done before in [1].) Combining this hermitization idea with the subordination formulation of operator-valued free convolution allows then to calculate the Brown measure of any (not just self-adjoint) polynomial in free variables.

In order to make this connection between Brown measure and operator-valued quantities more precise, we first have to rewrite our description of the Brown measure. In Section 11.5, we have seen that we get the Brown measure of a as the limit for $\varepsilon \to 0$ of

$$\nabla^2 f_\varepsilon(\lambda) = 2\varepsilon\tau\big((a_\lambda a_\lambda^* + \varepsilon)^{-1}(a_\lambda^* a_\lambda + \varepsilon)^{-1}\big), \qquad \text{where} \qquad a_\lambda := a - \lambda.$$

This can also be reformulated in the following form (compare [116], or Lemma 4.2 in [1]: Let us define

$$G_{\varepsilon,a}(\lambda) := \tau\big((\lambda-a)^*\big((\lambda-a)(\lambda-a)^*+\varepsilon^2\big)^{-1}\big). \tag{11.14}$$

Then

$$\mu_{\varepsilon,a} = \frac{1}{\pi}\frac{\partial}{\partial\bar{\lambda}}G_{\varepsilon,a}(\lambda) \tag{11.15}$$

is a probability measure on the complex plane (whose density is given by $\nabla^2 f_\varepsilon$), which converges weakly for $\varepsilon \to 0$ to the Brown measure of a.

In order to calculate the Brown measure, we need $G_{\varepsilon,a}(\lambda)$ as defined in (11.14). Let now

$$A = \begin{pmatrix} 0 & a \\ a^* & 0 \end{pmatrix} \in M_2(M).$$

Note that A is self-adjoint. Consider A in the $M_2(\mathbb{C})$-valued probability space with respect to $E = id \otimes \tau : M_2(M) \to M_2(\mathbb{C})$ given by

$$E\left[\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\right] = \begin{pmatrix} \tau(a_{11}) & \tau(a_{12}) \\ \tau(a_{21}) & \tau(a_{22}) \end{pmatrix}.$$

For the argument

$$\Lambda_\varepsilon = \begin{pmatrix} i\varepsilon & \lambda \\ \bar{\lambda} & i\varepsilon \end{pmatrix} \in M_2(\mathbb{C})$$

consider now the $M_2(\mathbb{C})$-valued Cauchy transform of A

$$G_A(\Lambda_\varepsilon) = E\big[(\Lambda_\varepsilon - A)^{-1}\big] =: \begin{pmatrix} g_{11}(\varepsilon,\lambda) & g_{12}(\varepsilon,\lambda) \\ g_{21}(\varepsilon,\lambda) & g_{22}(\varepsilon,\lambda) \end{pmatrix}.$$

One can easily check that $(\Lambda_\varepsilon - A)^{-1}$ is actually given by

$$\begin{pmatrix} -i\varepsilon((\lambda-a)(\lambda-a)^*+\varepsilon^2)^{-1} & (\lambda-a)((\lambda-a)^*(\lambda-a)+\varepsilon^2)^{-1} \\ (\lambda-a)^*((\lambda-a)(\lambda-a)^*+\varepsilon^2)^{-1} & -i\varepsilon((\lambda-a)^*(\lambda-a)+\varepsilon^2)^{-1} \end{pmatrix},$$

and thus we are again in the situation that our quantity of interest is actually one entry of an operator-valued Cauchy transform: $G_{\varepsilon,a}(\lambda) = g_{21}(\varepsilon,\lambda) = [G_A(\Lambda_\varepsilon)]_{21}$.

11.10 Brown measure of arbitrary polynomials in free variables

So in order to calculate the Brown measure of some polynomial p in self-adjoint free variables, we should first hermitize the problem by going over to self-adjoint 2×2 matrices over our underlying space, and then we should linearize the problem on this level and use finally our subordination description of operator-valued free convolution to deal with this linear problem. It might be not so clear whether hermitization and linearization go together well, but this is indeed the case. Essentially we do here a linearization of an operator-valued model instead of a scalar-valued one: we have to linearize a polynomial in matrices. But the linearization algorithm works in this case as well. As the end is near, let us illustrate this just with an example. For more details, see [95].

Example 11. Consider the polynomial $a = xy$ in the free self-adjoint variables $x = x^*$ and $y = y^*$. For the Brown measure of this a, we have to calculate the operator-valued Cauchy transform of

$$A = \begin{pmatrix} 0 & a \\ a^* & 0 \end{pmatrix} = \begin{pmatrix} 0 & xy \\ yx & 0 \end{pmatrix}.$$

In order to linearize this, we should first write it as a polynomial in matrices of x and matrices of y. This can be achieved as follows:

$$\begin{pmatrix} 0 & xy \\ yx & 0 \end{pmatrix} = \begin{pmatrix} x & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & y \\ y & 0 \end{pmatrix} \begin{pmatrix} x & 0 \\ 0 & 1 \end{pmatrix} = XYX,$$

which is a self-adjoint polynomial in the self-adjoint variables

$$X = \begin{pmatrix} x & 0 \\ 0 & 1 \end{pmatrix} \qquad \text{and} \qquad Y = \begin{pmatrix} 0 & y \\ y & 0 \end{pmatrix}.$$

This self-adjoint polynomial XYX has a self-adjoint linearization

$$\begin{pmatrix} 0 & 0 & X \\ 0 & Y & -1 \\ X & -1 & 0 \end{pmatrix}.$$

Plugging in back the 2×2 matrices for X and Y, we get finally the self-adjoint linearization of A as

$$\begin{pmatrix} 0 & 0 & 0 & 0 & x & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & y & -1 & 0 \\ 0 & 0 & y & 0 & 0 & -1 \\ x & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & x & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \\ x & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & y & 0 & 0 \\ 0 & 0 & y & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

We have written this as the sum of two $M_6(\mathbb{C})$-free matrices, both of them being self-adjoint. For calculating the Cauchy transform of this sum, we can then use

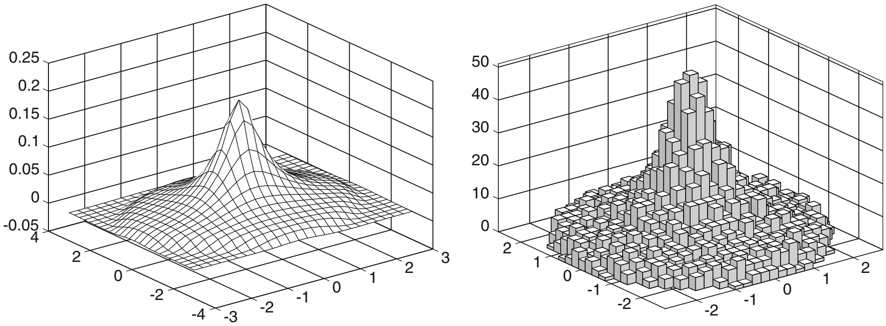


Fig. 11.1 Brown measure (*left*) of $p(x, y, z) = xyz - 2yzx + zxy$ with x, y, z free semicircles, compared to histogram (*right*) of the complex eigenvalues of $p(X, Y, Z)$ for independent Wigner matrices with $N = 5000$

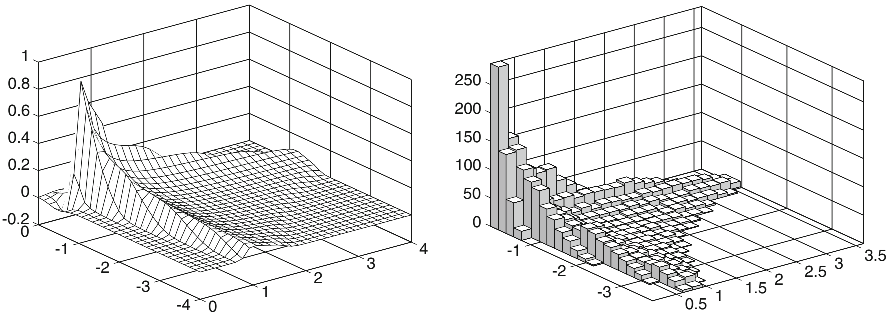


Fig. 11.2 Brown measure (*left*) of $p(x, y) = x + iy$ with x, y free Poissons of rate 1, compared to histogram (*right*) of the complex eigenvalues of $p(X, Y)$ for independent Wishart matrices X and Y with $N = 5000$

again the subordination algorithm for the operator-valued free convolution from Theorem 10.5. Putting all the steps together gives an algorithm for calculating the Brown measure of $a = xy$. One might note that in the case where both x and y are even elements (i.e. all odd moments vanish), the product is actually R-diagonal; see [137, Theorem 15.17]. Hence, in this case, we even have an explicit formula for the Brown measure of xy, given by Theorem 8 and the fact that we can calculate the S-transform of a^*a in terms of the S-transforms of x and of y.

Of course, we expect that the eigenvalue distribution of our polynomial evaluated in asymptotically free matrices (like independent Wigner or Wishart matrices) should converge to the Brown measure of the polynomial in the corresponding free variables. However, as was already pointed out before (see the discussion around Exercise 5), this is not automatic from the convergence of all $*$-moments, and one actually has to control probabilities of small eigenvalues during all of the calculations. Such controls have been achieved in the special cases of the circular law or the single ring theorem. However, for an arbitrary polynomial in asymptotically free matrices, this is an open problem at the moment.

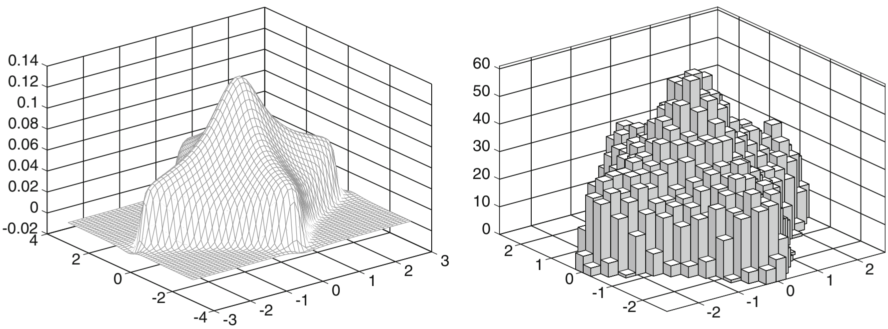


Fig. 11.3 Brown measure (*left*) of $p(x_1, x_2, x_3, x_4) = x_1x_2 + x_2x_3 + x_3x_4 + x_4x_1$ with x_1, x_2, x_3, x_4 free semicircles, compared to histogram (*right*) of the complex eigenvalues of $p(X_1, X_2, X_3, X_4)$ for independent Wigner matrices X_1, X_2, X_3, X_4 with $N = 7000$

In Figs. 11.1, 11.2 and 11.3, we give for some polynomials the Brown measure calculated according to the algorithm outlined above, and we also compare this with histograms of the complex eigenvalues of the corresponding polynomials in independent random matrices.

Chapter 12
Solutions to Exercises

12.1 Solutions to exercises in Chapter 1

1. Let ν be a probability measure on $\mathbb{R}$ such that $\int_{\mathbb{R}} |t|^n \, d\nu(t) < \infty$. For $m \leq n$,

$$\begin{aligned}
\int_{\mathbb{R}} |t|^m \, d\nu(t) &= \int_{|t|\leq 1} |t|^m \, d\nu(t) + \int_{|t|>1} |t|^m \, d\nu(t) \\
&\leq \int_{|t|\leq 1} 1 \, d\nu(t) + \int_{|t|>1} |t|^n \, d\nu(t) \\
&\leq \nu(\mathbb{R}) + \int_{\mathbb{R}} |t|^n \, d\nu(t) \\
&< \infty.
\end{aligned}$$

2. Since ν has a fifth moment, we can write

$$\varphi(t) = 1 + \alpha_1 \frac{(it)}{1!} + \alpha_2 \frac{(it)^2}{2!} + \alpha_3 \frac{(it)^3}{3!} + \alpha_4 \frac{(it)^4}{4!} + o(t^4)$$

and

$$\log(\varphi(t)) = k_1 \frac{(it)}{1!} + k_2 \frac{(it)^2}{2!} + k_3 \frac{(it)^3}{3!} + k_4 \frac{(it)^4}{4!} + o(t^4).$$

The expansion for $\log(1+x)$ is $x - x^2/2 + x^3/3 - x^4/4 + o(x^4)$. Let $s = it$. Thus

$$\begin{aligned}
\log(\varphi(t)) = &\left\{\alpha_1 \frac{s}{1!} + \alpha_2 \frac{s^2}{2!} + \alpha_3 \frac{s^3}{3!} + \alpha_4 \frac{s^4}{4!}\right\} - \frac{1}{2}\left\{\alpha_1 \frac{s}{1!} + \alpha_2 \frac{s^2}{2!} + \alpha_3 \frac{s^3}{3!}\right\}^2 \\
&+ \frac{1}{3}\left\{\alpha_1 \frac{s}{1!} + \alpha_2 \frac{s^2}{2!}\right\}^3 - \frac{1}{4}\left\{\alpha_1 \frac{s}{1!}\right\}^4 + o(s^4).
\end{aligned}$$

J.A. Mingo, R. Speicher, *Free Probability and Random Matrices*,
Fields Institute Monographs 35, DOI 10.1007/978-1-4939-6942-5_12

The only term of degree 1 is α_1 so $k_1 = \alpha_1$. The terms of degree 2 are

$$\frac{\alpha_2}{2!} - \frac{1}{2}\alpha_1^2 = \frac{1}{2!}(\alpha_2 - \alpha_1^2), \quad \text{so} \quad k_2 = \alpha_2 - \alpha_1^2.$$

The terms of degree 3 are

$$\frac{\alpha_3}{3!} - \frac{1}{2}\Big(2\alpha_1\frac{\alpha_2}{2!}\Big) + \frac{\alpha_1^3}{3} = \frac{1}{3!}(\alpha_3 - 3\alpha_1\alpha_2 + 2\alpha_1^3), \quad \text{so} \quad k_3 = \alpha_3 - 3\alpha_1\alpha_2 + 2\alpha_1^3.$$

The terms of degree 4 are

$$\begin{aligned}&\frac{\alpha_4}{4!} - \frac{1}{2}\Big(\frac{\alpha_2^2}{(2!)^2} + 2\alpha_1\frac{\alpha_3}{3!}\Big) + \frac{1}{3}\Big(3\alpha_1^2\frac{\alpha_2}{2!}\Big) - \frac{1}{4}\alpha_1^4\\ &\quad = \frac{1}{4!}\Big(\alpha_4 - 4\alpha_1\alpha_3 - 3\alpha_2^2 + 12\alpha_1^2\alpha_2 - 6\alpha_1^4\Big).\end{aligned}$$

Summarizing, let us put this in a table.

$$\begin{aligned}k_1 &= \alpha_1\\ k_2 &= \alpha_2 - \alpha_1^2\\ k_3 &= \alpha_3 - 3\alpha_2\alpha_1 + 2\alpha_1^3\\ k_4 &= \alpha_4 - 4\alpha_3\alpha_1 - 3\alpha_2^3 + 12\alpha_2\alpha_1^2 - 6\alpha_1^4;\end{aligned}$$

$$\begin{aligned}\alpha_1 &= k_1\\ \alpha_2 &= k_2 + k_1^2\\ \alpha_3 &= k_3 + 3k_2k_1 + k_1^3\\ \alpha_4 &= k_4 + 4k_3k_1 + 3k_2^2 + 6k_2k_1^2 + k_1^4.\end{aligned}$$

3. Suppose $(r_1, \dots, r_n)$ is a type, i.e. $r_1, \dots, r_n \geq 0$ and $1 \cdot r_1 + \cdots + n \cdot r_n = n$. Let us count the number of partitions of $[n]$ with type $(r_1, \dots, r_n)$. Let $m = r_1 + \cdots + r_n$ be the number of blocks and $l_1, \dots, l_m$ the size of the blocks. Then $(l_1, \dots, l_m)$ is a composition of the integer n with type $(r_1, \dots, r_n)$. There are $\binom{n}{l_1}$ ways of choosing the elements of the first block, $\binom{n-l_1}{l_2}$ ways of choosing the elements of the second block and finally $\binom{n-l_1-l_2-\cdots-l_{m-1}}{l_m}$ ways of choosing the elements of the last block. Multiplying these out we get

$$\binom{n}{l_1}\binom{n-l_1}{l_2} \times \cdots \times \binom{n-l_1-\cdots-l_{m-1}}{l_m} = \frac{n!}{l_1!l_2!\cdots l_m!}.$$

However, this overcounts because we don't distinguish between permutations of the r_1 blocks of size 1, the r_2 blocks of size 2, etc. Thus we must divide by $r_1!\cdots r_n!$. Also we may write $l_1!\cdots l_m!$ as $(1!)^{r_1}\cdots(n!)^{r_n}$. Hence the number of partitions of $[n]$ of type $(r_1,\ldots,r_n)$ is

$$\frac{n!}{(1!)^{r_1}(2!)^{r_2}\cdots(n!)^{r_n}r_1!\cdots r_n!}.$$

4. (*i*) Write

$$\log\Big(1+\sum_{n\ge 1}\alpha_n\frac{z^n}{n!}\Big)=\sum_{m\ge 1}\beta_m\frac{z^m}{m!}. \tag{12.1}$$

Then by differentiating both sides and multiplying by $1+\sum_{n\ge 1}\alpha_n\frac{z^n}{n!}$ we have

$$\sum_{n\ge 1}\alpha_n\frac{z^{n-1}}{(n-1)!}=\sum_{m\ge 1}\beta_m\frac{z^{m-1}}{(m-1)!}\Big(1+\sum_{n\ge 1}\alpha_n\frac{z^n}{n!}\Big)$$

and by reindexing

$$\sum_{n\ge 0}\alpha_{n+1}\frac{z^n}{n!}=\sum_{m\ge 0}\beta_{m+1}\frac{z^m}{m!}\Big(1+\sum_{n\ge 1}\alpha_n\frac{z^n}{n!}\Big).$$

Next let us expand the right-hand side. For convenience of notation we let $\alpha_0=1$.

$$\begin{aligned}
\sum_{m\ge 0}\beta_{m+1}\frac{z^m}{m!}\Big(1+\sum_{n\ge 1}\alpha_n\frac{z^n}{n!}\Big) &= \sum_{m\ge 0}\beta_{m+1}\frac{z^m}{m!}+\sum_{m\ge 0}\sum_{n\ge 1}\beta_{m+1}\alpha_n\frac{z^{m+n}}{m!n!}\\
&=\sum_{m\ge 0}\beta_{m+1}\frac{z^m}{m!}+\sum_{N\ge 1}\Bigg[\sum_{\substack{m\ge 0,n\ge 1\\ m+n=N}}\binom{N}{m}\beta_{m+1}\alpha_n\Bigg]\frac{z^N}{N!}\\
&=\sum_{N\ge 0}\beta_{N+1}\frac{z^N}{N!}+\sum_{N\ge 1}\Bigg[\sum_{m=0}^{N-1}\binom{N}{m}\beta_{m+1}\alpha_{N-m}\Bigg]\frac{z^N}{N!}\\
&=\sum_{N\ge 0}\Bigg[\sum_{m=0}^{N}\binom{N}{m}\beta_{m+1}\alpha_{N-m}\Bigg]\frac{z^N}{N!}.
\end{aligned}$$

Thus (12.1) is equivalent to

$$\alpha_n=\sum_{m=0}^{n-1}\binom{n-1}{m}\beta_{m+1}\alpha_{n-m-1}. \tag{12.2}$$

(*ii*) Now let us start with the equation $\alpha_n = \sum_{\pi \in \mathcal{P}(n)} k_\pi$. We shall show that this implies that

$$\alpha_n = \sum_{m=0}^{n-1} \binom{n-1}{m} k_{m+1}\alpha_{n-m-1}. \tag{12.3}$$

We shall adopt the following notation; given $\pi \in \mathcal{P}(n)$ we let V_1 denote the block of π containing 1.

$$\begin{aligned}\sum_{\pi \in \mathcal{P}(n)} k_\pi &= \sum_{m=0}^{n-1} \sum_{\substack{\pi \in \mathcal{P}(n) \\ |V_1|=m+1}} k_\pi \\ &= \sum_{m=0}^{n-1} k_{m+1} \binom{n-1}{m} \sum_{\sigma \in \mathcal{P}(n-m-1)} k_\sigma \\ &= \sum_{m=0}^{n-1} \binom{n-1}{m} k_{m+1}\alpha_{n-m-1}.\end{aligned}$$

Where the second inequality follows because there are $\binom{n-1}{m}$ ways to choose the m elements of $\{2, 3, 4, \ldots, n\}$ needed to make a block of size $m+1$ containing 1, and then σ, what remains of π after V_1 is removed, is a partition of the remaining $n-m-1$ elements.

(*iii*) Since $k_1 = \alpha_1 = \beta_1$ we can use equations (12.2) and (12.3) and induction to conclude that $\beta_n = k_n$ for all n.

5. (*i*) First note that for a Gaussian random vector, as we have defined it, the entries are centred, i.e.

$$\mathrm{E}(X_i) = \int_{\mathbb{R}^n} t_i \frac{\exp(-\langle Bt, t\rangle/2)}{(2\pi)^{n/2}\det(B)^{-1/2}}\, dt = 0$$

as the integrand is odd. Let $\sigma_i^2 = \mathrm{E}(X_i^2)$ be the variance of X_i.

If $\{X_1, \ldots, X_n\}$ are independent, then the joint distribution of $\{X_1, \ldots, X_n\}$ is

$$\frac{e^{-t_1^2/(2\sigma_1^2)}}{\sqrt{2\pi\sigma_1^2}} \cdots \frac{e^{-t_n^2/(2\sigma_n^2)}}{\sqrt{2\pi\sigma_n^2}} \times dt_1 \cdots dt_n = \frac{\exp(-\langle Bt, t\rangle/2)}{(2\pi)^{n/2}\sqrt{\sigma_1^2 \cdots \sigma_n^2}}\, dt$$

where B is the diagonal matrix with diagonal entries $\sigma_1^{-2}, \ldots, \sigma_n^{-2}$.

Conversely suppose that B is diagonal with diagonal entries $\sigma_1^{-2}, \ldots, \sigma_n^{-2}$. Then the density is the product:

$$\frac{e^{-t_1^2/(2\sigma_1^2)}}{\sqrt{2\pi\sigma_1^2}} \cdots \frac{e^{-t_n^2/(2\sigma_n^2)}}{\sqrt{2\pi\sigma_n^2}} \times dt_1 \cdots dt_n$$

and so $\{X_1, \ldots, X_n\}$ are independent.

(*ii*) Let $C = B^{-1}$. As noted above, when $\{X_1, \dots, X_n\}$ are independent $B_{ij} = \delta_{ij}\sigma_i^{-2} = (C^{-1})_{ij}$. So the result holds for independent X_i's.

B is a positive definite real symmetric matrix, so there is an orthogonal matrix O such that $D = O^{-1}BO$ is diagonal. Let $Y = O^{-1}X$. Write $O^{-1} = (p_{ij})$ and $s = O^{-1}t$ or $t = Os$. Then $dt = ds$ by the orthogonality of O.

$$\begin{aligned}
\mathrm{E}(Y_{i_1} \cdots Y_{i_k}) &= \sum_{j_1,\dots,j_k=1}^{n} p_{i_1 j_1} \cdots p_{i_k j_k} \mathrm{E}(X_{j_1} \cdots X_{j_k}) \\
&= \sum_{j_1,\dots,j_k=1}^{n} p_{i_1 j_1} \cdots p_{i_k j_k} \int_{\mathbb{R}^n} t_{j_1} \cdots t_{j_k} \frac{\exp(-\langle Bt, t\rangle/2)}{(2\pi)^{n/2} \det(B)^{-1/2}}\, dt \\
&= \int_{\mathbb{R}^n} s_{i_1} \cdots s_{i_k} \frac{\exp(-\langle BOs, Os\rangle/2)}{(2\pi)^{n/2} \det(B)^{-1/2}}\, ds \\
&= \int_{\mathbb{R}^n} s_{i_1} \cdots s_{i_k} \frac{\exp(-\langle Ds, s\rangle/2)}{(2\pi)^{n/2} \det(D)^{-1/2}}\, ds.
\end{aligned}$$

Thus $\{Y_1, \dots, Y_n\}$ are independent and Gaussian. Hence $\mathrm{E}(Y_i Y_j) = (D^{-1})_{ij}$. Thus

$$\begin{aligned}
c_{ij} &= \mathrm{E}(X_i X_j) = \sum_{k,l=1}^{n} o_{ik} o_{jl} \mathrm{E}(Y_k Y_l) = \sum_{k,l=1}^{n} o_{ik} o_{jl} (D^{-1})_{kl} \\
&= \sum_{k,l=1}^{n} o_{ik} (D^{-1})_{kl} o_{lj} = (OD^{-1}O^{-1})_{ij} = (B^{-1})_{ij}.
\end{aligned}$$

6. (*i*) We have

$$C = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}, \qquad \text{so} \qquad B = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}.$$

So the first claim follows from the formula for the density, and the second from the usual conversion to polar coordinates.

(*ii*) Note that the integral in polar coordinates factors as an integral over θ and one over r. Thus for any θ

$$\begin{aligned}
&\int_{\mathbb{R}^2} (t_1 + it_2)^m (t_1 - it_2)^n e^{-(t_1^2+t_2^2)}\, dt_1 dt_2 \\
&\qquad = e^{i\theta(m-n)} \int_{\mathbb{R}^2} (t_1 + it_2)^m (t_1 - it_2)^n e^{-(t_1^2+t_2^2)}\, dt_1 dt_2.
\end{aligned}$$

Hence

$$\mathrm{E}(Z^m \overline{Z}^n) = \int_{\mathbb{R}^2} (t_1 + it_2)^m (t_1 - it_2)^n e^{-(t_1^2+t_2^2)}\, dt_1 dt_2 = 0 \qquad \text{for } m \neq n.$$

Furthermore, we have

$$\mathrm{E}(|Z|^{2n}) = \frac{1}{\pi}\int_{\mathbb{R}^2}(t_1^2+t_2^2)^n e^{-(t_1^2+t_2^2)}\,dt_1dt_2 = \frac{1}{\pi}\int_0^{2\pi}\int_0^\infty r^{2n}e^{-r^2}\,r\,dr\,d\theta$$
$$= \int_0^\infty r^{2n}\,d(-e^{-r^2}) = n\int_0^\infty r^{2(n-1)}\,d(-e^{-r^2}) = \cdots = n!.$$

7. We have seen that $\mathrm{E}(Z_{i_1}\cdots Z_{i_n}\overline{Z_{j_1}}\cdots\overline{Z_{j_n}})$ is the number of pairings π of $[2n]$ such that for each pair (r,s) of π (with $r<s$) we have that $r\le n$ and $n+1\le s\le 2n$ and $i_r = j_{s-n}$. For such a π let σ be the permutation with $\sigma(r)=s-n$; we then have $i = j\circ\sigma$. Conversely let σ be a permutation of $[n]$ with $i=j\circ\sigma$. Let π be the pairing with pairs $(r, n+\sigma(r))$; then $i_r = j_{s-n}$ for $s = n+\sigma(r)$.

8. $\mathrm{E}(|f_{ij}|^2) = 1/N$, so $E(x_{ii}^2) = 1/N$ and for $i\ne j$, $\mathrm{E}(x_{ij}^2) = \mathrm{E}(y_{ij}^2) = 1/(2N)$. Thus the covariance matrix C is the $N^2\times N^2$ diagonal matrix with diagonal entries $(1/N,\dots,1/N,1/(2N),\dots,1/(2N))$ (here the entry $1/N$ appears N times). Thus the density matrix B is the diagonal matrix with diagonal entries $(N,\dots,N,2N,\dots,2N)$. Hence

$$\begin{aligned}\langle BX,X\rangle &= N\Big(\sum_{i=1}^N x_{ii}^2 + 2\big(\sum_{1\le i<j\le N}(x_{ij}^2+y_{ij}^2)\big)\Big)\\ &= N\Big(\sum_{i=1}^N x_{ii}^2 + \sum_{\substack{1\le i,j\le N\\ i\ne j}}(x_{ij}^2+y_{ij}^2)\Big)\\ &= N\Big(\sum_{i=1}^N x_{ii}^2 + \sum_{\substack{1\le i,j\le N\\ i\ne j}}(x_{ij}+\sqrt{-1}y_{ij})(x_{ij}-\sqrt{-1}y_{ij})\Big)\\ &= N\,\mathrm{Tr}(X^2).\end{aligned}$$

Thus $\exp(-\langle BX,X\rangle/2) = \exp(-N\mathrm{Tr}(X^2)/2)$. Next $\det(B) = N^{N^2}2^{N^2-N}$. Thus

$$c = \Big(\frac{N}{\pi}\Big)^{N^2/2}\Big(\frac{1}{2}\Big)^{N/2}.$$

9. Note that $(\mathcal{A}_1\vee\mathcal{A}_2)\ominus\mathcal{A}_1\subset\ker\varphi$ because $\mathcal{A}_1$ is unital. By the non-degeneracy of φ, $\mathring{\mathcal{A}}_1\cap((\mathcal{A}_1\vee\mathcal{A}_2)\ominus\mathcal{A}_1) = \{0\}$. So by equation (1.11), the left-hand side of (1.12) is contained in the right-hand side. To prove the reverse containment, let $a_1\cdots a_n\in\mathring{\mathcal{A}}_{\alpha_1}\cdots\mathring{\mathcal{A}}_{\alpha_n}$ for some $\alpha_1\ne\cdots\ne\alpha_n$. Let $a\in\mathcal{A}_1$; for $\alpha_n\ne 1$; we have $\varphi(a_1\cdots a_n a) = \varphi(a_1\cdots a_n\mathring{a}) + \varphi(a_1\cdots a_n)\varphi(a) = 0$ by freeness, and if $\alpha_n = 1$ we have $\varphi(a_1\cdots a_n a) = \varphi(a_1\cdots a_{n-1}(a_n a)^\circ) + \varphi(a_1\cdots a_{n-1})\varphi(a_n a) = 0$, again by freeness. Thus $a_1\cdots a_n\in(\mathcal{A}_1\vee\mathcal{A}_2)\ominus\mathcal{A}_1$.

10. (*i*) Let $\sum_{n=1}^{\infty} \beta_n z^n$ be a formal power series. Using the series for exp, we have

$$\begin{aligned}
\exp\Big(\sum_{n=1}^{\infty} \beta_n z^n\Big) &= \sum_{n=0}^{\infty} \frac{1}{n!}\Big(\sum_{l=1}^{\infty} \beta_l z^l\Big)^n \\
&= 1 + \sum_{n=1}^{\infty} \frac{1}{n!} \sum_{l_1=1}^{\infty} \cdots \sum_{l_n=1}^{\infty} \beta_{l_1} \cdots \beta_{l_n} z^{l_1+\cdots+l_n} \\
&= 1 + \sum_{n=1}^{\infty} \Big(\sum_{m=1}^{n} \sum_{\substack{l_1,\ldots,l_m \geq 1 \\ l_1+\cdots+l_m=n}} \frac{\beta_{l_1}\cdots\beta_{l_m}}{m!} \Big) z^n.
\end{aligned}$$

(*ii*) We continue from the solution to (*i*). First we shall work with the sum

$$S = \sum_{m=1}^{n} \sum_{\substack{l_1,\ldots,l_m \geq 1 \\ l_1+\cdots+l_m=n}} \frac{\beta_{l_1}\cdots\beta_{l_m}}{m!}.$$

We are summing over all tuples $l = (l_1, \ldots, l_m)$ of positive integers such that $l_1 + \cdots + l_m = n$, i.e. over all compositions of the integer n. By the type of the composition $l = (l_1, \ldots, l_n)$, we mean the n-tuple $r = (r_1, \ldots, r_n)$ where the r_i's are integers, $r_i \geq 0$, and r_i is the number of l_j's that equal i. We must have $1 \cdot r_1 + 2 \cdot r_2 + \cdots + n \cdot r_n = n$, and $m = r_1 + \cdots + r_n$ is the number of parts of $l = (l_1, \ldots, l_m)$. Note that $\beta_{l_1} \cdots \beta_{l_m} = \beta_1^{r_1} \cdots \beta_n^{r_n}$ depends only on the type of $l = (l_1, \ldots, l_m)$. Hence we can group the compositions by their type and thus S becomes

$$S = \sum_{1r_1+\cdots+nr_n=n} \frac{\beta_1^{r_1}\cdots\beta_n^{r_n}}{(r_1+\cdots+r_n)!} \times \text{no. compositions of } n \text{ of type } (r_1, \ldots, r_n).$$

Given a type $r = (r_1, \ldots, r_n)$, there are $r_1 + \cdots + r_n$ parts which can be permuted in $(r_1 + \cdots + r_n)!$ ways; however, we don't distinguish between permutations that change l_i's which are equal; thus, we must divide by $r_1!r_2!\cdots r_n!$. Hence the number of compositions of the integer n of type $(r_1, \ldots, r_n)$ is

$$\frac{(r_1+\cdots+r_n)!}{r_1!r_2!\cdots r_n!}$$

thus

$$S = \sum_{1r_1+\cdots+nr_n=n} \frac{\beta_1^{r_1}\cdots\beta_n^{r_n}}{r_1!r_2!\cdots r_n!}.$$

Hence

$$\exp\Big(\sum_{n=1}^{\infty}\beta_n z^n\Big) = 1 + \sum_{n=1}^{\infty} \sum_{\substack{r_1,\ldots,r_n\geq 0\\ 1r_1+\cdots+nr_n=n}} \frac{\beta_1^{r_1}\cdots\beta_n^{r_n}}{r_1!r_2!\cdots r_n!} z^n.$$

By replacing β_n by $\frac{k_n}{n!}$ we obtain the equation

$$\sum_{n=0}^{\infty} \sum_{\substack{r_1,\ldots,r_n\geq 0\\ 1\cdot r_1+\cdots+n\cdot r_n=n}} \frac{n!}{(1!)^{r_1}\cdots(n!)^{r_n}r_1!r_2!\cdots r_n!} k_1^{r_1}\cdots k_n^{r_n}\frac{z^n}{n!} = \exp\Big(\sum_{n=1}^{\infty} k_n\frac{z^n}{n!}\Big).$$

Then we compare this with the defining equation

$$\log\Big(1+\sum_{n\geq 1}\alpha_n\frac{z^n}{n!}\Big) = \sum_{n\geq 1} k_n\frac{z^n}{n!}$$

to conclude that equation (1.1) holds.

11. If we replace the ordinary generating function $\sum_{n\geq 1}\beta_n z^n$ by the exponential generating function $\sum_{n\geq 1}\beta_n z^n/(n!)$, we get from Exercise 10 (*ii*)

$$\begin{aligned}\exp\Big(\sum_{n=1}^{\infty}\frac{\beta_n}{n!} z^n\Big) &= 1 + \sum_{n=1}^{\infty} \sum_{\substack{r_1,\ldots,r_n\geq 0\\ 1r_1+\cdots+nr_n=n}} \frac{\beta_1^{r_1}\cdots\beta_n^{r_n}}{(1!)^{r_1}\cdots(n!)^{r_n}r_1!r_2!\cdots r_n!} z^n \\ &= 1 + \sum_{n=1}^{\infty} \sum_{\substack{r_1,\ldots,r_n\geq 0\\ 1r_1+\cdots+nr_n=n}} \frac{n!}{(1!)^{r_1}\cdots(n!)^{r_n}r_1!r_2!\cdots r_n!} \beta_1^{r_1}\cdots\beta_n^{r_n}\frac{z^n}{n!}.\end{aligned}$$

From Exercise 3 we know

$$\frac{n!}{(1!)^{r_1}\cdots(n!)^{r_n}r_1!r_2!\cdots r_n!}$$

counts the number of partitions of the set $[n]$ of type $(r_1,\ldots,r_n)$. If $\pi = \{V_1,\ldots,V_m\}$ is a partition of $[n]$, we let $\beta_\pi = \beta_{|V_1|}\beta_{|V_2|}\cdots\beta_{|V_m|}$ where $|V_i|$ is the number of elements in the block V_i. If the type of the partition π is $(r_1,\ldots,r_n)$, then $\beta_1^{r_1}\beta_2^{r_2}\cdots\beta_n^{r_n} = \beta_\pi$. Thus we can write

$$\exp\Big(\sum_{n=1}^{\infty}\frac{\beta_n}{n!} z^n\Big) = 1 + \sum_{n=1}^{\infty}\Big(\sum_{\pi\in\mathcal{P}(n)}\beta_\pi\Big)\frac{z^n}{n!}.$$

12. Using $\log(1-x) = -\sum_{n\geq 1} x^n/n$ we have

$$-\log(1-\sum_{n=1}^{\infty}\beta_n z^n) = \sum_{n=1}^{\infty}\frac{1}{n}\Big(\sum_{l=1}^{\infty}\beta_l z^l\Big)^n$$
$$= \sum_{n=1}^{\infty}\frac{1}{n}\sum_{l_1=1}^{\infty}\cdots\sum_{l_n=1}^{\infty}\beta_{l_1}\cdots\beta_{l_n}z^{l_1+\cdots+l_n}$$
$$= \sum_{m=1}^{\infty}\sum_{n=1}^{m}\frac{1}{n}\sum_{\substack{l_1,\ldots,l_n\geq 1\\ l_1+\cdots+l_n=m}}\beta_{l_1}\cdots\beta_{l_n}z^m$$
$$= \sum_{n=1}^{\infty}\sum_{m=1}^{n}\frac{1}{m}\sum_{\substack{l_1,\ldots,l_m\geq 1\\ l_1+\cdots+l_m=n}}\beta_{l_1}\cdots\beta_{l_m}z^n.$$

Now let S be the sum

$$S = \sum_{m=1}^{n}\frac{1}{m}\sum_{\substack{l_1,\ldots,l_m\geq 1\\ l_1+\cdots+l_m=n}}\beta_{l_1}\cdots\beta_{l_m}.$$

As with the exponential, this is a sum over all compositions of the integer n, so we group the terms according to their type, as was done in the solution to Exercise 10.

$$S = \sum_{1r_1+\cdots+nr_n=n}\frac{\beta_1^{r_1}\cdots\beta_n^{r_n}}{r_1+\cdots+r_n}\times \text{no. compositions of } n \text{ of type } (r_1,\ldots,r_n)$$
$$= \sum_{1r_1+\cdots+nr_n=n}\beta_1^{r_1}\cdots\beta_n^{r_n}\frac{(r_1+\cdots+r_n-1)!}{r_1!\cdots r_n!}.$$

Putting this in the equation for $-\log(1-\sum_{n\geq 1}\beta_n z^n)$, we get

$$-\log(1-\sum_{n\geq 1}\beta_n z^n) = \sum_{n=1}^{\infty}\sum_{1r_1+\cdots+nr_n=n}(r_1+\cdots+r_n-1)!\frac{\beta_1^{r_1}\cdots\beta_n^{r_n}}{r_1!\cdots r_n!}z^n.$$

Replacing β_n by $-\beta_n$ we obtain

$$\log(1+\sum_{n\geq 1}\beta_n z^n) = \sum_{n=1}^{\infty}\sum_{1r_1+\cdots+nr_n=n}(-1)^{r_1+\cdots+r_n-1}(r_1+\cdots+r_n-1)!\frac{\beta_1^{r_1}\cdots\beta_n^{r_n}}{r_1!\cdots r_n!}z^n.$$

13. (*i*) We replace β_n with $\alpha_n/(n!)$ in Exercise 12 to obtain

$$\log(1+\sum_{n\geq 1}\frac{\alpha_n}{n!}z^n)$$
$$=\sum_{n=1}^{\infty}\sum_{1r_1+\cdots+nr_n=n}(-1)^{r_1+\cdots+r_n-1}(r_1+\cdots+r_n-1)!\frac{\alpha_1^{r_1}\cdots\alpha_n^{r_n}\,n!}{(1!)^{r_1}\cdots(n!)^{r_n}\,r_1!\cdots r_n!}\frac{z^n}{n!}.$$

We then turn this into a sum over partitions recalling that

$$\frac{n!}{(1!)^{r_1}\cdots(n!)^{r_n}\,r_1!\cdots r_n!}$$

is the number of partitions of $[n]$ of type $(r_1,\ldots,r_n)$, and if π is a partition of $[n]$, we denote by $\#(\pi)$ the number of blocks of π. Then as

$$(-1)^{r_1+\cdots+r_n-1}(r_1+\cdots+r_n-1)!\alpha_1^{r_1}\cdots\alpha_n^{r_n}=(-1)^{\#(\pi)-1}(\#(\pi)-1)!\,\alpha_\pi$$

only depends on the type of π, we have

$$\log(1+\sum_{n=1}^{\infty}\frac{\alpha_n}{n!}z^n)=\sum_{n=1}^{\infty}\sum_{\pi\in\mathcal{P}(n)}(-1)^{\#(\pi)-1}(\#(\pi)-1)!\,\alpha_\pi\frac{z^n}{n!}. \qquad (12.4)$$

(*ii*) Note that α_n only appears once in

$$k_n=\sum_{\pi\in\mathcal{P}(n)}(-1)^{\#(\pi)-1}(\#(\pi)-1)!\,\alpha_\pi$$

so each of the sequences $\{\alpha_n\}_n$ and $\{k_n\}_n$ determines the other. Thus we may write the result of (*i*) as

$$\sum_{n=1}^{\infty}k_n\frac{z^n}{n!}=\log\Big(1+\sum_{n=1}^{\infty}\alpha_n\frac{z^n}{n!}\Big).$$

On the other hand, replacing the sequence $\{\beta_n\}_n$ by $\{k_n\}_n$ in Exercise 11, we have

$$1+\sum_{n=1}^{\infty}\alpha_n\frac{z^n}{n!}=\exp\Big(\sum_{n=1}^{\infty}k_n\frac{z^n}{n!}\Big)=1+\sum_{n=1}^{\infty}\Big(\sum_{\pi\in\mathcal{P}(n)}k_\pi\Big)\frac{z^n}{n!},$$

and so we get the other half of the moment-cumulant relation

$$\alpha_n=\sum_{\pi\in\mathcal{P}(n)}k_\pi.$$

14. Since ν has moments of all orders, φ, the characteristic function of ν, has derivatives of all orders. Fix $n > 0$. We may write

$$\varphi(t) = 1 + \sum_{r=1}^{n} \alpha_r \frac{s^r}{r!} + o(s^n)$$

where $s = it$ and α_r is the r^{th} moment of ν. We can also write

$$\log(1+z) = \sum_{r=1}^{n} (-1)^{r+1} \frac{z^r}{r} + o(z^n).$$

Now for $l \geq 1$

$$\Big(\sum_{r=1}^{n} \alpha_r \frac{s^r}{r!} + o(s^n) \Big)^l = \Big(\sum_{r=1}^{n} \alpha_r \frac{s^r}{r!} \Big)^l + o(s^n).$$

Thus

$$\log(\varphi(t)) = \sum_{l=1}^{n} \frac{(-1)^{l+1}}{l} \Big(\sum_{r=1}^{n} \alpha_r \frac{s^r}{r!} \Big)^l + o(s^n)$$

and hence

$$\sum_{l=1}^{n} k_l \frac{s^l}{l!} + o(s^n) = \sum_{l=1}^{n} \frac{(-1)^{l+1}}{l} \Big(\sum_{r=1}^{n} \alpha_r \frac{s^r}{r!} \Big)^l + o(s^n).$$

By Exercise 12 we have

$$k_n = \sum_{\pi \in \mathcal{P}(n)} (-1)^{\#(\pi)-1} (\#(\pi)-1)!\, \alpha_\pi$$

and

$$\alpha_n = \sum_{\pi \in \mathcal{P}(n)} k_\pi.$$

12.2 Solutions to exercises in Chapter 2

8. (*i*) This follows from applying a cyclic rotation to the moment-cumulant formula and observing that non-crossing partitions are mapped to non-crossing partitions under rotations.

(*ii*) This is not true, since the property non-crossing is not preserved under arbitrary permutations. For example, in the calculation of $\kappa_4(a_1, a_2, a_3, a_4)$, the crossing term $\varphi(a_1a_3)\varphi(a_2a_4)$ does not show up. However, in $\kappa_4(a_1, a_3, a_2, a_4)$ this term becomes non-crossing and will make a contribution. Hence $\kappa_4(a_1, a_2, a_3, a_4) \neq \kappa_4(a_1, a_3, a_2, a_4)$ in general, even if all a_i commute.

9. For the semi-circle law we have that all odd moments are 0 and the $2k^{th}$ moment is the k^{th} Catalan number $\frac{1}{k+1}\binom{2k}{k}$ which is also the cardinality of $NC_2(2k)$, the non-crossing pairings of $[2k]$. Since $\alpha_1 = 0$ we have $\kappa_1 = 0$; and since $\alpha_2 = \kappa_1^2 + \kappa_2$ we have $\kappa_2 = \alpha_2 = 1$. Now let $NC^*(n)$ be the set of non-crossing permutations which are not pairings. For $n = 2k$ we have

$$\alpha_n = \sum_{\pi \in NC(n)} \kappa_\pi = \sum_{\pi \in NC_2(n)} \kappa_\pi + \sum_{\pi \in NC^*(n)} \kappa_\pi = \alpha_n + \sum_{\pi \in NC^*(n)} \kappa_\pi .$$

Thus, for n even $\sum_{\pi \in NC^*(n)} \kappa_\pi = 0$ and also for n odd because there are no pairings of $[n]$. When $n = 3$, this forces $\kappa_3 = 0$. Then for general n we write

$$0 = \sum_{\pi \in NC^*(n)} \kappa_\pi = \kappa_n + \sum_{\pi \in NC^{**}(n)} \kappa_\pi ,$$

where $NC^{**}(n)$ is all the partitions in $NC^*(n)$ with more than one block. By induction $\sum_{\pi \in NC^{**}(n)} \kappa_\pi = 0$; so $\kappa_n = 0$ for $n \geq 3$.

11. (*iv*) We have

$$\sum_{\pi \in NC(n)} c^{\#(\pi)} = \alpha_n = \sum_{\pi \in NC(n)} \kappa_\pi . \tag{12.5}$$

When $n = 1$, this gives $\kappa_1 = c$. If we have shown that $\kappa_1 = \cdots = \kappa_{n-1} = c$, then

$$\sum_{\pi \in NC^{**}(n)} \kappa_\pi = \sum_{\pi \in NC^{**}(n)} c^{\#(\pi)}$$

where $NC^{**}(n)$ is all non-crossing partitions of $[n]$ with more than one block. Thus (12.5) shows that $\kappa_n = c$.

14. We have

$$\begin{aligned}
\omega_a(z) + \omega_b(z) &= 2z - \big(R_a(G_{a+b}(z)) + R_b(G_{a+b}(z))\big) \\
&= 2z - R_{a+b}(G_{a+b}(z)) \\
&= 2z - (z - 1/G_{a+b}(z)) \\
&= z + 1/G_{a+b}(z) \\
&= z + 1/G_a(\omega_a(z)).
\end{aligned}$$

15. By inverting the first equation in (2.32), we have $\omega_a(G^{\langle -1\rangle}(z)) = G_a^{\langle -1\rangle}(z)$ and $\omega_b(G^{\langle -1\rangle}(z)) = G_b^{\langle -1\rangle}(z)$. By the second equation in (2.32), we have

$$\begin{aligned} R(z) + 1/z &= G^{\langle -1\rangle}(z) \\ &= \omega_a(G^{\langle -1\rangle}(z)) + \omega_b(G^{\langle -1\rangle}(z)) - 1/G_a(\omega_a(G^{\langle -1\rangle}(z))) \\ &= G_a^{\langle -1\rangle}(z) + G_b^{\langle -1\rangle}(z) - 1/G_a(G_a^{\langle -1\rangle}(z)) \\ &= R_a(z) + 1/z + R_b(z) + 1/z - 1/z. \end{aligned}$$

Hence $R(z) = R_a(z) + R_b(z)$.

17. (i) Let $a_2 \in \mathring{\mathcal{A}}_2$ and $a_1 \in \mathcal{A}_1$. Then $\varphi(a_1 \mathrm{E}_x[a_2]) = \varphi(a_1 a_2) = \varphi(a_1)\varphi(a_2) = 0$, by freeness. Thus $\mathrm{E}_x[a_2] = 0$.

(ii) Let $a_1 \cdots a_n \in \mathring{\mathcal{A}}_{\alpha_1} \cdots \mathring{\mathcal{A}}_{\alpha_n}$ and $a \in \mathcal{A}_1$. First suppose $\alpha_1 \neq 1$. Then

$$\varphi(a\mathrm{E}_x[a_1 \cdots a_n]) = \varphi(aa_1 \cdots a_n) = \varphi(\mathring{a}a_1 \cdots a_n) + \varphi(a)\varphi(a_1 \cdots a_n) = 0$$

by freeness; thus, $\mathrm{E}_x[a_1 \cdots a_n] = 0$. If $\alpha_1 = 1$; then we write

$$\varphi(a\mathrm{E}_x[a_1 \cdots a_n]) = \varphi((aa_1)a_2 \cdots a_n) = \varphi((\mathring{aa_1})a_2 \cdots a_n) + \varphi(aa_1)\varphi(a_2 \cdots a_n) = 0$$

by freeness, and hence again $\mathrm{E}_x[a_1 \cdots a_n] = 0$.

18. Let $p(x, y) \in \mathbb{C}\langle x, y\rangle$ be given; we must show that using the definition of E_x given in the exercise, we have that Equation (2.36) holds for all $q(x) \in \mathbb{C}\langle x\rangle$, i.e. $\varphi(q(x)p(x, y)) = \varphi(q(x)\mathrm{E}_x[p(x, y)])$. This equation is linear in p so we only need to check it for p in each of the summands of the decomposition of $\mathcal{A}_1 \vee \mathcal{A}_2$. It is immediate for $p \in \mathcal{A}_1$. It is then an easy consequence of freeness that $\varphi(q(x)p(x, y)) = 0$ for p in any other of the summands.

19. (i) Recall the definition of $\tilde{\varphi}_\pi$. We have for $\pi = \{V_1, \ldots, V_s\}$ with $n \in V_s$,

$$\tilde{\varphi}_\pi(a_1, a_2, \ldots, a_n) = \varphi\Big(\prod_{i_1 \in V_1} a_{i_1}\Big) \cdots \varphi\Big(\prod_{i_{s-1} \in V_{s-1}} a_{i_{s-1}}\Big) \prod_{i_s \in V_s} a_{i_s}$$

so

$$\varphi(a_0\tilde{\varphi}_\pi(a_1, a_2, \ldots, a_n)) = \varphi\Big(\prod_{i_1 \in V_1} a_{i_1}\Big) \cdots \varphi\Big(\prod_{i_{s-1} \in V_{s-1}} a_{i_{s-1}}\Big) \varphi\Big(a_0 \prod_{i_s \in V_s} a_{i_s}\Big).$$

Now the right-hand side is exactly $\varphi_{\pi'}(a_0, a_1, a_2, \ldots, a_n)$ where π' is the non-crossing partition obtained by adding 0 to the block V_s of π containing n.

(ii) For the purposes of this solution, we shall introduce the following notation. Let $[\bar{n}] = \{\bar{1}, \bar{2}, \ldots, \bar{n}\}$. Let $[\bar{n}'] = \{\bar{0}, \bar{1}, \bar{2}, \ldots, \bar{n}\}$ and $[2n] = \{1, \bar{1}, 2, \bar{2}, \ldots, n, \bar{n}\}$ and $[2n'] = \{\bar{0}, 1, \bar{1}, \ldots, n, \bar{n}\}$. Let $\sigma \in NC(2n')$; since $x_0, x_1, \ldots, x_n$ are free from $y_1, \ldots, y_n$, we have that $\kappa_\sigma(x_0, y_1, x_1, \ldots, y_n, x_n) = 0$ unless we can write

$\sigma = \pi \cup \tau$ with $\pi \in NC(n)$ and $\tau \in NC(\bar{n}')$. Let us recall the definition of the Kreweras complement from section 2.3. For π a non-crossing partition of $[n]$, $K(\pi)$ is the largest non-crossing partition of $[\bar{n}]$ so that $\pi \cup K(\pi)$ is a non-crossing partition of $[2n]$. Thus $K(\pi)'$ is the largest non-crossing partition of $[\bar{n}']$ such that $\pi \cup K(\pi)'$ is a non-crossing partition of $[2n']$. Thus for $\pi \in NC(n)$ and $\tau \in NC(\bar{n}')$ we have that $\pi \cup \tau$ is a non-crossing partition of $[2n']$ if and only if $\tau \leq K(\pi)'$. Thus

$$\begin{aligned}
\varphi(x_0 y_1 x_1 \cdots y_n x_n) &= \sum_{\sigma \in NC(2n')} \kappa_\sigma(x_0, y_1, x_1, \dots, y_n, x_n) \\
&= \sum_{\pi \in NC(n)} \kappa_\pi(y_1, \dots, y_n) \sum_{\substack{\tau \in NC(\bar{n}') \\ \pi \cup \tau \in NC(2n')}} \kappa_\tau(x_0, x_1, \dots, x_n) \\
&= \sum_{\pi \in NC(n)} \kappa_\pi(y_1, \dots, y_n) \sum_{\substack{\tau \in NC(\bar{n}') \\ \tau \leq K(\pi)'}} \kappa_\tau(x_0, x_1, \dots, x_n) \\
&= \sum_{\pi \in NC(n)} \kappa_\pi(y_1, \dots, y_n) \varphi_{K(\pi)'}(x_0, x_1, \dots, x_n) \\
&= \sum_{\pi \in NC(n)} \kappa_\pi(y_1, \dots, y_n) \varphi(x_0 \tilde{\varphi}_{K(\pi)}(x_1, \dots, x_n)) \\
&= \varphi\left(x_0 \sum_{\pi \in NC(n)} \kappa_\pi(y_1, \dots, y_n) \tilde{\varphi}_{K(\pi)}(x_1, \dots, x_n) \right).
\end{aligned}$$

Hence by the non-degeneracy of φ, we have

$$\sum_{\pi \in NC(n)} \kappa_\pi(y_1, \dots, y_n) \tilde{\varphi}_{K(\pi)}(x_1, \dots, x_n) = \mathrm{E}_x(y_1 x_1 \cdots y_n x_n).$$

12.3 Solutions to exercises in Chapter 3

1. (*i*) Let δ_a be the probability measure with an atom of mass 1 at a. Then $\int 1/(z-t)\, d\delta_a(t) = 1/(z-a)$. We have

$$\mu = \sum_{i=1}^n \lambda_i \delta_{a_i}, \qquad \text{thus} \qquad G(z) = \sum_{i=1}^n \frac{\lambda_i}{z - a_i}.$$

(*ii*) Fix $z \in \mathbb{C}^+$. Let

$$f(w) = \frac{1}{\pi(z-w)(w-i)(w+i)}, \qquad \text{then} \qquad G(z) = \int_{-\infty}^{\infty} f(t)\, dt.$$

Since f is a rational function such that $\lim_{w\to\infty} wf(w) = 0$ and by the residue theorem we have

$$G(z) = \int_{-\infty}^{\infty} f(t)\,dt = \lim_{R\to\infty} \int_{C_R} f(w)\,dw = 2\pi i(\operatorname{Res}(f, z) + \operatorname{Res}(f, i)),$$

where C_R is the closed curve formed by joining part of the circle $|w| = R$ in $\mathbb{C}^+$ to the interval $[-R, R]$.

$$\operatorname{Res}(f, z) = \frac{-1}{\pi(z-i)(z+i)} \qquad \text{and} \qquad \operatorname{Res}(f, i) = \frac{1}{2\pi i(z-i)}.$$

Thus $G(z) = 1/(z+i)$.

4. (*iii*) We know that $w_1 w_2 = 1$, so one of $\{w_1, w_2\}$ is inside Γ, and the other is outside Γ. Let us show that $|w_1| < |w_2|$ by showing that $|\operatorname{Re}(w_1)| \le |\operatorname{Re}(w_2)|$ and $|\operatorname{Im}(w_1)| < |\operatorname{Im}(w_2)|$. Suppose $\operatorname{Re}(z) > 0$, the case $\operatorname{Re}(z) \le 0$ can be handled similarly. Then $\operatorname{Re}(\sqrt{z^2-4}) > 0$. By Exercise 3 we have

$$0 \le 2\operatorname{Re}(w_1) = \operatorname{Re}(z) - \operatorname{Re}(\sqrt{z^2-4}) < \operatorname{Re}(z) + \operatorname{Re}(\sqrt{z^2-4}) = 2\operatorname{Re}(w_2).$$

By Exercise 3 we have $\operatorname{Im}(w_1), \operatorname{Im}(w_2) < 0$; so we must show that $0 > -\operatorname{Im}(w_1) > -\operatorname{Im}(w_2)$. Now

$$-2\operatorname{Im}(w_1) = -\operatorname{Im}(z) + \operatorname{Im}(\sqrt{z^2-4}) > -\operatorname{Im}(z) - \operatorname{Im}(\sqrt{z^2-4}) = -\operatorname{Im}(w_2).$$

5. (*iii*) Use the same idea as in Exercise 3.4 (*iii*) to identify the roots inside Γ.

9. The density is given by

$$d\nu(t) = \frac{1}{\pi} \frac{-b}{b^2 + (t-a)^2}\,dt.$$

11. Let $0 < \alpha_1 < \alpha_2$ and $\beta_2 > 0$ be given; we must find $\beta_1 > 0$ so that $f(\Gamma_{\alpha_1,\beta_1}) \subset \Gamma_{\alpha_2,\beta_2}$. Choose $\epsilon > 0$ so that

$$\frac{\sqrt{1+\alpha_2^2}}{\sqrt{1+\alpha_1^2}} > \frac{1+\epsilon}{1-\epsilon\sqrt{1+\alpha_1^2}}.$$

Choose $\beta_1 > 0$ so that for $z \in \Gamma_{\alpha_1,\beta_1}$ we have $|f(z) - z| < \epsilon|z|$. Then

$$\begin{aligned}\operatorname{Im}(f(z)) &= \operatorname{Im}(z) + \operatorname{Im}(f(z) - z) \\ &> \operatorname{Im}(z) - |f(z) - z| \\ &> \operatorname{Im}(z) - \epsilon|z|\end{aligned}$$

$$\begin{aligned}
&> ((1+\alpha_1)^{-1/2}-\epsilon)|z| \\
&= ((1+\alpha_1)^{-1/2}-\epsilon)\frac{|z|+\epsilon|z|}{1+\epsilon} \\
&> ((1+\alpha_1)^{-1/2}-\epsilon)\frac{|z|+|f(z)-z|}{1+\epsilon} \\
&\geq \frac{((1+\alpha_1)^{-1/2}-\epsilon)}{1+\epsilon}|f(z)|.
\end{aligned}$$

Thus $\sqrt{1+\alpha_2^2}\,\mathrm{Im}(f(z)) > |f(z)|$, so $f(z) \in \Gamma_{\alpha_2}$. We now have

$$\mathrm{Im}(f(z)) > \mathrm{Im}(z) - \epsilon|z| > \left(1-\epsilon\sqrt{1+\alpha_1^2}\right)\mathrm{Im}(z) > \left(1-\epsilon\sqrt{1+\alpha_1^2}\right)\beta_1.$$

So by choosing β_1 still larger, we may have $\left(1-\epsilon\sqrt{1+\alpha_1^2}\right)\beta_1 > \beta_2$. Thus $f(z) \in \Gamma_{\alpha_2,\beta_2}$.

12. (i) The result is trivial when $t = 0$. By symmetry we only need to consider the case $t > 0$. Since Γ_α is convex, the minimum of $|z-t|$ occurs when z is in $\partial\Gamma_\alpha$. The distance from t to the line $x-\alpha y = 0$ is $t/\sqrt{1+\alpha^2}$. Hence $|z-t| \geq |t|/\sqrt{1+\alpha^2}$.

(ii) Write $z = |z|e^{i\theta}$ with $\tan^{-1}(\alpha^{-1}) < \theta < \pi - \tan^{-1}(\alpha^{-1})$. If $t = 0$, the inequality is trivially true. Suppose $t > 0$; then

$$|z-t| = |\bar{z}-t| = ||z| - te^{i\theta}| \geq |z|/\sqrt{1+\alpha^2}$$

by (i) since $te^{i\theta} \in \Gamma_\alpha$. If $t < 0$, then

$$|z-t| = ||z| - te^{-i\theta}| \geq |z|/\sqrt{1+\alpha^2}$$

by (i) since $te^{-i\theta} \in \Gamma_\alpha$.

(iii) By (i), $|t/(z-t)| \leq \sqrt{1+\alpha^2}$ for $z \in \Gamma_\alpha$. Since σ is a finite measure, we may apply the dominated convergence theorem.

(iv) Now

$$zG(z) = \int_{\mathbb{R}} \frac{z}{z-t}\,d\nu(t) \qquad \text{so} \qquad zG(z) - 1 = \int_{\mathbb{R}} \frac{t}{z-t}\,d\nu(t).$$

Thus we can apply the result from (iii).

13. By Exercise 12 we have, for $\mathrm{Im}(z) \geq 1$,

$$\left|\frac{1+tz}{z(t-z)}\right| \leq \frac{1}{|t-z|} + \frac{|t|}{|t-z|} \leq 2\sqrt{1+\alpha^2}.$$

Write

$$\frac{F(z)}{z} = \frac{a}{z} + b + \int \frac{1+tz}{z(t-z)}\,d\sigma(t).$$

For a fixed t we have

$$\frac{1+tz}{z(t-z)} = \frac{t+z^{-1}}{t-z} \longrightarrow 0$$

as $z \to \infty$. Since $|(1+tz)/(z(t-z))|$ is bounded independently of t and z, then we can apply the dominated convergence theorem to conclude that $F(z)/z \to b$ as $z \to \infty$ in Γ_α.

14. (i) By assumption the function $t \mapsto |t|^n$ is integrable with respect to ν. By Exercise 12 we have for $z \in \Gamma_\alpha$

$$\frac{|t|^{n+1}}{|z-t|} \leq |t|^n \sqrt{1+\alpha^2}.$$

Thus, by the dominated convergence theorem,

$$\lim_{z\to\infty} \int \frac{t^{n+1}}{z-t}\, d\nu(t) = 0.$$

(ii) We have

$$\begin{aligned} G(z) - \Big(\frac{1}{z} + \frac{\alpha_1}{z^2} + \cdots + \frac{\alpha_n}{z^{n+1}}\Big) &= \int_{\mathbb{R}} \frac{1}{z-t} - \Big(\frac{1}{z} + \frac{t}{z^2} + \cdots + \frac{t^n}{z^{n+1}}\Big)\, d\nu(t) \\ &= \frac{1}{z^{n+1}} \int_{\mathbb{R}} \frac{t^{n+1}}{z-t}\, d\nu(t). \end{aligned}$$

Thus

$$z^{n+1}\Big(G(z) - \Big(\frac{1}{z} + \frac{\alpha_1}{z^2} + \cdots + \frac{\alpha_n}{z^{n+1}}\Big)\Big) = \int_{\mathbb{R}} \frac{t^{n+1}}{z-t}\, d\nu(t)$$

and this integral converges to 0 as $z \to \infty$ in Γ_α by (i).

15. We shall proceed by induction on n. To begin the induction process, let us show that α_1 and α_2 are, respectively, the first and second moments of ν. Note that for any $1 \leq k \leq 2n$ we have that as $z \to \infty$ in Γ_α

$$\lim_{z\to\infty} z^{k+1}\Big(G(z) - \Big(\frac{1}{z} + \frac{\alpha_1}{z^2} + \cdots + \frac{\alpha_k}{z^{k+1}}\Big)\Big) = 0.$$

Also by Exercise 12, $\int_{\mathbb{R}} |t/(z-t)|\, d\nu(t) < \infty$, so we may let

$$G_1(z) = z\Big(G(z) - \frac{1}{z}\Big) = \int_{\mathbb{R}} \frac{t}{z-t}\, d\nu(t).$$

Then since n is a least 1, we have

$$\lim_{z\to\infty} z\Big(zG_1(z)-\alpha_1-\frac{\alpha_2}{z}\Big)=\lim_{z\to\infty} z^3\Big(G(z)-\Big(\frac{1}{z}+\frac{\alpha_1}{z^2}+\frac{\alpha_2}{z^3}\Big)\Big)=0.$$

Hence

$$\lim_{z\to\infty} z\Big(zG_1(z)-\alpha_1\Big)=\alpha_2.$$

Since α_1 and α_2 are real,

$$\lim_{z\to\infty} \mathrm{Re}\Big(z\Big(zG_1(z)-\alpha_1\Big)\Big)=\alpha_2.$$

Now let $z=iy$ with $y>0$; then

$$\begin{aligned}\mathrm{Re}\Big(z\Big(zG_1(z)-\alpha_1\Big)\Big)&=\mathrm{Re}\Big(-y^2G_1(iy)-i\alpha_1 y\Big)=-y^2\mathrm{Re}(G_1(iy))\\&=-y^2\int_{\mathbb{R}}\mathrm{Re}\Big(\frac{t}{iy-t}\Big)\,d\nu(t)=-y^2\int_{\mathbb{R}}\frac{-t^2}{y^2+t^2}\,d\nu(t)\\&=\int_{\mathbb{R}}\frac{t^2}{1+(t/y)^2}\,d\nu(t).\end{aligned}$$

Thus,

$$\lim_{y\to\infty}\int_{\mathbb{R}}\frac{t^2}{1+(t/y)^2}\,d\nu(t)=\alpha_2,$$

so by the monotone convergence theorem, $\int_{\mathbb{R}} t^2\,d\nu(t)=\alpha_2$. Hence, ν has a first and second moment, and the second moment is α_2.

Since $\lim_{z\to\infty} z(zG_1(z)-\alpha_1)=\alpha_2$, we must have $\lim_{z\to\infty} zG_1(z)=\alpha_1$. Letting $z=iy$ with $y>0$, we have $\alpha_1=\lim_{y\to\infty} iyG_1(iy)$ and thus

$$\begin{aligned}\alpha_1&=\lim_{y\to\infty}\mathrm{Re}(iyG_1(iy))=\lim_{y\to\infty}\int_{\mathbb{R}}\mathrm{Re}\Big(\frac{iyt}{iy-t}\Big)\,d\nu(t)\\&=\lim_{y\to\infty}\int_{\mathbb{R}}\frac{y^2t}{y^2+t^2}\,d\nu(t)=\lim_{y\to\infty}\int_{\mathbb{R}}\frac{t}{1+(t/y)^2}\,d\nu(t).\end{aligned}$$

Now $|t/(1+(t/y)^2)|\le|t|$ and $\int_{\mathbb{R}}|t|\,d\nu(t)<\infty$, so by the dominated convergence theorem, $\alpha_1=\int_{\mathbb{R}} t\,d\nu(t)$.

Suppose that we have shown that ν has moments up to order $2n-2$ and α_k, for $1\le k\le 2n-2$, is the k^{th} moment. Thus $\int_{\mathbb{R}}|t^{2n-1}/(z-t)|\,d\nu(t)<\infty$ by Exercise 12 (i). Let us write

$$\begin{aligned} G_{2n-1}(z) &= z^{2n-1}\Big(G(z) - \Big(\frac{1}{z} + \frac{\alpha_1}{z^2} + \cdots + \frac{\alpha_{2n-2}}{z^{2n-1}}\Big)\Big) \\ &= z^{2n-1}\int_{\mathbb{R}} \frac{1}{z-t} - \Big(\frac{1}{z} + \frac{t}{z^2} + \cdots + \frac{t^{2n-2}}{z^{2n-1}}\Big)\, d\nu(t) \\ &= \int_{\mathbb{R}} \frac{t^{2n-1}}{z-t}\, d\nu(t). \end{aligned}$$

By our hypothesis $\lim_{z\to\infty} z^2(G_{2n-1}(z) - (\frac{\alpha_{2n-1}}{z} + \frac{\alpha_{2n}}{z^2})) = 0$ or equivalently

$$\lim_{z\to\infty} z(zG_{2n-1}(z) - \alpha_{2n-1}) = \alpha_{2n}. \tag{12.6}$$

Let $z = iy$ with $y > 0$. Since α_{2n-1} and α_{2n} are real,

$$\begin{aligned} \alpha_{2n} &= \lim_{y\to\infty} \mathrm{Re}\Big(iy(iyG_{2n-1}(iy) - \alpha_{2n-1})\Big) = \lim_{y\to\infty} -y^2\mathrm{Re}(G_{2n-1}(iy)) \\ &= \lim_{y\to\infty} -y^2 \int_{\mathbb{R}} \mathrm{Re}\Big(\frac{t^{2n-1}}{iy-t}\Big)\, d\nu(t) = \lim_{y\to\infty} \int_{\mathbb{R}} \frac{y^2 t^{2n}}{y^2+t^2}\, d\nu(t) \\ &= \lim_{y\to\infty} \int_{\mathbb{R}} \frac{t^{2n}}{1+(t/y)^2}\, d\nu(t). \end{aligned}$$

So again by the monotone convergence theorem, we have $\int_{\mathbb{R}} t^{2n}\, d\nu(t) = \alpha_{2n}$, and thus ν has a moment of order $2n$, and this moment is α_{2n}. Thus, ν has a moment of order $2n-1$, and from Equation (12.6), we have $\lim_{z\to\infty} zG_{2n-1}(z) = \alpha_{2n-1}$. Then by letting $z = iy$ and taking real parts, we obtain that

$$\begin{aligned} \alpha_{2n-1} &= \lim_{y\to\infty} \mathrm{Re}\Big(iyG_{2n-1}(iy)\Big) = \lim_{y\to\infty} \int_{\mathbb{R}} \mathrm{Re}\Big(\frac{iyt^{2n-1}}{iy-t}\Big)\, d\nu(t) \\ &= \lim_{y\to\infty} \int_{\mathbb{R}} \frac{t^{2n-1}}{1+(t/y)^2}\, d\nu(t). \end{aligned}$$

Thus, by the dominated convergence theorem, $\alpha_{2n-1} = \int_{\mathbb{R}} t^{2n-1}\, d\nu(t)$. This completes the induction step.

16. Let us write

$$G(z) = \frac{1}{z} + \frac{\alpha_1}{z^2} + \frac{\alpha_2}{z^3} + \frac{\alpha_3}{z^4} + \frac{\alpha_4}{z^5} + r(z)$$

where $r(z) = o(\frac{1}{z^5})$. Then

$$z - \frac{1}{G(z)} = \frac{\frac{\alpha_1}{z} + \frac{\alpha_2}{z^2} + \frac{\alpha_3}{z^3} + \frac{\alpha_4}{z^4} + zr(z)}{\frac{1}{z} + \frac{\alpha_1}{z^2} + \frac{\alpha_2}{z^3} + \frac{\alpha_3}{z^4} + \frac{\alpha_4}{z^5} + r(z)}.$$

Let us equate this with

$$\alpha_1 + \frac{\beta_0}{z} + \frac{\beta_1}{z^2} + \frac{\beta_2}{z^3} + q(z)$$

and solve for β_0, β_1, β_2, and $q(z)$. After cross multiplication we find that

$$\alpha_2 = \alpha_1^2 + \beta_0 \qquad \alpha_3 = \alpha_1\alpha_2 + \beta_0\alpha_1 + \beta_1 \qquad \alpha_4 = \alpha_1\alpha_3 + \alpha_2\beta_0 + \alpha_1\beta_1 + \beta_2.$$

Thus,

$$\beta_0 = \alpha_2 - \alpha_1^2 \qquad \beta_1 = \alpha_3 - 2\alpha_1\alpha_2 + \alpha_1^3 \qquad \beta_2 = \alpha_4 - 2\alpha_1\alpha_3 - \alpha_2^2 + 3\alpha_1^2\alpha_2 - \alpha_1^4$$

and $q(z) = o(z^{-3})$.

17. (*i*) Note that since f is proper, each point has only a finite number of preimages. So let $w_0 \in \mathbb{C}$ and let $z_1, \ldots, z_r$ be the preimages of w_0. We shall treat the case when w_0 has no preimages separately. For each i choose a chart $(\mathcal{U}_i, \varphi_i)$ of z_i and an integer m_i so that $f(\varphi^{\langle -1 \rangle}(z)) = z^{m_i}$. By shrinking the $\mathcal{U}_i$, if necessary, we may assume that they are disjoint. If we can show that there is a neighbourhood $\mathcal{V}$ of w_0 such that all preimages of points in $\mathcal{V}$ are in the union $\mathcal{U}_1 \cup \cdots \cup \mathcal{U}_r$, then $\deg_f(w) = m_1 + \cdots + m_r$ for $w \in \mathcal{V}$. This will show that the integer-valued function $\deg_f$ is locally constant, and by the connectedness of $\mathbb{C}$, we shall have that $\deg_f$ is constant.

So let us suppose that no such $\mathcal{V}$ exists and reach a contradiction. If no such $\mathcal{V}$ exists, then there is a sequence $\{w_n\}_n$ converging to w_0 such that each w_n has a preimage z_n not in $\mathcal{U}_1 \cup \cdots \cup \mathcal{U}_r$. By shrinking $\mathcal{V}$ if necessary, we may suppose that $\overline{\mathcal{V}}$ is compact. By the properness of f, we must have a subsequence $\{z_{n_k}\}_k$ of $\{z_n\}_n$ which has a limit z, say. Then $f(z) = \lim_k f(z_{n_k}) = \lim_k w_{n_k} = w_0$. So $z = z_i$ for some i, and thus the subsequence $\{z_{n_k}\}_k$ must penetrate the open set $\mathcal{U}_i$ contradicting our assumption.

If w_0 has no preimages, then we must show that there is a neighbourhood of w_0 with no preimages. If not, then there is a sequence $\{w_n\}_n$ converging to w_0 such that each w_n has a preimage. But $\{w_0\} \cup \{w_n\}_n$ is a compact set so we can extract from these preimages a convergent sequence of preimages whose limit can only be a preimage of w_0, contradicting our assumption. This proves (*i*).

(*ii*) First let us note that since $F_i'(z) \neq 0$ for $i = 1, 2$ and $z \in \mathbb{C}^+$ for each $z \in \mathbb{C}^+$, there is neighbourhood of z on which both F_1 and F_2 are one to one. So for $(z_1, z_2) \in X$ let $w = F_1(z_1)$. Then there is $\mathcal{U}$, a neighbourhood of w, and two analytic maps f_1 and f_2 defined on $\mathcal{U}$ such that for $u \in \mathcal{U}$ we have $F_i \circ f_i = id$. We then let $\mathcal{V} = \{(f_1(u), f_2(u)) \mid u \in \mathcal{U}\}$ and define $\varphi : \mathcal{V} \to \mathcal{U}$ by $\varphi(w_1, w_2) = F_1(w_1)$.

To show these charts define a complex structure on X, we must show that given two charts $(\mathcal{V}, \varphi)$ and $(\mathcal{V}', \varphi')$, we have that $\varphi' \circ \varphi^{\langle -1 \rangle}$ is analytic on $\varphi(\mathcal{V} \cap \mathcal{V}')$. So by construction we have two points (z_1, z_2) and (z_1', z_2') in X and two neighbourhoods $\mathcal{U}$ and $\mathcal{U}'$ of $F_1(z_1)$ and $F_1(z_1')$, respectively, and on these neighbourhoods we have analytic maps $f_1, f_2 : \mathcal{U} \to \mathbb{C}$ and $f_1', f_2' : \mathcal{U}' \to \mathbb{C}$ such that $F_i \circ f_i = id$ and $F_i \circ f_i' = id$. Then on $\varphi(\mathcal{V} \cap \mathcal{V}')$, we have that $\varphi' \circ \varphi^{\langle -1 \rangle}(u) = \varphi(f_1'(u), f_2'(u)) = F_1(f_1'(u)) = u$. So $\varphi' \circ \varphi^{\langle -1 \rangle} = id$ is analytic.

(*iii*) To show that θ is proper, we must show that the inverse image of a compact subset of $\mathbb{C}$ is compact. So let $K = \overline{B(z,r)}$ be given. We must show that $\theta^{\langle -1 \rangle}(K)$ is compact. Since θ is continuous we have that $\theta^{\langle -1 \rangle}(K)$ is closed. So we only have to show that every sequence in $\theta^{\langle -1 \rangle}(K)$ contains a convergent subsequence. Let $\{(z_{1,n}, z_{2,n})\}_n$ be a sequence in $\theta^{\langle -1 \rangle}(K)$. Then

$$|z_{1,n}| \leq |\theta(z_{1,n}, z_{2n})| + |z_{2,n} - F_2(z_{2,n})| \leq |z| + r + \sigma_2^2/r.$$

Likewise $|z_{2,n}| \leq |z| + r + \sigma_1^2/r$. By Lemma 19,

$$\mathrm{Im}(z_{1,n}), \mathrm{Im}(z_{2,n}) \geq \mathrm{Im}(\theta(z_{1,n}, z_{2,n})) \geq \mathrm{Im}(z) - r.$$

So there is a subsequence $\{(z_{1,n_k}, z_{2,n_k})\}_k$ such that both $\{z_{1,n_k}\}_k$ converges to z_1, say and $\{z_{2,n_k}\}_k$ converges to z_2, say. Then

$$F_1(z_1) = \lim_k F_1(z_{1,n_k}) = \lim_k F_2(z_{2,n_k}) = F_2(z_2)$$

so $(z_1, z_2) \in X$. Also $\theta(z_1, z_2) = \lim_k \theta(z_{1,n_k}, z_{2,n_k}) \in K$. Hence $(z_1, z_2) \in \theta^{\langle -1 \rangle}(K)$ as required.

12.4 Solutions to exercises in Chapter 4

5. The commutativity of J_k and J_l is a special case of the fact that J_l commutes with $\mathbb{C}[S_{l-1}]$. For the latter note that for $k < l$ and $\sigma \in S_{l-1}$

$$\sigma \cdot (k, l) \cdot \sigma^{-1} = (\sigma(k), l).$$

Thus we have

$$\sigma J_l \sigma^{-1} = \sigma((1, l) + \cdots + (l-1, l))\sigma^{-1} = (\sigma(1), l) + \cdots + (\sigma(l-1), l) = J_l.$$

7. *Hint:* Using the convention that $J_1^0 = 1$, write

$$(1 + N^{-1}J_1)^{-1}(1 + N^{-1}J_2)^{-1} \cdots (1 + N^{-1}J_n)^{-1}$$

as

$$\sum_{l \geq 0} (-N)^{-l} \sum_{\substack{k_1, \ldots, k_n \geq 0 \\ k_1 + \cdots + k_n = l}} J_1^{k_1} J_2^{k_2} \cdots J_n^{k_n}$$

and observe that $J_1^{k_1} \cdots J_n^{k_n}$ is a linear combination of permutations of length at most $k_1 + \cdots + k_n$.

9. Recall that $\gamma = \gamma_m$. Given $i : [2m] \to [n]$ such that $\ker(i) \geq \pi$, let $j : [m] \to [n]$ be defined by $j(\gamma^{-1}(k)) = i(2k-1)$ and $j(\sigma(k)) = i(2k)$. To show that such a j is well defined, we must show that when $\sigma(k) = \gamma^{-1}(l)$ we have $i(2k) = i(2l-1)$. If $\sigma(k) = \gamma^{-1}(l)$, we have that $(k, \gamma^{-1}(l))$ is a pair of σ, and thus $(2l-1, 2k)$ is a pair of π. Since we have assumed that $\ker(i) \geq \pi$, we have $i(2l-1) = i(2k)$ as required. Conversely if we have $j : [m] \to [n]$, let $i(2k-1) = j(\gamma^{-1}(k))$ and $i(2k) = j(\sigma(k))$. Then $\ker i \geq \pi$. This gives us a bijection of indices so

$$\sum_{\substack{i_1,\ldots,i_{2m}=1 \\ \ker(i)\geq\pi}}^{n} d^{(1)}_{i_1 i_2} \cdots d^{(m)}_{i_{2m-1} i_{2m}} = \sum_{j_1,\ldots,j_m=1}^{n} d^{(1)}_{j_{\gamma^{-1}(1)} j_{\sigma(1)}} \cdots d^{(m)}_{j_{\gamma^{-1}(m)} j_{\sigma(m)}}.$$

By a change of variables, we have

$$\sum_{j_1,\ldots,j_m=1}^{n} d^{(1)}_{j_{\gamma^{-1}(1)} j_{\sigma(1)}} \cdots d^{(m)}_{j_{\gamma^{-1}(m)} j_{\sigma(m)}} = \sum_{j_1,\ldots,j_m=1}^{n} d^{(1)}_{j_1 j_{\gamma\sigma(1)}} \cdots d^{(m)}_{j_m j_{\gamma\sigma(m)}}.$$

12. The first part is just the expansion of the product of matrices. Now let us write $x_{\alpha(l)} = x_{i_l i_{-l}}$ and $x_{\beta(l)} = x_{i_{\gamma(l)}, i_{-l}}$, where γ is the permutation with one cycle $(1, 2, 3, \ldots, k)$. With this notation we have by Exercise 1.7

$$\begin{aligned} \mathrm{E}(x_{i_1 i_{-1}} \overline{x_{i_2 i_{-1}}} \cdots x_{i_k i_{-k}} \overline{x_{i_1 i_{-k}}}) &= \mathrm{E}(x_{\alpha(1)} \cdots x_{\alpha(k)} \overline{x_{\beta(1)}} \cdots \overline{x_{\beta(k)}}) \\ &= |\{\sigma \in S_k \mid \alpha = \beta \circ \sigma\}|. \end{aligned}$$

If $\alpha = \beta \circ \sigma$, then $i_l = i_{\gamma(\sigma(l))}$ and $i_{-l} = i_{-\sigma(l)}$ for $1 \leq l \leq k$. Thus, for a fixed σ, there are $N^{\#(\gamma\sigma)}$ ways to choose the k-tuple $(i_1, \ldots, i_k)$ so that $i_l = i_{\gamma(\sigma(l))}$ and $M^{\#(\sigma)}$ ways of choosing the k-tuple $(i_{-1}, \ldots, i_{-k})$ so that $i_{-l} = i_{-\sigma(l)}$. Hence

$$\mathrm{E}(\mathrm{Tr}(A^k)) = \sum_{\sigma\in S_k} N^{\#(\gamma\sigma)-k} M^{\#(\sigma)} = \sum_{\sigma\in S_k} N^{\#(\sigma)+\#(\sigma^{-1}\gamma)-k} \left(\frac{M}{N}\right)^{\#(\sigma)}.$$

Thus

$$\mathrm{E}(\mathrm{tr}(A^k)) = \sum_{\sigma\in S_k} N^{\#(\sigma)+\#(\sigma^{-1}\gamma)-(k+1)} \left(\frac{M}{N}\right)^{\#(\sigma)}$$

and by Proposition 1.5 the only σ's for which the exponent of N is not negative are those σ's which are non-crossing partitions. Thus $\lim \mathrm{E}(\mathrm{tr}(A^k)) = \sum_{\sigma\in NC(k)} c^{\#(\sigma)}$.

12.5 Solutions to exercises in Chapter 5

1. One has to realize that the order on a through-cycle of a non-crossing annular permutation has to be of the following form: one moves at one point p from the first circle to the second circle, moves then on the second circle in cyclically increasing order, moves then back to the first circle, and moves then on the first circle in cyclically increasing order, until we are back to the first point p.

(i) If one has at least two through-cycles, then the positions where one has to move to the other circle, as well as the order on the cycles lying on just one circle, are uniquely determined by the annular non-crossing condition.

(*ii*) In the case of just one through-cycle, the order on this is not uniquely determined, but depends on the choice of a point p on the first circle and a point q on the second circle of this through-cycle. We can then fix the order by sending p to q, and the order on this through block as well as the order on all other blocks is then determined. So we have mn choices, each giving a different permutation.

10. (*i*) Calculate the free cumulants with the help of the product formula (2.19), and observe that, in both cases, there is for each n exactly one contributing pairing in (2.19); thus $\kappa_n(s^2, \dots, s^2) = 1 = \kappa_n(cc^*, \dots, cc^*)$.

(*ii*) In Example 5.33 (and in Example 5.36), it was shown that $\kappa_{1,1}(s^2, s^2) = 1$.

(*iii*) Use the second order version (5.16) of the product formula to see that all second order moments of cc^* are zero. It is instructive to do this for the case $\kappa_{1,1}(cc^*, cc^*) = 0$ and compare this with the calculation of $\kappa_{1,1}(s^2, s^2) = 1$ in Example 5.36. In both cases we have the term corresponding to π_1; whereas it makes the contribution $\kappa_2(s, s)\kappa_2(s, s) = 1$ in the first case, in the second case its contribution is $\kappa_2(c, c)\kappa_2(c^*, c^*) = 0$.

12.6 Solutions to exercises in Chapter 6

1. (*i*) Begin by recalling that every element $x \in \mathcal{L}(G)$ defines a function on G as follows: $\lambda(x)\delta_e \in \ell^2(G)$ and for convenience we call this square summable function x. If x is in the centre of $\mathcal{L}(G)$, then x must be constant on all conjugacy classes; so if G has the ICC property, then every x in the centre of $\mathcal{L}(G)$ vanishes on all conjugacy classes, except possibly the one containing e. Such an x must then be a scalar multiple of the identity. This shows that if G has the ICC property, then $\mathcal{L}(G)$ is a factor. If G does not have the ICC property and $X \subset G$ is a finite conjugacy class not containing e, then the indicator function of X is in the centre of $\mathcal{L}(G)$ and is not a scalar multiple of the identity, and thus $\mathcal{L}(G)$ is not a factor.

(*ii*) Suppose we are given $\sigma \in S_n$ with $\sigma \neq e$. Then there is $k \leq n$ such that $\sigma(k) \neq k$. Let $\tau_m = (k, m)$ for $m > n$. Note that $\tau_m \sigma \tau_m^{-1}$ moves m but fixes all $l > m$. Thus $\{\tau_m \sigma \tau_m^{-1}\}_m$ is infinite.

2. It suffices to consider the case $\mathbb{F}_2$. A reduced word in $\mathbb{F}_2$ can be written as $g = a_{i_1}^{\epsilon_1} \cdots a_{i_n}^{\epsilon_n}$ where whenever $i_r = i_{r+1}$ for $1 \leq r < n$ we must have $\epsilon_r = \epsilon_{r+1}$. Let us show that the conjugacy class of g is infinite. If g is a power of a_1, then $a_2^m g a_2^{-m}$ are all distinct for $m = 1, 2, 3, \dots$, and hence g has an infinite conjugacy class, likewise if g is a power of a_2. So now we can suppose that there is k such that $i_1 = \cdots = i_k \neq i_{k+1}$. Let $h_m = a_{i_1}^{m\epsilon_1}$. We claim that all $h_m g h_m^{-1}$ $(m \geq 1)$ are distinct. If we could find $r < s$ with $h_r g h_r^{-1} = h_s g h_s^{-1}$, then $g = h_p g h_p^{-1}$ with $p = s - r$. Let us consider the reduced form of $h_p g h_p^{-1}$. The p copies of $a_{i_1}^{-\epsilon_1}$ on the right of $h_p g h_p^{-1}$ cannot cancel off $a_{i_{k+1}}^{\epsilon_{k+1}}$ because $i_1 = i_k \neq i_{k+1}$. Thus in reduced form $h_p g h_p^{-1}$ starts with the letter $a_{i_1}^{\epsilon_1}$ repeated $p + k$ times. However in reduced form g starts with the letter a_{i_1} repeated k times. Hence, in reduced form, the words in $\{h_m g h_m^{-1}\}_m$ are distinct, and thus the conjugacy class of g is infinite.

4. We shall just compute $\mathrm{tr} \otimes \varphi(x^n)$ directly using the moment-cumulant formula. For this calculation we will need to rewrite

$$x = \frac{1}{\sqrt{2}} \begin{pmatrix} s_1 & c \\ c^* & s_2 \end{pmatrix} \qquad \text{as} \qquad \frac{1}{\sqrt{2}} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

Then

$$\mathrm{tr} \otimes \varphi(x^n) = \frac{1}{2}\varphi(\mathrm{Tr}(x^n)) = 2^{-(1+n/2)} \sum_{i_1,\dots,i_n=1}^{2} \varphi(a_{i_1 i_2} \cdots a_{i_n i_1}).$$

Now given $i_1, \dots, i_n$

$$\varphi(a_{i_1 i_2} \cdots a_{i_n i_1}) = \sum_{\pi \in NC(n)} \kappa_\pi(a_{i_1 i_2}, \dots, a_{i_n i_1}) = \sum_{\pi \in NC_2(n)} \kappa_\pi(a_{i_1 i_2}, \dots, a_{i_n i_1})$$

because (i) all mixed cumulants vanish so each block of π must consist either of all a_{11}'s or all a_{22}'s or a mixture of a_{21} and a_{12}, and (ii) the only non-zero cumulant of a_{ii} is κ_2 so any blocks that contain a_{ii} must be κ_2 and (iii) the only non-zero $*$-cumulants of a_{ij} (for $i \neq j$) are $\kappa_2(a_{ij}, a_{ij}^*)$ and $\kappa_2(a_{ij}^*, a_{ij})$. Thus we have a sum over pairings. Moreover, if $\pi \in NC_2(n)$ is a pairing and if (r, s) is a pair of π, then $\kappa_\pi(a_{i_1 i_2}, \dots, a_{i_n i_1})$ will be 0 unless $a_{i_r i_{r+1}} = (a_{i_s i_{s+1}})^*$, i.e. $i_r = i_{s+1}$ and $i_s = i_{r+1}$. For such a π the contribution is 1 since s_1, s_2 and c all have variance 1. Hence, letting $\gamma = (1, 2, 3, \dots, n)$ as in Chapter 1, we have $\varphi(a_{i_1 i_2} \cdots a_{i_n i_1}) = |\{\pi \in NC_2(n) \mid i = i \circ \gamma \circ \pi\}|$. Thus

$$\begin{aligned} \mathrm{tr} \otimes \varphi(x^n) &= 2^{-(1+n/2)} \sum_{\pi \in NC_2(n)} |\{i : [n] \to [2] \mid i = i \circ \gamma \circ \pi\}| \\ &= 2^{-(1+n/2)} \sum_{\pi \in NC_2(n)} 2^{\#(\gamma\pi)}. \end{aligned}$$

Now recall from Chapter 1 that for any pairing (interpreted as a permutation in S_n)

$$\#(\pi) + \#(\gamma\pi) + \#(\gamma) = n + 2(1 - g)$$

and $\pi \in NC_2(n)$ if and only if $g = 0$. Thus for any $\pi \in NC_2(n)$

$$\#(\gamma\pi) = n + 2 - \#(\gamma) - \#(\pi) = 1 + n/2.$$

Hence $\mathrm{tr} \otimes \varphi(x^n) = |NC_2(n)|$ is the n^{th} moment of a semi-circular operator.

5. (i) A product of alternating centred elements from $\mathcal{A}_{11}, \mathcal{A}_{12}, \mathcal{A}_{21}, \mathcal{A}_{22}$ can, by multiplying neighbours from $\mathcal{A}_1$ and from $\mathcal{A}_2$, be read as a product of alternating elements from $\mathcal{A}_1$ and $\mathcal{A}_2$; that those elements are also centred follows from the freeness of $\mathcal{A}_{11}$ and $\mathcal{A}_{12}$ in $\mathcal{A}_1$ and the freeness of $\mathcal{A}_{21}$ and $\mathcal{A}_{22}$ in $\mathcal{A}_2$.

(*ii*) Compare the remarks after Theorem 4.8

(*iii*). It is clear that $u_2u_1^*$ is a unitary and, by $*$-freeness of u_1 and u_2 and the centredness of u_1 and u_2, that $\varphi((u_2u_1^*)^p) = \delta_{0p}$ for any $p \in \mathbb{Z}$. For the $*$-freeness between $u_2u_1^*$ and $u_1\mathcal{A}u_1^*$, it suffices to show that alternating products in elements of the form $(u_2u_1^*)^p$ $(p \in \mathbb{Z}\setminus\{0\})$ and centred elements from $u_1\mathcal{A}u_1^*$ are also centred. But this is clear, since $\varphi(u_1au_1^*) = \varphi(a)\varphi(u_1u_1^*) = \varphi(a)$ and thus centred elements from $u_1\mathcal{A}u_1^*$ are of the form $u_1au_1^*$ with centred a.

12.7 Solutions to exercises in Chapter 7

1. Let μ be the distribution of X, then $\mathrm{E}(e^{\lambda X}) = \int e^{\lambda x}\, d\mu(x)$. We notice that for $\lambda > 0$ we have $\int_{-\infty}^0 e^{\lambda x}\, d\mu(x) < \infty$ because the integrand is bounded by 1 and μ is a probability measure. If $\mathrm{E}(e^{\lambda X}) < \infty$ for some $\lambda > 0$, we must have $\int_0^\infty e^{\lambda x}\, d\mu(x) < \infty$. Now expand $e^{\lambda x}$ into a power series; for $\lambda \geq 0$ all the terms are positive. Hence $\int_0^\infty x^n\, d\mu(x) < \infty$ for all n. Likewise if for some $\lambda < 0$ we have $\mathrm{E}(e^{\lambda X}) < \infty$, then for all n, $\int_{-\infty}^0 x^n\, d\mu(x) < \infty$. Hence, if $\mathrm{E}(e^{\lambda X}) < \infty$ for all $|\lambda| \leq \lambda_0$, then X has moments of all orders and $\mathrm{E}(e^{\lambda_0|X|}) < \infty$. Thus, by the dominated convergence theorem, $\lambda \mapsto \mathrm{E}(e^{\lambda X})$ has a convergent power series expansion in λ with a radius of convergence of at least λ_0. In fact the proof shows that if there are $\lambda_1 < 0$ and $\lambda_2 > 0$ with $\mathrm{E}(e^{\lambda_1 X}) < \infty$ and $\mathrm{E}(e^{\lambda_2 X}) < \infty$, then for all $\lambda_1 \leq \lambda \leq \lambda_2$, we have $\mathrm{E}(e^{\lambda X}) < \infty$, and we may choose $\lambda_0 = \min\{-\lambda_1, \lambda_2\}$.

3. (*i*) We have

$$\begin{aligned} a_{m+n} &= \log[P(X_1 + \cdots + X_{m+n} > (m+n)a)] \\ &\geq \log[P(X_1 + \cdots + X_m > ma \text{ and } X_{m+1} + \cdots + X_{m+n} > na)] \\ &= \log[P(X_1 + \cdots + X_m > ma) \cdot P(X_{m+1} + \cdots + X_{m+n} > na)] \\ &= a_m + a_n. \end{aligned}$$

(*ii*) Fix m; for $n \geq m$ write $n = rm + s$ with $0 \leq s < m$, then

$$\frac{a_n}{n} \geq \frac{ra_m + a_s}{n} = \frac{rm}{n}\frac{a_m}{m} + \frac{a_s}{n} \to \frac{a_m}{m}.$$

(*iii*) We have

$$\limsup_n \frac{a_n}{n} \leq \sup_m \frac{a_m}{m} \leq \liminf_n \frac{a_n}{n}.$$

5. We have learned this statement and its proof from an unpublished manuscript of Uffe Haagerup.

(*i*) By using the Taylor series expansion

$$\log(1-z) = -\sum_{n=1}^\infty \frac{z^n}{n},$$

which converges for every complex number $z \neq 1$ with $|z| \leq 1$, we derive an expansion for $\log |s-t|$, by substituting $s = 2\cos u$ and $t = 2\cos v$:

$$\begin{aligned}
\log|s-t| &= \log|e^{iu}+e^{-iu}-e^{iv}-e^{-iv}| \\
&= \log|e^{-iu}(1-e^{i(u+v)})(1-e^{i(u-v)})| \\
&= \log|1-e^{i(u+v)}| + \log|1-e^{i(u-v)}| \\
&= \mathrm{Re}(\log(1-e^{i(u+v)}) + \log(1-e^{i(u-v)})) \\
&= -\mathrm{Re}\sum_{n=1}^{\infty}\frac{1}{n}(e^{in(u+v)}+e^{in(u-v)}) \\
&= -\mathrm{Re}\sum_{n=1}^{\infty}\frac{2}{n}e^{inu}\cos nv \\
&= -\sum_{n=1}^{\infty}\frac{2}{n}\cos nu \cos nv \\
&= -\sum_{n=1}^{\infty}\frac{1}{2n}C_n(s)C_n(t).
\end{aligned}$$

Then one has to show (which is not trivial) that the convergence is strong enough to allow term-by-term integration.

(*ii*) For this one has to show that

$$\int_{-2}^{+2} C_n(t)\,d\mu_W(t) = \begin{cases} 2, & n=0 \\ -1, & n=2 \\ 0, & \text{otherwise} \end{cases}.$$

6. (*i*) Let us first see that the mapping $T \otimes I_N : (M_N^{sa})^n \to (M_N^{sa})^n$ transports microstates for $(x_1,\ldots,x_n)$ into microstates for $(y_1,\ldots,y_n)$. Namely, let $A = (A_1,\ldots,A_n) \in \Gamma(x_1,\ldots,x_n; N, r, \epsilon)$ be a microstate for $(x_1,\ldots,x_n)$, and consider $B = (B_1,\ldots,B_n) := (T \otimes I_N)A$, i.e., $B_i = \sum_{j=1}^n t_{ij}A_j$. Then we have for each $k \leq r$:

$$\begin{aligned}
&|\tau(y_{i_1}\cdots y_{i_k}) - \mathrm{tr}(B_{i_1}\cdots B_{i_k})| \\
&= |\tau(\sum_{j_1=1}^{n} t_{i_1j_1}x_{j_1}\cdots\sum_{j_k=1}^{n} t_{i_kj_k}x_{j_k}) - \mathrm{tr}(\sum_{j_1=1}^{n} t_{i_1j_1}A_{j_1}\cdots\sum_{j_k=1}^{n} t_{i_kj_k}A_{j_k})| \\
&\leq \sum_{j_1,\ldots,j_k=1}^{n} |t_{i_1j_1}\cdots t_{i_kj_k}|\cdot|\tau(x_{j_1}\cdots x_{j_k}) - \mathrm{tr}(A_{j_1}\cdots A_{j_k})| \\
&\leq (cn)^r\epsilon,
\end{aligned}$$

where $c := \max_{i,j}\{|t_{ij}|\}$. Thus we have shown

$$(T \otimes I_N)(\Gamma(x_1, \dots, x_n; N, r, \epsilon)) \subseteq \Gamma(y_1, \dots, y_n; N, r, (cn)^r \epsilon).$$

The Lebesgue measure Λ on $M_N(\mathbb{C})^n_{sa} \simeq \mathbb{R}^{nN^2}$ scales under the linear mapping $T \otimes I_N$ as

$$\begin{aligned}\Lambda[(T \otimes I_N)(\Gamma(x_1, \dots, x_n; N, r, \epsilon))] &= \Lambda[\Gamma(x_1, \dots, x_n; N, r, \epsilon)] \cdot |\det(T \otimes I_N)| \\ &= \Lambda[\Gamma(x_1, \dots, x_n; N, r, \epsilon)] \cdot |\det T|^{N^2}.\end{aligned}$$

This yields then for the free entropies the estimate

$$\chi(y_1, \dots, y_n) \geq \chi(x_1, \dots, x_n) + \log|\det T|.$$

In order to get the reverse inequality, we do the same argument for the inverse map, $(x_1, \dots, x_n) = T^{-1}(y_1, \dots, y_n)$, which gives

$$\chi(x_1, \dots, x_n) \geq \chi(y_1, \dots, y_n) + \log|\det T^{-1}| = \chi(y_1, \dots, y_n) - \log|\det T|.$$

(*ii*) If $(x_1, \dots, x_n)$ are linear dependent, there are $(\alpha_1, \dots, \alpha_n) \in \mathbb{C}^n \setminus \{0\}$ such that $0 = \alpha_1 x_1 + \cdots + \alpha_n x_n$. Since the x_i are self-adjoint, the α_i can be chosen real. Without restriction, we can assume that $\alpha_1 \neq 0$.

Now consider $T = I_n + \beta T'$, where $T' = (t_{ij})_{i,j=1}^n$ with $t_{ij} = \delta_{1i}\alpha_j$. Then T is invertible for any $\beta \neq -\alpha_1^{-1}$ and $\det T = 1 + \alpha_1 \beta$.

On the other hand, we also have $T(x_1, \dots, x_n) = (x_1, \dots, x_n)$. Hence, by (*i*),

$$\chi(x_1, \dots, x_n) = \chi(x_1, \dots, x_n) + \log|\det T| = \chi(x_1, \dots, x_n) + \log|1 + \alpha_1 \beta|.$$

Since β is arbitrary and $\chi \in [-\infty, +\infty)$, we must have $\chi(x_1, \dots, x_n) = -\infty$.

12.8 Solutions to exercises in Chapter 8

2. We have

$$\partial_j(X_{i_1} \cdots X_{i_k})(X_j \otimes 1) = \sum_{l=1}^{k} \delta_{j,i_l} X_{i_1} \cdots X_{i_l} \otimes X_{i_{l+1}} \cdots X_{i_k}$$

where we have adopted the convention that we have $X_{i_1} \cdots X_{i_l} \otimes X_{i_{l+1}} \cdots X_{i_k} = X_{i_1} \cdots X_{i_k} \otimes 1$ when $l = k$. Similarly

$$(1 \otimes X_j)\partial_j(X_{i_1} \cdots X_{i_k}) = \sum_{l=1}^{k} \delta_{j,i_l} X_{i_1} \cdots X_{i_{l-1}} \otimes X_{i_l} \cdots X_{i_k},$$

and we have adopted the convention that $X_{i_1} \cdots X_{i_{l-1}} \otimes X_{i_l} \cdots X_{i_k} = 1 \otimes X_{i_1} \cdots X_{i_k}$ when $l = 1$. Thus

$$
\begin{aligned}
&\sum_j \partial_j (X_{i_1} \cdots X_{i_k})(X_j \otimes 1) - (1 \otimes X_j)\partial_j (X_{i_1} \cdots X_{i_k}) \\
&\quad = \sum_j \sum_{l=1}^{k} \delta_{j,i_l} X_{i_1} \cdots X_{i_l} \otimes X_{i_{l+1}} \cdots X_{i_k} - \delta_{j,i_l} X_{i_1} \cdots X_{i_{l-1}} \otimes X_{i_l} \cdots X_{i_k} \\
&\quad = \sum_{l=1}^{k} X_{i_1} \cdots X_{i_l} \otimes X_{i_{l+1}} \cdots X_{i_k} - X_{i_1} \cdots X_{i_{l-1}} \otimes X_{i_l} \cdots X_{i_k} \\
&\quad = X_{i_1} \cdots X_{i_k} \otimes 1 - 1 \otimes X_{i_1} \cdots X_{i_k}
\end{aligned}
$$

because $\sum_j \delta_{j,i_l} = 1$ for all l.

3. (i) By linearity we are reduced to checking identities on monomials. So consider $p = x_{i_1} \cdots x_{i_k}$; hence, $p^* = x_{i_k} \cdots x_{i_1}$. Then

$$
\partial_i p = \sum_{l=1}^{k} \delta_{i,i_l} x_{i_1} \cdots x_{i_{l-1}} \otimes x_{i_{l+1}} \cdots x_{i_k}, \quad \partial_i p^* = \sum_{l=1}^{k} \delta_{i,i_l} x_{i_k} \cdots x_{i_{l+1}} \otimes x_{i_{l-1}} \cdots x_{i_1}.
$$

Thus

$$
\langle \xi_i, p \rangle = \langle \partial_i^*(1 \otimes 1), p \rangle = \langle 1 \otimes 1, \partial_i p \rangle = \sum_{l=1}^{k} \delta_{i,i_l} \tau(x_{i_1} \cdots x_{i_{l-1}}) \tau(x_{i_{l+1}} \cdots x_{i_k})
$$

and

$$
\langle \xi_i, p^* \rangle = \langle \partial_i^*(1 \otimes 1), p^* \rangle = \langle 1 \otimes 1, \partial_i p^* \rangle \sum_{l=1}^{k} \delta_{i,i_l} \tau(x_{i_k} \cdots x_{i_{l+1}}) \tau(x_{i_{l-1}} \cdots x_{i_1}).
$$

(ii) Consider again a monomial $p = x_{i_1} \cdots x_{i_k}$ as above. Then

$$
(\partial_i p^*)^* = \sum_{l=1}^{k} \delta_{i,i_l} x_{i_{l+1}} \cdots x_{i_k} \otimes x_{i_1} \cdots x_{i_{l-1}}.
$$

(iii) First we note that for $r \in \mathbb{C}\langle x_1, \ldots, x_n \rangle$, we have by the Leibniz rule

$$
\begin{aligned}
\langle p \cdot \partial_i^*(1 \otimes 1) \cdot q, r \rangle &= \langle \partial_i^*(1 \otimes 1), p^* r q^* \rangle = \langle 1 \otimes 1, \partial_i (p^* r q^*) \rangle \\
&= \langle 1 \otimes 1, \partial_i (p^*) \cdot 1 \otimes r q^* \rangle + \langle 1 \otimes 1, p^* \otimes 1 \cdot \partial_i r \cdot 1 \otimes q^* \rangle \\
&\quad + \langle 1 \otimes 1, p^* r \otimes 1 \cdot \partial_i (q^*) \rangle.
\end{aligned}
$$

The first term becomes $\langle (id \otimes \tau)(\partial_i p) \cdot q, r\rangle$, the middle term becomes $\langle \partial_i^*(p \otimes q), r\rangle$, and the last term becomes $\langle p \cdot (\tau \otimes id)(\partial_i q), r\rangle$.

(iv) We write $p = x_{i_1} \cdots x_{i_k}$ and $q = x_{j_1} \cdots x_{j_n}$. Then using the expansion in (i), we have

$$\begin{aligned}&\langle id \otimes \tau(\partial_i p), id \otimes \tau(\partial_i q)\rangle\\&\quad = \sum_{l=1}^{k}\sum_{m=1}^{n} \delta_{i,i_l}\delta_{i,j_m}\tau[\tau(x_{j_n}\cdots x_{j_{m+1}})x_{j_{m-1}}\cdots x_{j_1}x_{i_1}\cdots x_{i_{l-1}}\tau(x_{i_{l+1}}\cdots x_{i_k})].\end{aligned}$$

Next

$$\begin{aligned}&\langle 1 \otimes \xi_i, \partial_i(p^*) \cdot 1 \otimes q\rangle\\&\quad = \sum_{l=1}^{k}\sum_{m=1}^{n} \delta_{i,i_l}\delta_{i,j_m}\tau[\tau(x_{j_n}\cdots x_{j_{m+1}})x_{j_{m-1}}\cdots x_{j_1}x_{i_1}\cdots x_{i_{l-1}}\tau(x_{i_{l+1}}\cdots x_{i_k})]\\&\qquad + \sum_{l=1}^{k}\sum_{r=1}^{l-1} \delta_{i,i_l}\delta_{i,i_r}\tau[x_{j_n}\cdots x_{j_1}x_{i_1}\cdots x_{i_{r-1}}\tau(x_{i_{r+1}}\cdots x_{i_{l-1}})\tau(x_{i_{l+1}}\cdots x_{i_k})]\end{aligned}$$

and

$$\begin{aligned}&\langle \xi_i \otimes 1, \partial_i(p^*) \cdot 1 \otimes q\rangle\\&\quad = \sum_{l=1}^{k}\sum_{r=l+1}^{k} \delta_{i,i_l}\delta_{i,i_r}\tau[x_{j_n}\cdots x_{j_1}x_{i_1}\cdots x_{i_{l-1}}\tau(x_{i_{l+1}}\cdots x_{i_{r-1}})\tau(x_{i_{r+1}}\cdots x_{i_k})].\end{aligned}$$

(v) Check that for $p, r \in \mathbb{C}\langle x_1, \ldots, x_n\rangle$, we have

$$\langle (id \otimes \tau)(\partial_i r), p\rangle = \langle r, \partial_i^*(p \otimes 1)\rangle.$$

This shows that $\mathbb{C}\langle x_1, \ldots, x_n\rangle$ is in the domain of the adjoint of $(id \otimes \tau) \circ \partial_i$; hence, this adjoint has a dense domain, and thus $(id \otimes \tau) \circ \partial_i$ is itself closable.

5. (i) Since we assumed that we have $p \in L^3(\mathbb{R})$, we have that h_ϵ and $H(p)$ are in $L^3(\mathbb{R})$ and $\|h_\epsilon - H(p)\|_3 \to 0$. Thus by Hölder's inequality

$$\int |h_\epsilon(s) - H(p)(s)|^2 p(s)\, ds \le \|h_\epsilon - H(p)\|_3^2\, \|p\|_3.$$

If f is a polynomial, it is bounded on the support of p which is contained in $[-\|x\|, \|x\|]$. Thus

$$\int |f(s)|^2 |h_\epsilon(s) - H(p)(s)|^2 p(s)\, ds \to 0$$

as $\epsilon \to 0^+$. Thus

$$2\pi \int f(s)h_\epsilon(s)p(s)\,ds \to 2\pi \int f(s)H(p)(s)p(s)\,ds = 2\pi\,\tau(f(x)H(p)(x)) \\ = \tau(f(x)\xi).$$

For s and t real and f a polynomial, we have for $\epsilon > 0$

$$\left|\frac{(s-t)(f(s)-f(t))}{(s-t)^2+\epsilon^2}\right| \le \left|\frac{f(s)-f(t)}{s-t}\right|$$

and the right-hand side is bounded on compact subsets of $\mathbb{R}^2$. Thus

$$\lim_{\epsilon\to 0^+} \iint \frac{(s-t)(f(s)-f(t))}{(s-t)^2+\epsilon^2} p(s)p(t)\,ds\,dt \\ = \iint \frac{f(s)-f(t)}{s-t} p(s)p(t)\,ds\,dt = \tau\otimes\tau(\partial f(x)).$$

On the other hand,

$$\begin{aligned}
&\iint \frac{(s-t)(f(s)-f(t))}{(s-t)^2+\epsilon^2} p(s)p(t)\,ds\,dt \\
&= \iint \frac{(s-t)f(s)}{(s-t)^2+\epsilon^2} p(s)p(t)\,ds\,dt - \iint \frac{(s-t)f(t)}{(s-t)^2+\epsilon^2} p(s)p(t)\,ds\,dt \\
&= \iint \frac{(s-t)f(s)}{(s-t)^2+\epsilon^2} p(s)p(t)\,ds\,dt - \iint \frac{(t-s)f(s)}{(s-t)^2+\epsilon^2} p(s)p(t)\,ds\,dt \\
&= 2\iint \frac{(s-t)f(s)}{(s-t)^2+\epsilon^2} p(s)p(t)\,ds\,dt \\
&= 2\int f(s)p(s)\left[\int p(t)\frac{s-t}{(s-t)^2+\epsilon^2}\,dt\right] ds \\
&= 2\pi \int f(s)p(s)h_\epsilon(s)\,ds \\
&\to \tau(f(x)\xi) \qquad \text{for } \epsilon \to 0.
\end{aligned}$$

Thus $\tau(f(x)\xi) = \tau\otimes\tau(\partial f(x))$ so ξ satisfies the conjugate relation. Since ξ is a function of x, ξ is the conjugate variable for x.

(*ii*) Let γ be the curve $\{i\epsilon + Re^{i\theta} \mid 0 \le \theta \le \pi\} \cup \{x + i\epsilon \mid -R \le x \le R\} \subset \mathbb{C}^+$. As G is analytic on $\mathbb{C}^+$, we have that the integral $\int_\gamma G(z)^3\,dz = 0$. Thus

$$\int_{-R}^{R} G(x+i\epsilon)^3\,dx = -i\int_0^\pi G(i\epsilon + Re^{i\theta})^3\,Re^{i\theta}d\theta.$$

Now for $c = \|x\|$ and for $R > c$, we have

$$|G(i\epsilon + Re^{i\theta})| \leq \int_{-c}^{c} \frac{p(t)}{|i\epsilon + Re^{i\theta} - t|}\,dt \leq \frac{1}{R-c}\int_{-c}^{c} p(t)\,dt = \frac{1}{R-c}.$$

Hence

$$\left|\int_{-R}^{R} G(x+i\epsilon)^3\,dx\right| = \left|\int_{0}^{\pi} G(i\epsilon+Re^{i\theta})^3\,Re^{i\theta}\,d\theta\right| \leq \frac{R\pi}{(R-c)^3} \to 0 \quad \text{as } R\to\infty.$$

Thus $\int G(x+i\epsilon)^3\,dx = 0$. By taking the imaginary part of this equality, we get that

$$\int h_\epsilon(s)^2 p(s)\,ds = 3\int p(s)^3\,ds.$$

6. We begin by extending τ to vectors in L^2 by setting $\tau(\eta) = \langle \eta, 1\rangle$. If $\eta \in M$, then $\langle \eta, 1\rangle = \tau(1^*\eta)$, so the two ways of computing $\tau(\eta)$ agree. If π is any partition, we define $\tau_\pi(\eta, a_2, \ldots, a_n)$ to be the product, along the blocks of π, of τ applied to the product of elements of each block. One block will contain η, but η is the only argument that is unbounded, and it is in L^2, so all factors are defined and finite. We can also use the cumulant-moment formula

$$\kappa_n(\eta, a_2, \ldots, a_n) = \sum_{\pi\in NC(n)} \mu(\pi, 1_n)\tau_\pi(\eta, a_2, \ldots, a_n)$$

to extend the definition of $\kappa_n(\eta, a_2, \ldots, a_n)$; $\kappa_\pi(\eta, a_2, \ldots, a_n)$ is then defined as the product of cumulants along the blocks of π.

Let us first show that (*ii*) implies (*i*). If we have (*ii*), then $\kappa_\pi(\xi_i, x_{i(1)}, \ldots, x_{i(m)})$ is only different from 0 if 1 belongs to a block of size 2. This means that the only contributing partitions in the moment-cumulant formula for $\tau(\xi_i x_{i(1)} \cdots x_{i(m)})$ are of the form $\pi = \{(1,k)\} \cup \sigma_1 \cup \sigma_2$, where σ_1 is a non-crossing partition of $[1, k-1]$ and σ_2 is a non-crossing partition of $[k+1, m]$. Then we have

$$\begin{aligned}
\tau(\xi_i x_{i(1)} \cdots x_{i(m)}) &= \sum_{(1,k)\cup\sigma_1\cup\sigma_2} \kappa_2(\xi_i, x_{i(k)})\kappa_{\sigma_1}(x_{i(1)}, \ldots, x_{i(k-1)})\kappa_{\sigma_2}(x_{i(k+1)}, \ldots, x_{i(m)}) \\
&= \sum_k \kappa_2(\xi_i, x_{i(k)}) \left(\sum_{\sigma_1} \kappa_{\sigma_1}(x_{i(1)}, \ldots, x_{i(k-1)})\right)\left(\sum_{\sigma_2} \kappa_{\sigma_2}(x_{i(k+1)}, \ldots, x_{i(m)})\right) \\
&= \sum_k \delta_{i\,i(k)}\tau(x_{i(1)} \cdots x_{i(k-1)})\tau(x_{i(k+1)} \cdots x_{i(m)}).
\end{aligned}$$

Let us now show that (*i*) implies (*ii*). We do this by induction on m. It is clear that the conjugate relations (8.12) for $m = 0$ and $m = 1$ are equivalent to the cumulant relations for $m = 0$ and $m = 1$ from (*ii*). So it remains to consider the cases $m \geq 3$.

Assume (i) and that we have already shown the conditions (ii) up to $m-1$. We have to show it for m. By our induction hypothesis we know that in

$$\tau(\xi_i x_{i(1)} \cdots x_{i(m)}) = \sum_{\pi \in NC(m+1)} \kappa_\pi(\xi_i, x_{i(1)}, \dots, x_{i(m)})$$

the cumulants involving ξ_i are either of length 2 or they are the maximal one, $\kappa_{m+1}(\xi_i, x_{i(1)}, \dots, x_{i(m)})$; hence

$$\begin{aligned}\tau(\xi_i x_{i(1)} \cdots x_{i(m)}) &= \sum_{\pi=(1,k)\cup\sigma_1\cup\sigma_2} \kappa_\pi(\xi_i, x_{i(1)}, \dots, x_{i(m)}) + \kappa_{m+1}(\xi_i, x_{i(1)}, \dots, x_{i(m)}) \\ &= \sum_k \delta_{ii(k)} \tau(x_{i(1)} \cdots x_{i(k-1)}) \tau(x_{i(k+1)} \cdots x_{i(m)}) + \kappa_{m+1}(\xi_i, x_{i(1)}, \dots, x_{i(m)}).\end{aligned}$$

Since the first sum gives by our assumption (i) the value $\tau(\xi_i x_{i(1)} \cdots x_{i(m)})$, it follows that $\kappa_{m+1}(\xi_i, x_{i(1)}, \dots, x_{i(m)}) = 0$.

7. By Theorem 8.20 we have to show that $\kappa_1(\xi) = 0$, $\kappa_2(\xi, x_1 + x_2) = 1$ and $\kappa_{m+1}(\xi, x_1+x_2, \dots, x_1+x_2) = 0$ for all $m \geq 2$. However, this follows directly from the facts that ξ is conjugate variable for x_1 (hence we have $\kappa_1(\xi) = 0$, $\kappa_2(\xi, x_1) = 1$ and $\kappa_{m+1}(\xi, x_1, \dots, x_1) = 0$ for all $m \geq 2$) and that mixed cumulants in $\{x_1, \xi\}$ and x_2 vanish; for this note that ξ as a conjugate variable is in $L^2(x_1)$ and the vanishing of mixed cumulants in free variables goes also over to a situation, where one of the variables is in L^2.

8. By Theorem 8.20, the condition that for a conjugate system we have $\xi_i = x_i$ is equivalent to the cumulant conditions: $\kappa_1(x_i) = 0$, $\kappa_2(x_i, x_{i(1)}) = \delta_{ii(1)}$ and $\kappa_{m+1}(x_i, x_{i(1)}, \dots, x_{i(m)}) = 0$ for $m \geq 2$ and all $1 \leq i, i(1), \dots, i(m) \leq n$. But these are just the cumulants of a free semi-circular family.

9. Note that in the special case where $i \notin \{i(1), \dots, i(k-1), i(k+1), \dots, i(m)\}$, we have

$$\partial_i^* s_{i(1)} \cdots s_{i(k-1)} \otimes s_{i(k+1)} \cdots s_{i(m)} = s_{i(1)} \cdots s_{i(k-1)} s_i s_{i(k+1)} \cdots s_{i(m)}.$$

This follows by noticing that in this case in the formula (8.6) for the action of ∂_i^*, only the first term is different from zero and gives, by also using $\partial_i^*(1 \otimes 1) = s_i$, exactly the above result.

Thus, we get in the case where all $i(1), \dots, i(m)$ are different:

$$\begin{aligned}\sum_{i=1}^n \partial_i^* \partial_i s_{i(1)} \cdots s_{i(m)} &= \sum_{i=1}^n \sum_{k=1}^m \delta_{ii(k)} \partial_i^* s_{i(1)} \cdots s_{i(k-1)} \otimes s_{i(k+1)} \cdots s_{i(m)} \\ &= \sum_{k=1}^m \partial_{i(k)}^* s_{i(1)} \cdots s_{i(k-1)} \otimes s_{i(k+1)} \cdots s_{i(m)}\end{aligned}$$

$$
\begin{aligned}
&= \sum_{k=1}^{m} s_{i(1)} \cdots s_{i(k-1)} s_{i(k)} s_{i(k+1)} \cdots s_{i(m)} \\
&= m s_{i(1)} \cdots s_{i(k-1)} s_{i(k)} s_{i(k+1)} \cdots s_{i(m)}.
\end{aligned}
$$

Thus we have $\sum_{i=1}^{n} \partial_i^* \partial_i p = mp$.

12. (*ii*) We have to show that $\tau(\xi p(x)) = \tau \otimes \tau(\partial p(x))$ for all $p(x) \in \mathbb{C}\langle x \rangle$. By linearity, it suffices to treat the cases $p(x) = U_m(x)$ for all $m \geq 0$. So fix such an m. Thus we have to show

$$
\sum_{n \geq 1} \alpha_n \tau(C_n(x) U_m(x)) = \tau \otimes \tau(\partial U_m(x)).
$$

For the left-hand side, we have

$$
\begin{aligned}
\sum_n \alpha_n \tau(C_n U_m) &= \sum_{n \leq m} \alpha_n \big(\tau(U_{n+m}) + \tau(U_{m-n})\big) + \alpha_{m+1} \tau(U_{2m+1}) \\
&\qquad + \sum_{n \geq m+2} \alpha_n \big(\tau(U_{n+m}) - \tau(U_{n-m-2})\big) \\
&= \sum_{n \leq m} \alpha_n \big(\alpha_{n+m+1} + \alpha_{m-n+1}\big) + \alpha_{m+1} \alpha_{2m+2} \\
&\qquad + \sum_{n \geq m+2} \alpha_n \big(\alpha_{n+m+1} - \alpha_{n-m-1}\big) \\
&= \sum_n \alpha_n \alpha_{n+m+1} - \sum_{n \geq m+2} \alpha_n \alpha_{n-m-1} + \sum_{n \leq m} \alpha_n \alpha_{m-n+1}.
\end{aligned}
$$

But the first two sums cancel, and thus we remain with exactly the same as in

$$
\tau \otimes \tau(\partial U_m(x)) = \sum_{k=0}^{m-1} \tau(U_k) \tau(U_{m-k-1}) = \sum_{k=0}^{m-1} \alpha_{k+1} \alpha_{m-k}.
$$

For the relevance of this in the context of Schwinger-Dyson equations, see [130].

12.9 Solutions to exercises in Chapter 9

2. We have

$$
E_{\mathcal{B}}(x d_1 \cdots x d_{n-1} x) = \sum_{\pi \in NC(n)} \kappa_\pi^{\mathcal{B}}(x d_1, \ldots, x d_{n-1}, x).
$$

Note that the assumption implies that also all $\kappa_\pi^{\mathcal{B}}(xd_1,\ldots,xd_{n-1},x)$ for $\pi \in NC(n)$ are in $\mathcal{D}$. Applying $E_{\mathcal{D}}$ to the equation above gives thus

$$E_{\mathcal{D}}(xd_1\cdots xd_{n-1}x) = \sum_{\pi\in NC(n)} \kappa_\pi^{\mathcal{B}}(xd_1,\ldots,xd_{n-1},x).$$

If we compare this with the moment-cumulant formula on the $\mathcal{D}$-level,

$$E_{\mathcal{D}}(xd_1\cdots xd_{n-1}x) = \sum_{\pi\in NC(n)} \kappa_\pi^{\mathcal{D}}(xd_1,\ldots,xd_{n-1},x),$$

then we get the equality of the $\mathcal{B}$-valued and the $\mathcal{D}$-valued cumulants by induction.

12.10 Solutions to exercises in Chapter 10

2. Note that in general

$$\begin{aligned}\mathbb{H}^+(M_n(\mathbb{C})) &= \{B\in M_n(\mathbb{C}) \mid \exists\epsilon>0:\ \mathrm{Im}(B)\geq \epsilon 1\}\\ &= \{B\in M_n(\mathbb{C}) \mid \mathrm{Im}(B) \text{ is positive definite}\}.\end{aligned}$$

(i) Recall that any self-adjoint matrix

$$\begin{pmatrix}\alpha & \beta\\ \overline{\beta} & \gamma\end{pmatrix} \in M_2(\mathbb{C})$$

is positive definite if and only if $\alpha>0$ and $\alpha\gamma-|\beta|^2>0$.

Now, for

$$B=\begin{pmatrix} b_{11} & b_{12}\\ b_{21} & b_{22}\end{pmatrix}\in M_2(\mathbb{C}) \qquad \text{we have} \qquad \mathrm{Im}(B)=\begin{pmatrix}\mathrm{Im}(b_{11}) & \frac{1}{2i}(b_{12}-\overline{b_{21}})\\ \frac{1}{2i}(b_{21}-\overline{b_{12}}) & \mathrm{Im}(b_{22})\end{pmatrix}.$$

Hence $\mathrm{Im}(B)$ is positive definite, if and only if

$$\mathrm{Im}(b_{11})>0 \qquad \text{and} \qquad \mathrm{Im}(b_{11})\mathrm{Im}(b_{22})-\frac{1}{4}|b_{12}-\overline{b_{21}}|^2>0.$$

(ii) Assume that $\lambda\in\mathbb{C}$ is an eigenvalue of $B\in\mathbb{H}^+(M_n(\mathbb{C}))$. We want to show that $\mathrm{Im}(\lambda)>0$. Let $\eta\in\mathbb{C}^n$ with $\|\eta\|=1$ be a corresponding eigenvector of B, i.e. $B\eta=\lambda\eta$. Since $\mathrm{Im}(B)$ is positive definite, it follows

$$0<\langle\mathrm{Im}(B)\eta,\eta\rangle=\frac{1}{2i}\big(\langle B\eta,\eta\rangle-\langle B^*\eta,\eta\rangle\big)=\frac{1}{2i}\big(\langle B\eta,\eta\rangle-\langle \eta,B\eta\rangle\big)=\mathrm{Im}(\lambda),$$

as desired.

The converse is not true as shown by the following counterexample for $n = 2$. Take a matrix of the form

$$B = \begin{pmatrix} \lambda_1 & \rho \\ 0 & \lambda_2 \end{pmatrix}$$

with $\mathrm{Im}(\lambda_1) > 0$, $\mathrm{Im}(\lambda_2) > 0$ and some $\rho \in \mathbb{C}$. B satisfies the condition that all its eigenvalues belong to the upper half-plane $\mathbb{C}^+$. However, if in addition $|\rho| \geq 2\sqrt{\mathrm{Im}(\lambda_1)\mathrm{Im}(\lambda_2)}$ holds, it cannot belong to $\mathbb{H}^+(M_2(\mathbb{C}))$, since the second characterizing condition of $\mathbb{H}^+(M_2(\mathbb{C}))$, $\mathrm{Im}(b_{11})\mathrm{Im}(b_{22}) > |b_{12} - \overline{b_{21}}|^2/4$, is violated.

12.11 Solutions to exercises in Chapter 11

1. We shall show that while $\nabla^2 \log|z| = 0$ as a function, $\nabla^2 \log|z| = 2\pi\delta_0$ as a distribution, where δ_0 is the distribution which evaluates a test function at $(0, 0)$. In other words, $G(z, w) = \frac{1}{2\pi} \log|z - w|$ is the Green function of the the Laplacian on $\mathbb{R}^2$. To see what this means, first note that by writing $\log|z|\, dxdy = r \log r\, drd\theta$, where (r, θ) are polar coordinates, we see that $\log|z|$ is a locally integrable function on $\mathbb{R}^2$. Thus it determines (see Rudin [152, Ch. 6]) a distribution

$$f \mapsto \iint_{\mathbb{R}^2} f(x, y) \log\sqrt{x^2 + y^2}\, dxdy$$

where f is a test function, i.e. a C^∞-function with compact support. By definition, the Laplacian of this distribution, $\nabla^2 \log|z|$, is the distribution

$$f \mapsto \iint_{\mathbb{R}^2} \nabla^2 f(x, y) \log\sqrt{x^2 + y^2}\, dxdy.$$

Hence our claim is that for a test function f

$$\iint_{\mathbb{R}^2} \nabla^2 f(x, y) \log\sqrt{x^2 + y^2}\, dxdy = 2\pi f(0, 0).$$

We denote the gradient of f by $\nabla f = (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y})$ and the divergence of a vector field F by $\nabla \cdot F$. Let $D_r = \{(x, y) \mid \sqrt{x^2 + y^2} < r\}$, and

$$D_{r,R} = \{(x, y) \mid r < \sqrt{x^2 + y^2} < R\}.$$

We proceed in three steps.

(*i*) Let f, g be C^2-functions on $\mathbb{R}^2$; then

$$\nabla \cdot f\nabla g = \frac{\partial f}{\partial x}\frac{\partial g}{\partial x} + f\frac{\partial^2 g}{\partial x^2} + \frac{\partial f}{\partial y}\frac{\partial g}{\partial y} + f\frac{\partial^2 g}{\partial y^2} = f\nabla^2 g + \nabla f \cdot \nabla g$$

so that

$$f\nabla^2 g - g\nabla^2 f = \nabla\cdot(f\nabla g - g\nabla f).$$

(*ii*) Let $g(x,y) = \log\sqrt{x^2+y^2}$ and f be a test function. Choose R large enough so that $\operatorname{supp}(f)\subset D_R$. We show that for all $0<r<R$

$$\iint_{D_{r,R}} \nabla^2 f(x,y)\,\log\sqrt{x^2+y^2}\,dxdy = \int_{\partial D_r}\Big(\frac{1}{r}f - \log r\,\frac{\partial f}{\partial r}\Big)ds.$$

Let D be an open connected region in $\mathbb{R}^2$ and ∂D its boundary. Suppose that ∂D is the union of a finite number of Jordan curves which do not intersect each other. Green's theorem asserts that for a vector field F

$$\iint_D \nabla\cdot F(x,y)\,dxdy = \int_{\partial D} F\cdot\mathbf{n}\,ds$$

where $\mathbf{n}$ is the outward pointing unit normal of ∂D. So in particular, if we let $F=\nabla f$, we have

$$\iint_D \nabla^2 f(x,y)\,dxdy = \int_{\partial D}\nabla f\cdot\mathbf{n}\,ds.$$

By assumption both f and ∇f vanish on ∂D_R, and by our earlier observation that $\log|z|$ is harmonic, $\nabla^2 g = 0$ on $D_{r,R}$. Hence

$$\begin{aligned}\iint_{D_{r,R}} \nabla^2 f(x,y)\,\log\sqrt{x^2+y^2}dxdy &= -\iint_{D_{r,R}}(f\nabla^2 g - g\nabla^2 f)\,dxdy\\ &= -\iint_{D_{r,R}}\nabla\cdot(f\nabla g - g\nabla f)\,dxdy\\ &= \int_{\partial D_r}(f\nabla g - g\nabla f)\cdot\mathbf{n}\,ds\\ &\quad -\int_{\partial D_R}(f\nabla g - g\nabla f)\cdot\mathbf{n}\,ds\\ &= \int_{\partial D_r}(f\nabla g - g\nabla f)\cdot\mathbf{n}\,ds.\end{aligned}$$

Now $\nabla g = (x,y)/(x^2+y^2)$ and on ∂D_r we have $\mathbf{n} = (x,y)/\sqrt{x^2+y^2}$, so $\nabla g\cdot\mathbf{n} = 1/r$. Also $g=\log r$ on ∂D_r, and on D_r, $\nabla f\cdot\mathbf{n} = \frac{\partial f}{\partial r}$, by the chain rule. Thus

$$\iint_{D_{r,R}} \nabla^2 f(x,y)\,\log\sqrt{x^2+y^2}dxdy = \frac{1}{r}\int_{\partial D_r} f\,ds - \log r\int_{\partial D_r}\frac{\partial f}{\partial r}\,ds.$$

(*iii*) Finally we show that for a test function f

$$\iint_{\mathbb{R}^2} \nabla^2 f(x,y) \log \sqrt{x^2+y^2}\, dxdy = 2\pi f(0,0).$$

To calculate the integrals above, let us parameterize ∂D_r with $x(\theta) = r\cos\theta$ and $y(\theta) = r\sin\theta$. Then $ds = \sqrt{x'(\theta)^2 + y'(\theta)^2}\, d\theta = r\, d\theta$. So

$$\frac{1}{r}\int_{\partial D_r} f\, ds = \int_0^{2\pi} f(r\cos\theta, r\sin\theta)\, d\theta$$

which converges to $2\pi f(0,0)$ as $r \to 0$. Also

$$\log r \int_{\partial D_r} \frac{\partial f}{\partial r}\, ds = r \log r \int_0^{2\pi} \frac{\partial f}{\partial r}(r\cos\theta, r\sin\theta)\, d\theta.$$

Now as $r \to 0$, $\int_0^{2\pi} \frac{\partial f}{\partial r}(r\cos\theta, r\sin\theta)\, d\theta$ converges to $2\pi \frac{\partial f}{\partial r}(0,0)$ and $r\log r$ converges to 0. Thus

$$\begin{aligned}\iint_{\mathbb{R}^2} \nabla^2 f(x,y) \log \sqrt{x^2+y^2} dxdy &= \iint_{D_R} \nabla^2 f(x,y) \log \sqrt{x^2+y^2} dxdy \\ &= \lim_{r\to 0} \iint_{D_{r,R}} \nabla^2 f(x,y) \log \sqrt{x^2+y^2} dxdy \\ &= 2\pi f(0,0)\end{aligned}$$

as claimed.

4. Let us put

$$A := \begin{pmatrix} 0 & a \\ a^* & 0 \end{pmatrix}, \qquad \Lambda := \begin{pmatrix} 0 & \lambda \\ \bar{\lambda} & 0 \end{pmatrix}.$$

Note that both A and Λ are self-adjoint and have with respect to $\mathrm{tr} \otimes \tau$ the distributions $\tilde{\mu}_{|a|}$ and $(\delta_\alpha + \delta_{-\alpha})/2$, respectively, and that $A - \Lambda$ has the distribution $\tilde{\mu}_{|a-\lambda|}$. (It is of course important that we are in a tracial setting, so that aa^* and a^*a have the same distribution.)

It remains to show that A and Λ are free with respect to $\mathrm{tr} \otimes \tau$. For this note that the kernel of $\mathrm{tr} \otimes \tau$ on the unital algebra generated by A is spanned by matrices of the form

$$\begin{pmatrix} 0 & (aa^*)^{k-1}a \\ (a^*a)^{k-1}a^* & 0 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} (aa^*)^k - \tau((aa^*)^k) & 0 \\ 0 & (a^*a)^k - \tau((a^*a)^k) \end{pmatrix} \tag{12.7}$$

for some $k \geq 1$, whereas the kernel of $\mathrm{tr} \otimes \tau$ on the algebra generated by Λ is just spanned by the off-diagonal matrices of the form

$$\begin{pmatrix} 0 & |\lambda|^k \lambda \\ |\lambda|^k \bar{\lambda} & 0 \end{pmatrix} = |\lambda|^k \Lambda$$

for some $k \geq 1$. Hence we have to check that we have

$$\mathrm{tr} \otimes \tau[A_1 \Lambda A_2 \Lambda \cdots A_n \Lambda] = 0 \qquad \text{and} \qquad \mathrm{tr} \otimes \tau[A_1 \Lambda A_2 \Lambda \cdots A_n] = 0,$$

for all n and all choices of $A_1, \dots, A_n$ from the collection (12.7). Multiplication with Λ has on the A_i the effect that we get matrices from the collection

$$\begin{pmatrix} (aa^*)^{k-1}a & 0 \\ 0 & (a^*a)^{k-1}a^* \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} 0 & (aa^*)^k - \tau((aa^*)^k) \\ (a^*a)^k - \tau((a^*a)^k) & 0 \end{pmatrix}. \tag{12.8}$$

Hence, we have to see that whenever we multiply matrices from the collection (12.8) in any order, we get only matrices where all entries vanish under the application of τ. Let us denote the non-trivial entries in the matrices from (12.8) as follows:

$$p_{11}^k := (aa^*)^{k-1}a, \qquad p_{12}^k := (aa^*)^k - \tau((aa^*)^k),$$
$$p_{21}^k := (a^*a)^k - \tau((a^*a)^k), \qquad p_{22}^k := (a^*a)^{k-1}a^*.$$

With this notation we have to show that $\tau(p_{i_1 i_2}^{k_1} p_{i_2 i_3}^{k_2} \cdots p_{i_{n-1} i_n}^{k_{n-1}} p_{i_n i_{n+1}}^{k_n}) = 0$ for all $n, k \geq 1$ and all $i_1, \dots, i_{n+1} \in \{1, 2\}$. Now we use the fact that an R-diagonal element a has the property that its $*$-distribution is invariant under the multiplication with a free Haar unitary; this means we can replace a by au, where u is a Haar unitary which is $*$-free from a. But then our operators p_{ij}^k go over to $p_{11}^k u$, p_{12}^k, $u^* p_{21}^k u$ and $u^* p_{22}^k$. If we multiply those elements as required, then we always get words which are alternating in factors from the p_{ij}^k and $\{u, u^*\}$; all those factors are centred; hence, by the $*$-freeness between a and u, the whole product is centred.

For more details, see also [138].

References

1. L. Aagaard, U. Haagerup, Moment formulas for the quasi-nilpotent DT-operator. Int. J. Math. **15**(6), 581–628 (2004)
2. G. Akemann, J. Baik, P. Di Francesco, *The Oxford Handbook of Random Matrix Theory* (Oxford University Press, Oxford, 2011)
3. N.I. Akhiezer, *The Classical Moment Problem and Some Related Questions in Analysis*. Translated by N. Kemmer (Hafner Publishing Co., New York, 1965)
4. N.I. Akhiezer, I.M. Glazman, *Theory of Linear Operators in Hilbert Space. Vol. II* (Pitman (Advanced Publishing Program), Boston, MA/London, 1981)
5. G.W. Anderson, Convergence of the largest singular value of a polynomial in independent Wigner matrices. Ann. Probab. **41**(3B), 2103–2181 (2013)
6. G.W. Anderson, O. Zeitouni, A CLT for a band matrix model. Probab. Theory Relat. Fields **134**(2), 283–338 (2006)
7. G.W. Anderson, A. Guionnet, O. Zeitouni, *An Introduction to Random Matrices*. Cambridge Studies in Advanced Mathematics, vol. 118 (Cambridge University Press, Cambridge, 2010)
8. M. Anshelevich, Free martingale polynomials. J. Funct. Anal. **201**(1), 228–261 (2003)
9. M. Anshelevich, S.T. Belinschi, M. Bożejko, F. Lehner, Free infinite divisibility for g-Gaussians. Math. Res. Lett. **17**(5), 905–916 (2010)
10. O. Arizmendi, T. Hasebe, N. Sakuma, On the law of free subordinators. ALEA, Lat. Am. J. Probab. Math. Stat. **10**(1), 271–291 (2013)
11. O. Arizmendi, T. Hasebe, F. Lehner, C. Vargas, Relations between cumulants in noncommutative probability. Adv. Math. **282**, 56–92 (2015)
12. O. Arizmendi, I. Nechita, C. Vargas, On the asymptotic distribution of block-modified random matrices. J. Math. Phys. **57**(1), 015216 (2016)
13. D.H. Armitage, S.J. Gardiner, *Classical Potential Theory* (Springer, London, 2001)
14. Z.D. Bai, J.W. Silverstein, CLT for linear spectral statistics of large-dimensional sample covariance matrices. Ann. Probab. **32**(1A), 553–605 (2004)
15. Z. Bai, J.W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices*. Springer Series in Statistics, 2nd edn. (Springer, New York, 2010)
16. T. Banica, R. Speicher, Liberation of orthogonal Lie groups. Adv. Math. **222**(4), 1461–1501 (2009)
17. T. Banica, S. Teodor Belinschi, M. Capitaine, B. Collins, Free Bessel laws. Can. J. Math **63**(1), 3–37 (2011)
18. S.T. Belinschi, Complex analysis methods in noncommutative probability. Ph.D. thesis, Indiana University, 2005

J.A. Mingo, R. Speicher, *Free Probability and Random Matrices*,
Fields Institute Monographs 35, DOI 10.1007/978-1-4939-6942-5

19. S.T. Belinschi, The Lebesgue decomposition of the free additive convolution of two probability distributions. Probab. Theory Relat. Fields **142**(1–2), 125–150 (2008)
20. S.T. Belinschi, H. Bercovici, A property of free entropy. Pac. J. Math. **211**(1), 35–40 (2003)
21. S.T. Belinschi, H. Bercovici, A new approach to subordination results in free probability. J. d'Anal. Math. **101**(1), 357–365 (2007)
22. S.T. Belinschi, M. Bożejko, F. Lehner, R. Speicher, The normal distribution is $\boxplus$-infinitely divisible. Adv. Math. **226**(4), 3677–3698 (2011)
23. S.T. Belinschi, T. Mai, R. Speicher, Analytic subordination theory of operator-valued free additive convolution and the solution of a general random matrix problem. J. Reine Angew. Math. (2013). Published online: 2015-04-12
24. S.T. Belinschi, P. Śniady, R. Speicher, Eigenvalues of non-hermitian random matrices and Brown measure of non-normal operators: Hermitian reduction and linearization method. arXiv:1506.02017 (2015)
25. S.T. Belinschi, R. Speicher, J. Treilhard, C. Vargas, Operator-valued free multiplicative convolution: analytic subordination theory and applications to random matrix theory. Int. Math. Res. Not. IMRN **2015**(14), 5933–5958 (2015)
26. G. Ben Arous, A. Guionnet, Large deviations for Wigner's law and Voiculescu's non-commutative entropy. Probab. Theory Relat. Fields **108**(4), 517–542 (1997)
27. F. Benaych-Georges, Taylor expansions of R-transforms: application to supports and moments. Indiana Univ. Math. J. **55**(2), 465–481 (2006)
28. F. Benaych-Georges, Rectangular random matrices, related convolution. Probab. Theory Relat. Fields **144**(3–4), 471–515 (2009)
29. H. Bercovici, V. Pata, Stable laws and domains of attraction in free probability theory. Ann. Math. **149**, 1023–1060 (1999)
30. H. Bercovici, D. Voiculescu, Free convolution of measures with unbounded support. Indiana Univ. Math. J. **42**(3), 733–773 (1993)
31. H. Bercovici, D.-V. Voiculescu, Lévy-Hinčin type theorems for multiplicative and additive free convolution. Pac. J. Math. **153**(2), 217–248 (1992)
32. H. Bercovici, D. Voiculescu, Superconvergence to the central limit and failure of the Cramér theorem for free random variables. Probab. Theory Relat. Fields **103**(2), 215–222 (1995)
33. P. Biane, Some properties of crossings and partitions. Discrete Math. **175**(1–3), 41–53 (1997)
34. P. Biane, Processes with free increments. Math. Z. **227**(1), 143–174 (1998)
35. P. Biane, Representations of symmetric groups and free probability. Adv. Math. **138**(1), 126–181 (1998)
36. P. Biane, Entropie libre et algèbres d'opérateurs. Séminaire Bourbaki **43**, 279–300 (2002)
37. P. Biane, Free probability and combinatorics, in *Proc. ICM 2002, Beijing*, vol. II, 2002, pp. 765–774
38. P. Biane, F. Lehner, Computation of some examples of Brown's spectral measure in free probability. Colloq. Math. **90**(2), 181–211 (2001)
39. P. Biane, R. Speicher, Stochastic calculus with respect to free Brownian motion and analysis on Wigner space. Probab. Theory Relat. Fields **112**(3), 373–409 (1998)
40. P. Biane, M. Capitaine, A. Guionnet, Large deviation bounds for matrix Brownian motion. Invent. Math. **152**(2), 433–459 (2003)
41. P. Billingsley, *Probability and Measure*. Wiley Series in Probability and Mathematical Statistics, 3rd edn. (Wiley, New York, 1995). A Wiley-Interscience Publication
42. C. Bordenave, D. Chafaï, Around the circular law. Probab. Surv. **9**, 1–89 (2012)
43. M. Bożejko, On $\Lambda(p)$ sets with minimal constant in discrete noncommutative groups. Proc. Am. Math. Soc. **51**(2), 407–412 (1975)
44. M. Brannan, Approximation properties for free orthogonal and free unitary quantum groups. Journal für die reine und angewandte Mathematik (Crelles J.) **2012**(672), 223–251 (2012)
45. E. Brézin, C. Itzykson, G. Parisi, J.-B. Zuber, Planar diagrams. Commun. Math. Phys. **59**(1), 35–51 (1978)

46. L.G. Brown, Lidskiĭ's theorem in the type II case, in *Geometric Methods in Operator Algebras (Kyoto, 1983)*. Pitman Research Notes in Mathematics Series, vol. 123 (Longman Science Technology, Harlow, 1986), pp. 1–35
47. T. Cabanal-Duvillard, Fluctuations de la loi empirique de grandes matrices aléatoires. Ann. Inst. H. Poincaré Probab. Statist. **37**(3), 373–402 (2001)
48. M. Capitaine, C. Donati-Martin, Strong asymptotic freeness for Wigner and Wishart matrices. Indiana Univ. Math. J. **56**(2), 767–803 (2007)
49. D.I. Cartwright, P.M. Soardi et al., Random walks on free products, quotients and amalgams. Nagoya Math. J. **102**, 163–180 (1986)
50. G. Cébron, A. Dahlqvist, C. Male, Universal constructions for spaces of traffics. arXiv 1601.00168, January 2016
51. I. Charlesworth, D. Shlyakhtenko, Free entropy dimension and regularity of non-commutative polynomials. J. Funct. Anal. **271**(8), 2274–2292 (2016)
52. I. Charlesworth, B. Nelson, P. Skoufranis, Combinatorics of bi-freeness with amalgamation. Commun. Math. Phys. **338**(2), 801–847 (2015)
53. G.P. Chistyakov, F. Götze, Limit theorems in free probability theory. I. Ann. Probab., 54–90 (2008)
54. G. Chistyakov, F. Götze, The arithmetic of distributions in free probability theory. Open Math. **9**(5), 997–1050 (2011)
55. K.L. Chung, *A Course in Probability Theory*, 3rd edn. (Academic, San Diego, CA, 2001)
56. P.M. Cohn, *Free Ideal Rings and Localization in General Rings.* (Cambridge University Press, Cambridge, 2006)
57. B. Collins, Moments and cumulants of polynomial random variables on unitary groups, the Itzykson-Zuber integral, and free probability. Int. Math. Res. Not. **2003**(17), 953–982 (2003)
58. B. Collins, C. Male, Liberté asymptotique forte pour les matrices de Haar et les matrices déterministes. Ann. Sci. Éc. Norm. Supér. (4) **47**(1), 147–163 (2014)
59. B. Collins, P. Śniady, Integration with respect to the Haar measure on unitary, orthogonal and symplectic group. Commun. Math. Phys. **264**(3), 773–795 (2006)
60. B. Collins, J.A. Mingo, P. Śniady, R. Speicher, Second order freeness and fluctuations of random matrices. III. Higher order freeness and free cumulants. Doc. Math. **12**, 1–70 (2007) (electronic)
61. R. Cori, *Un code pour les graphes planaires et ses applications* (Société Mathématique de France, Paris, 1975). With an English abstract, Astérisque, No. 27
62. R. Couillet, M. Debbah, *Random Matrix Methods for Wireless Communications* (Cambridge University Press, Cambridge, 2011)
63. Y. Dabrowski, A note about proving non-Γ under a finite non-microstates free Fisher information assumption. J. Funct. Anal. **258**(11), 3662–3674 (2010)
64. A. Dembo, O. Zeitouni, *Large Deviations Techniques and Applications*. Stochastic Modelling and Applied Probability, vol. 38 (Springer, Berlin, 2010). Corrected reprint of the second (1998) edition
65. P.L. Duren, *Univalent Functions*. Grundlehren der Mathematischen Wissenschaften (Fundamental Principles of Mathematical Sciences) , vol. 259 (Springer, New York, 1983)
66. K. Dykema, On certain free product factors via an extended matrix model. J. Funct. Anal. **112**(1), 31–60 (1993)
67. K. Dykema, Interpolated free group factors. Pac. J. Math. **163**(1), 123–135 (1994)
68. K.J. Dykema, Multilinear function series and transforms in free probability theory. Adv. Math. **208**(1), 351–407 (2007)
69. A. Edelman, N. Raj Rao, Random matrix theory. Acta Numer. **14**, 233–297 (2005)
70. W. Ejsmont, U. Franz, K. Szpojankowski, Convolution, subordination and characterization problems in noncommutative probability. Ind. Univ. Math. J. **66**(1), 237–257 (2017)
71. J. Feinberg, A. Zee, Non-Hermitian random matrix theory: method of Hermitian reduction. Nuclear Phys. B **504**(3), 579–608 (1997)
72. V. Féray, P. Śniady, Asymptotics of characters of symmetric groups related to Stanley character formula. Ann. Math. **173**(2), 887–906 (2011)

73. A. Freslon, M. Weber, On the representation theory of partition (easy) quantum groups. J. Reine Angew. Math. **720**, 155–197 (2016)
74. R.M. Friedrich, J. McKay, Homogeneous Lie groups and quantum probability. arXiv 1506.07089, June 2015
75. B. Fuglede, R.V. Kadison, Determinant theory in finite factors. Ann. Math. (2) **55**, 520–530 (1952)
76. L. Ge, Applications of free entropy to finite von Neumann algebras. II. Ann. Math. (2) **147**(1), 143–157 (1998)
77. V.L. Girko, The circular law. Teor. Veroyatnost. i Primenen. **29**(4), 669–679 (1984)
78. V.L. Girko, *Theory of Stochastic Canonical Equations. Vol. I.* Mathematics and Its Applications, vol. 535 (Kluwer Academic Publishers, Dordrecht, 2001)
79. R.L. Graham, D.E. Knuth, O. Patashnik, *Concrete Mathematics: A Foundation for Computer Science*, 2nd edn. (Addison-Wesley Publishing Company, Reading, MA, 1994)
80. D.S. Greenstein, On the analytic continuation of functions which map the upper half plane into itself. J. Math. Anal. Appl. **1**, 355–362 (1960)
81. Y. Gu, H.W. Huang, J.A. Mingo, An analogue of the Lévy-Hinčin formula for bi-free infinitely divisible distributions. Ind. Univ. Math. J. **65**(5), 1795–1831 (2016)
82. A. Guionnet, M. Krishnapur, O. Zeitouni, The single ring theorem. Ann. Math. (2) **174**(2), 1189–1217 (2011)
83. A. Guionnet, D. Shlyakhtenko, Free monotone transport. Invent. Math. **197**(3), 613–661 (2014)
84. U. Haagerup, Random matrices, free probability and the invariant subspace problem relative to a von Neumann algebra. Proc. ICM 2002, Beijing **I**, 273–290 (2002)
85. U. Haagerup, F. Larsen, Brown's spectral distribution measure for R-diagonal elements in finite von Neumann algebras. J. Funct. Anal. **176**(2), 331–367 (2000)
86. U. Haagerup, H. Schultz, Brown measures of unbounded operators affiliated with a finite von Neumann algebra. Math. Scand. **100**(2), 209–263 (2007)
87. U. Haagerup, H. Schultz, Invariant subspaces for operators in a general II_1-factor. Publ. Math. Inst. Hautes Études Sci. **109**(1), 19–111 (2009)
88. U. Haagerup, S. Thorbjørnsen, A new application of random matrices: $\mathrm{Ext}(C^*_{\mathrm{red}}(F_2))$ is not a group. Ann. Math. (2) **162**(2), 711–775 (2005)
89. U. Haagerup, H. Schultz, S. Thorbjørnsen, A random matrix approach to the lack of projections in $C^*_{\mathrm{red}}(\mathbb{F}_2)$. Adv. Math. **204**(1), 1–83 (2006)
90. W. Hachem, P. Loubaton, J. Najim, Deterministic equivalents for certain functionals of large random matrices. Ann. Appl. Probab. **17**(3), 875–930 (2007)
91. J. Harer, D. Zagier, The Euler characteristic of the moduli space of curves. Invent. Math. **85**(3), 457–485 (1986)
92. W.K. Hayman, P.B. Kennedy, *Subharmonic Functions. Vol. I.* London Mathematical Society Monographs, vol. 9 (Academic, London/New York, 1976)
93. J.W. Helton, S.A. McCullough, V. Vinnikov, Noncommutative convexity arises from linear matrix inequalities. J. Funct. Anal. **240**(1), 105–191 (2006)
94. J.W. Helton, R. Rashidi Far, R. Speicher, Operator-valued semicircular elements: solving a quadratic matrix equation with positivity constraints. Int. Math. Res. Not. IMRN **2007**(22), 15 (2007). Art. ID rnm086
95. J.W. Helton, T. Mai, R. Speicher, Applications of realizations (aka linearizations) to free probability. arXiv preprint arXiv:1511.05330, 2015
96. F. Hiai, D. Petz, Asymptotic freeness almost everywhere for random matrices. Acta Sci. Math. (Szeged). **66**(3–4), 809–834 (2000)
97. F. Hiai, D. Petz, *The Semicircle Law, Free Random Variables and Entropy*. Mathematical Surveys and Monographs, vol. 77 (American Mathematical Society, Providence, RI, 2000)
98. G. Higman, The units of group-rings. Proc. Lond. Math. Soc. **2**(1), 231–248 (1940)
99. K. Hoffman, *Banach Spaces of Analytic Functions* (Dover Publications, Inc., New York, 1988)

100. H.-W. Huang, J.-C. Wang, Analytic aspects of the bi-free partial R-transform. J. Funct. Anal. **271**(4), 922–957 (2016)
101. L. Isserlis, On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. Biometrika **12**(1/2), 134–139 (1918)
102. A. Jacques, Sur le genre d'une paire de substitutions. C. R. Acad. Sci. Paris Sér. A-B **267**, A625–A627 (1968)
103. R.A. Janik, M.A. Nowak, G. Papp, I. Zahed, Non-Hermitian random matrix models. Nuclear Phys. B **501**(3), 603–642 (1997)
104. K. Johansson, On fluctuations of eigenvalues of random Hermitian matrices. Duke Math. J. **91**(1), 151–204 (1998)
105. V.F.R. Jones, Index for subfactors. Invent. Math. **72**(1), 1–25 (1983)
106. R.V. Kadison, J.R. Ringrose, *Fundamentals of the Theory of Operator Algebras,* vol. II (American Mathematical Society, Providence, 1997)
107. D.S. Kaliuzhnyi-Verbovetskyi, V. Vinnikov, *Foundations of Free Noncommutative Function Theory*. Mathematical Surveys and Monographs, vol. 199 (American Mathematical Society, Providence, RI, 2014)
108. T. Kemp, I. Nourdin, G. Peccati, R. Speicher, Wigner chaos and the fourth moment. Ann. Probab. **40**(4), 1577–1635 (2012)
109. P. Koosis, *Introduction to H_p Spaces*. Cambridge Tracts in Mathematics, vol. 115, 2nd edn. (Cambridge University Press, Cambridge, 1998).
110. C. Köstler, R. Speicher, A noncommutative de Finetti theorem: invariance under quantum permutations is equivalent to freeness with amalgamation. Commun. Math. Phys. **291**(2), 473–490 (2009)
111. B. Krawczyk, R. Speicher, Combinatorics of free cumulants. J. Combin. Theory Ser. A **90**(2), 267–292 (2000)
112. M. Krbalek, P. Šeba, P. Wagner, Headways in traffic flow: remarks from a physical perspective. Phys. Rev. E **64**(6), 066119 (2001)
113. G. Kreweras, Sur les partitions non croisées d'un cycle. Discr. Math. **1**(4), 333–350 (1972)
114. B. Kümmerer, R. Speicher, Stochastic integration on the Cuntz algebra O_∞. J. Funct. Anal. **103**(2), 372–408 (1992)
115. T. Kusalik, J.A. Mingo, R. Speicher, Orthogonal polynomials and fluctuations of random matrices. J. Reine Angew. Math. **604**, 1–46 (2007)
116. F. Larsen, Brown measures and R-diagonal elements in finite von Neumann algebras. Ph.D. thesis, University of Southern Denmark, 1999
117. F. Lehner, Cumulants in noncommutative probability theory I. Noncommutative exchangeability systems. Math. Zeitschrift **248**(1), 67–100 (2004)
118. F. Lemeux, P. Tarrago, Free wreath product quantum groups: the monoidal category, approximation properties and free probability. J. Funct. Anal. **270**(10), 3828–3883 (2016)
119. E. Lukacs, *Characteristic Functions*. Griffin's Statistical Monographs & Courses, vol. 5 (Hafner Publishing Co., New York, 1960)
120. H. Maassen, Addition of freely independent random variables. J. Funct. Anal. **106**(2), 409–438 (1992)
121. T. Mai, R. Speicher, M. Weber, Absence of algebraic relations and of zero divisors under the assumption of full non-microstates free entropy dimension. Adv. Math. **304**, 1080–1107 (2017)
122. C. Male, The distributions of traffics and their free product. arXiv:1111.4662, 2015
123. V.A. Marchenko, L.A. Pastur, Distribution of eigenvalues for some sets of random matrices. Mat. Sbornik **114**(4), 507–536 (1967)
124. A. Marcus, D.A. Spielman, N. Srivastava, Finite free convolutions of polynomials. arXiv:1504.00350, 2015
125. J.C. McLaughlin, Random walks and convolution operators on free products. Ph.D. thesis, New York University, 1986
126. J.A. Mingo, A. Nica, Annular noncrossing permutations and partitions, and second-order asymptotics for random matrices. Int. Math. Res. Not. **2004**(28), 1413–1460 (2004)

127. J.A. Mingo, M. Popa, Real second order freeness and Haar orthogonal matrices. J. Math. Phys. **54**(5), 051701, 35 (2013)
128. J.A. Mingo, R. Speicher, Second order freeness and fluctuations of random matrices. I. Gaussian and Wishart matrices and cyclic Fock spaces. J. Funct. Anal. **235**(1), 226–270 (2006)
129. J.A. Mingo, R. Speicher, Sharp bounds for sums associated to graphs of matrices. J. Funct. Anal. **262**(5), 2272–2288 (2012)
130. J.A. Mingo, R. Speicher, Schwinger-Dyson equations: classical and quantum. Probab. Math. Stat. **33**(2), 275–285 (2013)
131. J.A. Mingo, P. Śniady, R. Speicher, Second order freeness and fluctuations of random matrices. II. Unitary random matrices. Adv. Math. **209**(1), 212–240 (2007)
132. J.A. Mingo, R. Speicher, E. Tan, Second order cumulants of products. Trans. Am. Math. Soc. **361**(9), 4751–4781 (2009)
133. R. Miranda, Algebraic curves and Riemann surfaces. *Graduate Studies in Mathematics*, vol. 5 (American Mathematical Society, Providence, 1995)
134. F.J. Murray, J. Von Neumann, On rings of operators. Ann. Math. (2) **37**(1), 116–229 (1936)
135. A. Nica, R-transforms of free joint distributions and non-crossing partitions. J. Funct. Anal. **135**(2), 271–296 (1996)
136. A. Nica, R. Speicher, R-diagonal pairs—a common approach to Haar unitaries and circular elements, in *Free Probability Theory (Waterloo, ON, 1995)*. Fields Institute Communications, vol. 12 (American Mathematical Society, Providence, RI, 1997), pp. 149–188
137. A. Nica, R. Speicher, *Lectures on the Combinatorics of Free Probability*. London Mathematical Society Lecture Note Series, vol. 335 (Cambridge University Press, Cambridge, 2006)
138. A. Nica, D. Shlyakhtenko, R. Speicher, R-diagonal elements and freeness with amalgamation. Can. J. Math. **53**(2), 355–381 (2001)
139. A. Nica, D. Shlyakhtenko, R. Speicher, Operator-valued distributions. I. Characterizations of freeness. Int. Math. Res. Not. **2002**(29), 1509–1538 (2002)
140. A. Nica, D. Shlyakhtenko, R. Speicher, R-cyclic families of matrices in free probability. J. Funct. Anal. **188**(1), 227–271 (2002)
141. J. Novak, Three lectures on free probability, in *Random Matrix Theory, Interacting Particle Systems and Integrable Systems* (Cambridge University Press, New York, NY, 2014), pp. 309–383
142. J. Novak, P. Śniady, What is ... a free cumulant? Not. AMS **58**(2), 300–301 (2011)
143. M. Pimsner, D. Voiculescu, K-groups of reduced crossed products by free groups. J. Oper. Theory **8**(1), 131–156 (1982)
144. F. Rădulescu, The fundamental group of the von Neumann algebra of a free group with infinitely many generators is $\mathbb{R}_+\backslash\{0\}$. J. Am. Math. Soc. **5**(3), 517–532 (1992)
145. F. Rădulescu, Random matrices, amalgamated free products and subfactors of the von Neumann algebra of a free group, of noninteger index. Invent. Math., **115**(2), 347–389 (1994)
146. N. Raj Rao, R. Speicher, Multiplication of free random variables and the S-transform: the case of vanishing mean. Electron. Commun. Probab. **12**, 248–258 (2007)
147. R. Rashidi Far, T. Oraby, W. Bryc, R. Speicher, On slow-fading MIMO systems with nonseparable correlation. IEEE Trans. Inf. Theory **54**(2), 544–553 (2008)
148. S. Raum, M. Weber, The full classification of orthogonal easy quantum groups. Commun. Math. Phys. **341**(3), 751–779 (2016)
149. C.E.I. Redelmeier, Real second-order freeness and the asymptotic real second-order freeness of several real matrix models. Int. Math. Res. Not. **2014**(12), 3353–3395 (2014)
150. M. Reed, B. Simon, *Methods of Modern Mathematical Physics. I: Functional Analysis* (Academic, New York/London, 1980)
151. W. Rudin, *Real and Complex Analysis*, 3rd edn. (McGraw-Hill, New York, NY, 1987)
152. W. Rudin, *Functional Analysis* International Series in Pure and Applied Mathematics, 2nd edn. (McGraw-Hill, Inc., New York, 1991)
153. M. Schürmann, S. Voß, Schoenberg correspondence on dual groups. Commun. Math. Phys. **328**(2), 849–865 (2014)

154. M.P. Schützenberger. On the definition of a family of automata. Inf. Control **4**, 245–270 (1961)
155. D. Shlyakhtenko, Random Gaussian band matrices and freeness with amalgamation. Int. Math. Res. Not. **1996**(20), 1013–1025 (1996)
156. D. Shlyakhtenko, Free probability, planar algebras, subfactors and random matrices. Proc. ICM 2010 Hyderabad **III**, 1603–1623 (2010)
157. J.A. Shohat, J.D. Tamarkin, *The Problem of Moments, Mathematical Surveys. I* (American Mathematical Society, New York, 1943)
158. P. Śniady, R. Speicher, Continuous family of invariant subspaces for R–diagonal operators. Invent. Math. **146**(2), 329–363 (2001)
159. R. Speicher, A new example of "independence" and "white noise". Probab. Theory Relat. Fields **84**(2), 141–159 (1990)
160. R. Speicher, Free convolution and the random sum of matrices. Publ. Res. Inst. Math. Sci. **29**(5), 731–744 (1993)
161. R. Speicher, Multiplicative functions on the lattice of non-crossing partitions and free convolution. Math. Ann. **298**(1), 611–628 (1994)
162. R. Speicher, Free probability theory and non-crossing partitions. Sém. Lothar. Combin **39**, 38 (1997)
163. R. Speicher, Combinatorial theory of the free product with amalgamation and operator-valued free probability theory. Mem. Am. Math. Soc. **627**, 88 (1998)
164. R. Speicher, Free probability and random matrices. Proc. ICM 2014 **III**, 477–501 (2014)
165. R. Speicher, Free probability theory: and its avatars in representation theory, random matrices, and operator algebras; also featuring: non-commutative distributions. Jahresber. Dtsch. Math.-Ver. **119**(1), 3–30 (2017)
166. R. Speicher, C. Vargas, Free deterministic equivalents, rectangular random matrix models, and operator-valued free probability theory. Rand. Matrices: Theory Appl. **1**(02), 1150008 (2012)
167. R.P. Stanley, *Enumerative Combinatorics*, vol. 1 (Cambridge University Press, Cambridge, 1997)
168. E.M. Stein, G. Weiss, *Introduction to Fourier Analysis on Euclidean Spaces* (Princeton University Press, Princeton, 1971)
169. S.J. Szarek, Nets of Grassmann manifold and orthogonal group, in *Banach Space Theory, Proc. Res. Workshop*, Univ. Iowa 1981 (1982), pp. 169–185
170. M. Takesaki. *Theory of Operator Algebras I* (Springer, Berlin, 2002)
171. T. Tao, *Topics in Random Matrix Theory*, vol. 132 (American Mathematical Society Providence, RI, 2012)
172. G. 't Hooft, A planar diagram theory for strong interactions. Nuclear Phys. B **72**, 461–473 (1974)
173. S. Thorbjørnsen, Mixed moments of Voiculescu's Gaussian random matrices. J. Funct. Anal. **176**(2), 213–246 (2000)
174. A.M. Tulino, S. Verdú, Random matrix theory and wireless communications. Found. Trends Commun. Inf. Theory **1**(1), 184 (2004)
175. C. Vargas Obieta, Free probability theory: deterministic equivalents and combinatorics. Ph.D. thesis, Saarbrücken, Universität des Saarlandes, Diss., 2015
176. D. Voiculescu, Symmetries of some reduced free product C^*-algebras, in *Operator Algebras and Their Connections with Topology and Ergodic Theory (Buşteni, 1983)*. Lecture Notes in Mathematics, vol. 1132 (Springer, Berlin, 1985), pp. 556–588
177. D. Voiculescu, Addition of certain noncommuting random variables. J. Funct. Anal. **66**(3), 323–346 (1986)
178. D. Voiculescu, Multiplication of certain non-commuting random variables. J. Oper. Theory **18**(2), 223–235 (1987)
179. D. Voiculescu, Circular and semicircular systems and free product factors, in *Operator Algebras, Unitary Representations, Enveloping Algebras, and Invariant Theory (Paris, 1989)*. Progress in Mathematics, vol. 92 (Birkhäuser Boston, Boston, MA, 1990), pp. 45–60

180. D. Voiculescu, Limit laws for random matrices and free products. Invent. Math. **104**(1), 201–220 (1991)
181. D. Voiculescu, The analogues of entropy and of Fisher's information measure in free probability theory. I. Commun. Math. Phys. **155**(1), 71–92 (1993)
182. D. Voiculescu, The analogues of entropy and of Fisher's information measure in free probability theory. II. Invent. Math. **118**(3), 411–440 (1994)
183. D. Voiculescu, Free probability theory: Random matrices and von Neumann algebras, in *Proceedings of the International Congress of Mathematicians, ICM*, vol. 94 (1994), pp. 3–11
184. D. Voiculescu, Operations on certain non-commutative operator-valued random variables. Recent advances in operator algebras (Orléans, 1992). Astérisque **232**, 243–275 (1995)
185. D. Voiculescu, The analogues of entropy and of Fisher's information measure in free probability theory. III. The absence of Cartan subalgebras. Geom. Funct. Anal. **6**(1), 172–199 (1996)
186. D. Voiculescu, The analogues of entropy and of Fisher's information measure in free probability theory. IV. Maximum entropy and freeness, in *Free Probability Theory (Waterloo, ON, 1995)*. Fields Institute Communications, vol. 12 (American Mathematical Society, Providence, RI, 1997), pp. 293–302
187. D. Voiculescu, The analogues of entropy and of Fisher's information measure in free probability theory. V. Noncommutative Hilbert transforms. Invent. Math. **132**(1), 189–227 (1998)
188. D. Voiculescu, A strengthened asymptotic freeness result for random matrices with applications to free entropy. Int. Math. Res. Not. **1998**, 41–63 (1998)
189. D. Voiculescu, The analogues of entropy and of Fisher's information measure in free probability theory. VI. Liberation and mutual free information. Adv. Math. **146**(2), 101–166 (1999)
190. D. Voiculescu, The coalgebra of the free difference quotient and free probability. Int. Math. Res. Not. **2000**(2), 79–106 (2000)
191. D. Voiculescu, Lectures on free probability. Lect. Notes Math. **1738**, 279–349 (2000)
192. D. Voiculescu, Free entropy. Bull. Lond. Math. Soc. **34**(3), 257–278 (2002)
193. D. Voiculescu, Free analysis questions. I. Duality transform for the coalgebra of $\partial_{X:B}$. Int. Math. Res. Not. **2004**(16), 793–822 (2004)
194. D.-V. Voiculescu, Aspects of free analysis. Jpn. J. Math. **3**(2), 163–183 (2008)
195. D.-V. Voiculescu, Free analysis questions II: the Grassmannian completion and the series expansions at the origin. Journal für die reine und angewandte Mathematik (Crelles J.) **2010**(645), 155–236 (2010)
196. D.-V. Voiculescu, Free probability for pairs of faces I. Commun. Math. Phys. **332**(3), 955–980 (2014)
197. D.V. Voiculescu, K.J. Dykema, A. Nica, *Free Random Variables: A Noncommutative Probability Approach to Free Products with Applications to Random Matrices, Operator Algebras and Harmonic Analysis on Free Groups*. CRM Monograph Series, vol. 1 (American Mathematical Society, Providence, RI, 1992)
198. D.-V. Voiculescu, N. Stammeier, M. Weber, *Free Probability and Operator Algebras*. EMS Münster Lectures in Mathematics, vol. 1 (European Mathematical Society, Zürich, 2016)
199. J.-C. Wang, Local limit theorems in free probability theory. Ann. Probab. **38**(4), 1492–1506 (2010)
200. G.-C. Wick, The evaluation of the collision matrix. Phys. Rev. **80**(2), 268 (1950)
201. E.P. Wigner, Characteristic vectors of bordered matrices with infinite dimensions. Ann. Math. (2) **62**, 548–564 (1955)
202. J.D. Williams, Analytic function theory for operator-valued free probability. J. Reine Angew. Math. (2013). Published online: 2015-01-20
203. J. Wishart, The generalised product moment distribution in samples from a normal multivariate population. Biometrika **20A**(1–2), 32–52 (1928)

204. W. Woess, Nearest neighbour random walks on free products of discrete groups. Boll. Unione Mat. Ital. VI. Ser. B **5**, 961–982 (1986)
205. W. Woess, *Random Walks on Infinite Graphs and Groups*, vol. 138 (Cambridge University Press, Cambridge, 2000)
206. Y.Q. Yin, P.R. Krishnaiah, A limit theorem for the eigenvalues of product of two random matrices. J. Multivariate Anal. **13**, 489–507 (1983)
207. A. Zvonkin, Matrix integrals and map enumeration: an accessible introduction. Math. Comput. Modell. **26**(8), 281–304 (1997)

Index of Exercises

J.A. Mingo, R. Speicher, *Free Probability and Random Matrices*,
Fields Institute Monographs 35, DOI 10.1007/978-1-4939-6942-5

Chapter 8

Chapter 9

Chapter 10

Chapter 11

Index

Symbols

J.A. Mingo, R. Speicher, *Free Probability and Random Matrices*,
Fields Institute Monographs 35, DOI 10.1007/978-1-4939-6942-5