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

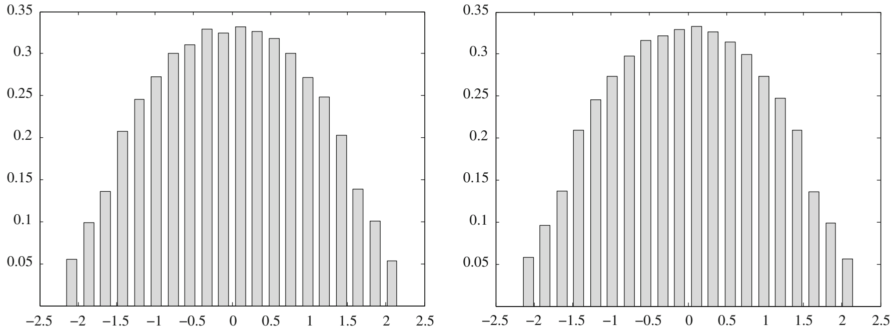


Fig. 9.1 Histogram of the dN eigenvalues of a random matrix X_N, for $N = 1000$, for two different realizations

$$X_N = \frac{1}{\sqrt{3}} \begin{pmatrix} A_N & B_N & C_N \\ B_N & A_N & B_N \\ C_N & B_N & A_N \end{pmatrix}, \tag{9.1}$$

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 $\{c_1, \ldots, c_n\}$ be operators in $(\mathcal{A}, \varphi)$. If $\{\mathrm{Re}(c_1), \mathrm{Im}(c_1), \ldots, \mathrm{Re}(c_n), \mathrm{Im}(c_n)\}$ is a semi-circular family, we say that $\{c_1, \ldots, c_n\}$ 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 $\{c_1, \ldots, c_n\}$ to be a circular family, it is necessary and sufficient that for every $i_1, \ldots, i_m \in [n]$ and every $\epsilon_1, \ldots, \epsilon_m \in \{-1, 1\}$ we have

$$\varphi(c_{i_1}^{(\epsilon_1)} \cdots c_{i_m}^{(\epsilon_m)}) = \sum_{\pi \in NC_2(m)} \kappa_\pi(c_{i_1}^{(\epsilon_1)}, \ldots, c_{i_m}^{(\epsilon_m)}).$$

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

$$\begin{aligned} \mathrm{tr}_d \otimes \varphi(X^m) &= \frac{1}{d} \sum_{i(1),\ldots,i(m)=1}^{d} \varphi\big[s_{i_1 i_2} \cdots s_{i_m i_1}\big] \\ &= \frac{1}{d} \sum_{\pi \in NC_2(m)} \sum_{i(1),\ldots,i(m)=1}^{d} \prod_{(p,q)\in\pi} \sigma\big(i_p, i_{p+1}; i_q, i_{q+1}\big). \end{aligned}$$

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

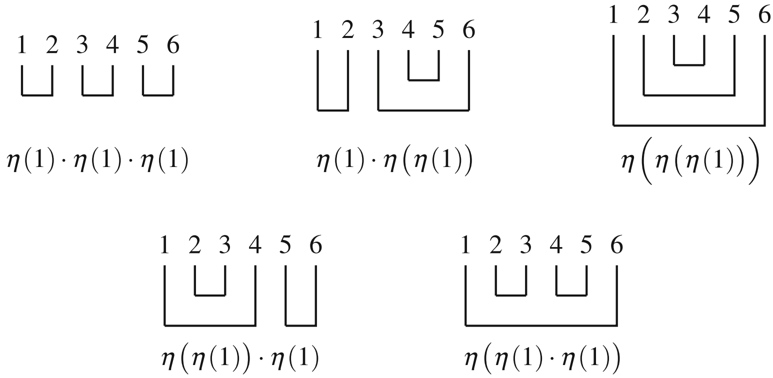
$$\begin{aligned}[\kappa_\pi]_{ij} &= \sum_{i_5,i_6=1}^{d} \sigma(i_5,i_6;i_6,j)\cdot\sum_{i_2,i_4=1}^{d}\sigma(i,i_2;i_4,i_5)\cdot[\eta(1)]_{i_2i_4}\\ &= \sum_{i_5,i_6=1}^{d} \sigma(i_5,i_6;i_6,j)\cdot\big[\eta\big(\eta(1)\big)\big]_{ii_5}.\end{aligned}$$

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

$$\|\Phi(B)\| \leq |z|^{-1}(1 + \|\eta\|\|B\|^2)$$

and

$$\|\Phi(B_1) - \Phi(B_2)\| \leq |z|^{-1}\|\eta\|(\|B_1\| + \|B_2\|)\|B_1 - B_2\|.$$

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 $\|\Phi(B_1) - \Phi(B_2)\| \leq c\|B_1 - B_2\|$ 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:

$$\begin{aligned} zf &= 1 + \frac{e\,(f + h) + 2\,(f^2 + h^2)}{3}, \\ ze &= 1 + \frac{e\,(e + 2\,(f + h))}{3}, \\ zh &= \frac{4\,f\,h + e\,(f + h)}{3}. \end{aligned} \tag{9.9}$$

This system of equations can be solved numerically for z close to the real axis; then

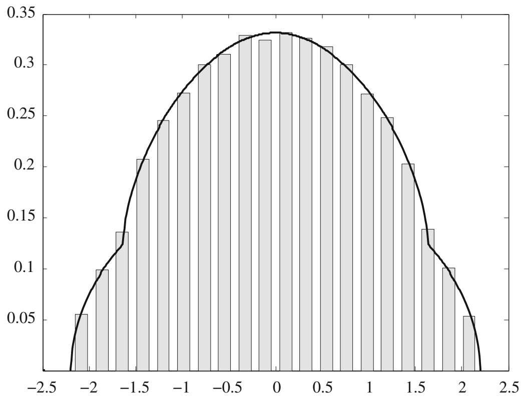
$$g(z) = \mathrm{tr}_3\big(G(z)\big) = (2f(z) + e(z))/3, \quad \frac{d\mu(t)}{dt} = -\frac{1}{\pi}\lim_{s\to 0} \mathrm{Im}\, g\,(t + is) \tag{9.10}$$

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),

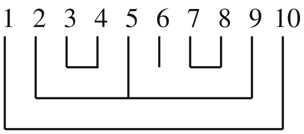
$$\begin{aligned} E(x_1y_1x_2y_2) = {} & E\big(x_1E(y_1)x_2\big)\cdot E(y_2) + E(x_1)\cdot E\big(y_1E(x_2)y_2\big) \\ & -E(x_1)E(y_1)E(x_2)E(y_2). \end{aligned} \tag{9.15}$$

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

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

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

$$\kappa_n^{\mathcal{B}}(b_0a_1,b_1a_2,\ldots,b_na_nb_{n+1}) = b_0\kappa_n^{\mathcal{B}}(a_1b_1,a_2b_2,\ldots,a_n)b_{n+1}$$

for all $a_1,\ldots,a_n\in\mathcal{A}$ and $b_0,\ldots,b_{n+1}\in\mathcal{B}$. 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 $T_n(x_1,\ldots,x_n)$. 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:

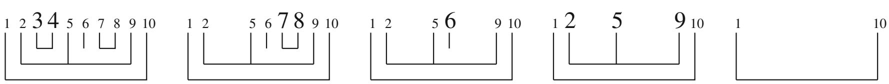
   (i) when $\pi = 1_n$, we have $T_\pi = T_n$;
   (ii) whenever $\pi\in NC(n)$ and $V=\{l+1,\ldots,l+k\}$ is an interval in π then

$$\begin{aligned}T_\pi(x_1,\ldots,x_n) &= T_{\pi'}(x_1,\ldots,x_lT_k(x_{l+1},\ldots,x_{l+k}),x_{l+k+1},\ldots,x_n)\\ &= T_{\pi'}(x_1,\ldots,x_l,T_k(x_{l+1},\ldots,x_{l+k})x_{l+k+1},\ldots,x_n),\end{aligned}$$

   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

$$T_\pi(x_1,\ldots,x_n) = T_{\pi'}(T_k(x_1,\ldots,x_k)x_{k+1},\ldots,x_n).$$

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

$$S_\pi = \sum_{\substack{\sigma \in NC(n) \\ \sigma \leq \pi}} T_\sigma \qquad (\pi \in NC(n)). \tag{9.17}$$

   Then the family $\{S_\pi\}_\pi$ is also multiplicative.

   ∘ The relation (9.17) between two multiplicative families is via Möbius inversions also equivalent to

$$T_\pi = \sum_{\substack{\sigma \in NC(n) \\ \sigma \leq \pi}} \mu(\sigma, \pi) S_\sigma \qquad (\pi \in NC(n)), \tag{9.18}$$

   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"

$$E_n : \mathcal{A}^{\otimes_{\mathcal{B}} n} \to \mathcal{B}, \qquad E_n(a_1, a_2, \ldots, a_n) = E(a_1 a_2 \cdots a_n).$$

   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

$$E_\pi(a_1, \ldots, a_{10}) = E\Big(a_1 \cdot E\big(a_2 \cdot E(a_3 a_4) \cdot a_5 \cdot E(a_6) \cdot E(a_7 a_8) \cdot a_9\big) \cdot a_{10}\Big).$$

   Then we define the multiplicative family $\{\kappa_\pi^{\mathcal{B}}\}_\pi$ by

$$\kappa_\pi^{\mathcal{B}} = \sum_{\substack{\sigma \in NC(n) \\ \sigma \leq \pi}} \mu(\sigma, \pi) E_\sigma \qquad (\pi \in NC(n)),$$

   which is equivalent to (9.16). In particular, this means that the $\kappa_n^{\mathcal{B}}$ are given by

$$\kappa_n^{\mathcal{B}}(a_1, \ldots, a_n) = \sum_{\pi \in NC(n)} \mu(\pi, 1_n) E_\pi(a_1, \ldots, a_n). \tag{9.19}$$

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

$$\begin{aligned} R_a(b) : &= \sum_{n \geq 0} \kappa_{n+1}^{\mathcal{B}}(ab, ab, \dots, ab, a) \\ &= \kappa_1^{\mathcal{B}}(a) + \kappa_2^{\mathcal{B}}(ab, a) + \kappa_3^{\mathcal{B}}(ab, ab, a) + \cdots . \end{aligned}$$

2) We say that $s \in \mathcal{A}$ is $\mathcal{B}$*-valued semi-circular* if $\kappa_n^{\mathcal{B}}(sb_1, sb_2, \dots, sb_{n-1}, s) = 0$ for all $n \neq 2$, and all $b_1, \dots, b_{n-1} \in \mathcal{B}$.

If $s \in \mathcal{A}$ is $\mathcal{B}$-valued semi-circular, then by the moment-cumulant formula, we have

$$E(s^n) = \sum_{\pi \in NC_2(n)} \kappa_\pi(s, \dots, s).$$

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*

$$bG(b) = 1 + R(G(b)) \cdot G(b) \qquad \textit{or} \qquad G(b) = (b - R(G(b)))^{-1}. \tag{9.20}$$

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 $x_1, \dots, x_r \in \mathcal{A}$, we actually mean all expressions of the form $\kappa_n^{\mathcal{B}}(x_{i_1}b_1, x_{i_2}b_2, \dots, x_{i_{n-1}}b_{n-1}, x_{i_n})$ ($n \in \mathbb{N}$, $1 \leq i(1), \dots, i(n) \leq r$, $b_1, \dots, b_{n-1} \in \mathcal{B}$).

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* $\mathbb{H}^+(\mathcal{B}) := \{b \in \mathcal{B} \mid \operatorname{Im}(b) > 0 \text{ and invertible}\}$. 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* $a^k_{ij} \in \mathcal{C}$ $(i, j = 1, \ldots, d;\ k = 1, \ldots, n)$ *and put, for each* $k = 1, \ldots, n$, $A_k = (a^k_{ij})^d_{i,j=1} \in M_d(\mathcal{C})$. *Then the operator-valued cumulants of the* A_k *are given in terms of the cumulants of their entries as follows:*

$$[\kappa^{\mathcal{B}}_n(A_1, A_2, \ldots, A_n)]_{ij} = \sum_{i_2,\ldots,i_n=1}^{d} \kappa_n\big(a^1_{i i_2}, a^2_{i_2 i_3}, \ldots, a^n_{i_n j}\big). \tag{9.23}$$

Proof: Let us begin by noting that

$$[E(A_1 A_2 \cdots A_n]_{ij} = \sum_{i_2,\ldots,i_n=1}^{d} \varphi\big(a^1_{i i_2} a^2_{i_2 i_3} \cdots a^n_{i_n j}\big).$$

Let $\pi \in NC(n)$ be a non-crossing partition; we claim that

$$[E_\pi(A_1, A_2, \ldots, A_n]_{ij} = \sum_{i_2,\ldots,i_n=1}^{d} \varphi_\pi\big(a^1_{i i_2}, a^2_{i_2 i_3}, \ldots, a^n_{i_n j}\big).$$

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

$$\begin{aligned}
[\kappa^{\mathcal{B}}_n(A_1, A_2, \ldots, A_n]_{ij} &= \sum_{\pi \in NC(n)} \mu(\pi, 1_n)[E_\pi(A_1, A_2, \ldots, A_n]_{ij}. \\
&= \sum_{i_2,\ldots,i_n=1}^{d} \sum_{\pi \in NC(n)} \mu(\pi, 1_n)\varphi_\pi\big(a^1_{i i_2}, a^2_{i_2 i_3}, \ldots, a^n_{i_n j}\big) \\
&= \sum_{i_2,\ldots,i_n=1}^{d} \kappa_\pi\big(a^1_{i i_2}, a^2_{i_2 i_3}, \ldots, a^n_{i_n j}\big).
\end{aligned}$$

□

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

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

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*

$$\kappa_n^{\mathcal{B}}(xd_1, xd_2, \ldots, xd_{n-1}, x) \in \mathcal{D} \qquad \forall n \geq 1, \quad \forall d_1, \ldots, d_{n-1} \in \mathcal{D}.$$

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 $d_1, \ldots, d_{n-1} \in \mathcal{D}$, we have

$$\kappa_n^{\mathcal{D}}(xd_1, xd_2, \ldots, xd_{n-1}, x) = \kappa_n^{\mathcal{B}}(xd_1, xd_2, \ldots, xd_{n-1}, x).$$

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.