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 $P\,(s_1, \ldots, s_{i_1}, c_1, \ldots, c_{i_2}, u_1, \ldots, u_{i_3}, d_1, \ldots, d_{i_4}) =: p_\infty$, 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" $P\big(s_1,\dots,s_{i_1},c_1,\dots,c_{i_2},u_1,\dots,u_{i_3},d_1^{(N)},\dots,d_{i_4}^{(N)}\big)$ $=: p_N^\square$, 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 $d_1^{(N)},\dots,d_{i_4}^{(N)}$ be exactly the same as the one of $D_1^{(N)},\dots,D_{i_4}^{(N)}$ 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 $M_2(\mathbb{C}) \otimes \mathcal{A}$,

$$\begin{pmatrix} 1 & bd^{-1} \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & -bd^{-1} \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & 0 \\ d^{-1}c & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 \\ -d^{-1}c & 1 \end{pmatrix},$$

the stated equivalence of (*i*) and (*ii*), as well as formula (10.3), follows from (10.4). □

What we now need, given our operator $p = P(x_1, \ldots, x_n)$, 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 $b_0, \ldots, b_n \in M_N(\mathbb{C})$ 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

$$\Lambda(z) = \begin{pmatrix} z & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 \end{pmatrix} \quad \text{for all } z \in \mathbb{C}. \tag{10.5}$$

As we will see in the following, the linearization in terms of the dimension $N \in \mathbb{N}$ and the matrices $b_0, \ldots, b_n \in M_N(\mathbb{C})$ usually depends only on the given polynomial $P \in \mathbb{C}\langle X_1, \ldots, X_n \rangle$ and not on the special choice of elements $x_1, \ldots, x_n \in \mathcal{A}$.

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 $P \in \mathbb{C}\langle X_1, \ldots, X_n \rangle$ be given. A matrix

$$\hat{P} := \begin{pmatrix} 0 & U \\ V & Q \end{pmatrix} \in M_N(\mathbb{C}) \otimes \mathbb{C}\langle X_1, \ldots, X_n \rangle,$$

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 $P \in \mathbb{C}\langle X_1, \ldots, X_n\rangle$ 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 $X_j \in \mathbb{C}\langle X_1, \ldots, X_n\rangle$ has a linearization

$$\hat{X}_j = \begin{pmatrix} 0 & X_j \\ 1 & -1 \end{pmatrix} \in M_2(\mathbb{C}) \otimes \mathbb{C}\langle X_1, \ldots, X_n\rangle.$$

(This statement looks simplistic taken for itself, but it will be useful when combined with the third part.)

(ii) A monomial of the form $P := X_{i_1} X_{i_2} \cdots X_{i_k} \in \mathbb{C}\langle X_1, \ldots, X_n\rangle$ for $k \geq 2$, $i_1, \ldots, i_k \in \{1, \ldots, n\}$ has a linearization

$$\hat{P} = \begin{pmatrix} & & & X_{i_1} \\ & & X_{i_2} & -1 \\ & \iddots & \iddots & \\ X_{i_k} & -1 & & \end{pmatrix} \in M_k(\mathbb{C}) \otimes \mathbb{C}\langle X_1, \ldots, X_n\rangle.$$

(iii) If the polynomials $P_1, \ldots, P_k \in \mathbb{C}\langle X_1, \ldots, X_n\rangle$ have linearizations

$$\hat{P}_j = \begin{pmatrix} 0 & U_j \\ V_j & Q_j \end{pmatrix} \in M_{N_j}(\mathbb{C}) \otimes \mathbb{C}\langle X_1, \ldots, X_n\rangle$$

for $j = 1, \ldots, n$, then their sum $P := P_1 + \cdots + P_k$ has the linearization

$$\hat{P} = \begin{pmatrix} 0 & U_1 & \ldots & U_k \\ V_1 & Q_1 & & \\ \vdots & & \ddots & \\ V_k & & & Q_k \end{pmatrix} \in M_N(\mathbb{C}) \otimes \mathbb{C}\langle X_1, \ldots, X_n\rangle$$

with $N := (N_1 + \cdots + N_k) - k + 1$.

(iv) If

$$\begin{pmatrix} 0 & U \\ V & Q \end{pmatrix} \in M_N(\mathbb{C}) \otimes \mathbb{C}\langle X_1, \ldots, X_n\rangle$$

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*) $\mathrm{Im}(\omega_j(b)) \geq \mathrm{Im}(b)$ *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 $\omega_2 \colon \{b \in \mathcal{B} \colon \mathrm{Im}(bx) > 0\} \to \mathbb{H}^+(\mathcal{B})$, such that*

(*i*) $\eta_y(\omega_2(b)) = \eta_{xy}(b)$, $\mathrm{Im}(bx) > 0$;
(*ii*) $\omega_2(b)$ *and* $b^{-1}\omega_2(b)$ *are analytic around zero;*
(*iii*) *for any $b \in \mathcal{B}$ so that $\mathrm{Im}(bx) > 0$, 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

$$\eta_{xy}(b) = \omega_2(b)\eta_x(\omega_1(b))\omega_2(b)^{-1}, \quad \mathrm{Im}(bx) > 0.$$

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

$$b = \begin{pmatrix} z & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{by} \quad \begin{pmatrix} z & 0 & 0 \\ 0 & i\varepsilon & 0 \\ 0 & 0 & i\varepsilon \end{pmatrix}$$

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.

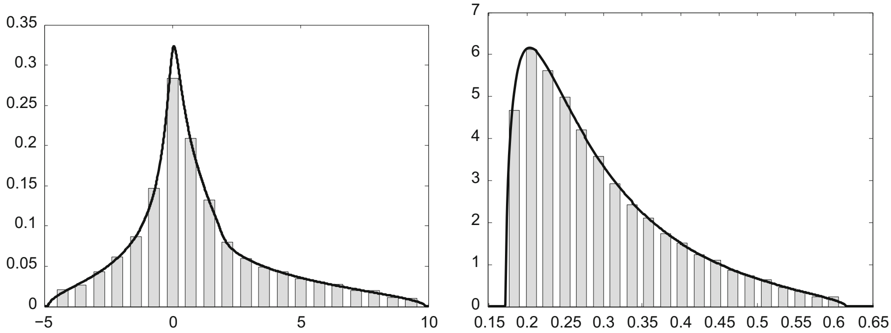

Fig. 10.1 Plots of the distribution of $p(x, y) = xy + yx + x^2$ (left) for free x, y, where x is semi-circular and y Marchenko-Pastur, and of the rational function $r(x_1, x_2)$ (right) for free semi-circular elements x_1 and x_2; in both cases the theoretical limit curve is compared with the histogram 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

$$\mathbb{H}^+(M_n(\mathbb{C})) := \{B \in M_n(\mathbb{C}) \mid \exists \varepsilon > 0 : \operatorname{Im}(B) \geq \varepsilon 1\},$$

where $\operatorname{Im}(B) := (B - B^*)/(2i)$.

(i) In the case $n = 2$, show that in fact

$$\mathbb{H}^+(M_2(\mathbb{C})) := \left\{ \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \middle| \operatorname{Im}(b_{11}) > 0, \operatorname{Im}(b_{11})\operatorname{Im}(b_{22}) > \frac{1}{4}|b_{12} - \overline{b_{21}}|^2 \right\}.$$

(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?