



Fluctuation Moments for Regular Functions of Wigner Matrices

Jana Reker¹

Received: 23 August 2023 / Accepted: 3 May 2024
© The Author(s) 2024

Abstract

We compute the deterministic approximation for mixed fluctuation moments of products of deterministic matrices and general Sobolev functions of Wigner matrices. Restricting to polynomials, our formulas reproduce recent results of Male et al. (Random Matrices Theory Appl. 11(2):2250015, 2022), showing that the underlying combinatorics of non-crossing partitions and annular non-crossing permutations continue to stay valid beyond the setting of second-order free probability theory. The formulas obtained further characterize the variance in the functional central limit theorem given in the recent companion paper (Reker in Preprint, [arXiv:2204.03419](https://arxiv.org/abs/2204.03419), 2023), and thus allow identifying the fluctuation around the thermal value in certain thermalization problems.

Keywords Wigner matrix · Global fluctuations · Fluctuation moments · Annular non-crossing permutations · Free probability

Mathematics Subject Classification 60B20 · 15B52 · 46L54

1 Introduction

In his seminal work [32], Wigner established that the empirical spectral measure of certain random matrix ensembles converges, as the dimension goes to infinity, to the semicircle distribution. Since then, many variations and extensions of this result have been considered, yielding a variety of asymptotic phenomena for a wide range of random matrix models. One particular example is the fact that the resolvent $G(z) = (W - z)^{-1}$ of a large Hermitian random matrix W tends to concentrate around a deterministic matrix $M = M(z)$ for spectral parameters $z \in \mathbb{C}$ even just slightly away from the real axis (see, e.g., [5] and references therein for a collection of recent results). It was recently shown (see [5, 6]) that a similar concentration holds for alternating

✉ Jana Reker
jana.reker@ist.ac.at

¹ IST Austria, Am Campus 1, 3400 Klosterneuburg, Austria

products of the form

$$F_{[1,k]} := f_1(W)A_1 \dots f_k(W)A_k, \tag{1.1}$$

where A_1, \dots, A_k are bounded deterministic matrices and f_1, \dots, f_k are regular test functions, allowing in particular for $f_j(W) = G(z_j)$. Apart from resolvents, two choices of f_j are of special interest in this setting. First, consider the case $f_j(x) = e^{it_j x}$ with $t_j \in \mathbb{R}$. Interpreting the Wigner matrix W as the Hamiltonian of a mean-field quantum system and the deterministic bounded matrix A as an observable, the quantity

$$A(t) := e^{itW} A e^{-itW}$$

describes the Heisenberg time evolution of A . For $k \geq 2$, (1.1) hence describes the time evolution of multiple observables, possibly to different times, in the same quantum system. Considering the regime where the differences between the individual times are large links the matrix product to thermalization problems in mathematical physics (see [6, Sect. 2.1] and [5, Rem. 2.8]). Second, products of the form (1.1) with $f_j(W)$ replaced by (polynomials of) the random matrix itself play a key role in free probability theory, as they characterize the joint non-commutative probability distribution of Wigner and deterministic matrices.

We remind the reader that a (tracial *first-order*) non-commutative probability space is a pair (\mathcal{A}, φ_1) consisting of a complex unital algebra \mathcal{A} and a tracial linear functional $\varphi_1 : \mathcal{A} \rightarrow \mathbb{C}$ with $\varphi_1(1_{\mathcal{A}}) = 1$, where $1_{\mathcal{A}}$ is the unit element of the algebra. One particular example is the space $(\mathcal{A}, \varphi_1) = (\mathcal{M}_{N \times N}(L^{\infty-}(\Omega, \mathbb{P})), \mathbb{E}\langle \cdot \rangle)$ of $N \times N$ random matrices, where (Ω, \mathbb{P}) is a classical probability space, $\mathcal{M}_{N \times N}(S)$ denotes the $N \times N$ -matrices with entries in S , the space

$$L^{\infty-}(\Omega, \mathbb{P}) := \bigcap_{1 \leq p < \infty} L^p(\Omega, \mathbb{P})$$

contains all random variables with all finite moments, and $\langle \cdot \rangle$ denotes the normalized trace. Note that this definition includes deterministic and Wigner matrices. In this context, the non-commutative probability distribution of $a \in \mathcal{A}$ is characterized in terms of its moments $(\varphi_1(a^k))_k$ with the joint distribution of multiple elements of \mathcal{A} being defined analogously. Recent work by Cipolloni, Erdős and Schröder [6] established that the structure of the limit of $\mathbb{E}\langle F_{[1,k]} \rangle$ as in (1.1) matches the formulas obtained in free probability, and reproduces known results for the alternating moments $\mathbb{E}\langle W_1 D_1 \dots W_k D_k \rangle$ of a finite family of independent Wigner matrices $(W_j)_j$ and a finite family of deterministic matrices $(D_j)_j$ (see, e.g., [22, Sect. 4.4]) in the case $f_j(x) = x$. More precisely, in the large N limit, the leading-order term $m_1[F_{[1,k]}]$ of $\mathbb{E}\langle F_{[1,k]} \rangle$ is of the form

$$m_1[F_{[1,k]}] := \sum_{\pi \in NCP([k])} \left(\prod_{B \in \pi} \left\langle \prod_{j \in B} A_j \right\rangle \right) \Phi_{\pi}^{(1)}(f_1, \dots, f_k), \tag{1.2}$$

where $NCP([k])$ denotes the non-crossing partitions of the cyclically ordered set $\{1, \dots, k\}$ and the functions $\Phi_{\pi}^{(1)}$ only depend on f_1, \dots, f_k and $\pi \in NCP([k])$.

Hence, the right-hand side of (1.2) is a sum of terms that factorize into a contribution of the deterministic matrices resp. the test functions appearing in the product (1.1) with the underlying combinatorics matching the results obtained for the case $f_j(x) = x$ in free probability theory. Note, however, that resolvents and functions with an N -dependent mesoscopic scaling are typically not accessible in free probability as many of the standard techniques rely on explicit moment computations for polynomials. The results in [6] thus show that the underlying combinatorics continue to apply in a more general context. In particular, evaluating (1.2) for $f_j(x) = e^{it_j x}$ with $t_j \in \mathbb{R}$ implies a closed formula for the thermal value of quantities such as $\langle e^{it_1 W} A_1 e^{-it_2 W} A_2 \rangle$, which allows for a direct analysis of the $t \gg 1$ regime relevant for thermalization (see, e.g., [6, Cor. 2.9]).

After considering the concentration of (1.1), the next natural step is to study the fluctuations around the deterministic value. It is well-known that the linear statistics $\text{Tr} f(W) = \sum_{j=1}^N f(\lambda_j)$ with a regular test function $f : \mathbb{R} \rightarrow \mathbb{R}$ have a variance of order one (first observed in [14]) and, in fact, satisfy a central limit theorem (CLT) with a Gaussian limit, as shown, e.g., in [15] for the Wigner case and in [13] for invariant ensembles. By now, the statistics $\text{Tr} f(W)$ are well-studied on both macroscopic and mesoscopic scales (see, e.g., [1–3, 11, 12, 16, 18, 19, 29–31] for the Wigner case and [7, 27] for further references on previous results for Wigner matrices and other models). However, while the fluctuations of $\text{Tr}[f(W)A]$ are known for general regular functions f (see [17] and [7]), traces of products of the form (1.1) for $k \geq 2$ have so far only been studied for f_j being polynomials in the context of second-order freeness (see, e.g., [22, Chap. 5] or [8, 20, 21]).

We remind the reader that a *second-order* non-commutative probability space is a triplet $(\mathcal{A}, \varphi_1, \varphi_2)$, where the functional $\varphi_2 : \mathcal{A} \times \mathcal{A} \rightarrow \mathbb{C}$ is bilinear, tracial in both arguments, symmetric under the interchanging of its arguments, and satisfies $\varphi_2(a, 1_{\mathcal{A}}) = \varphi_2(1_{\mathcal{A}}, a) = 0$ for all $a \in \mathcal{A}$. The second-order probability distribution of $a \in \mathcal{A}$ is characterized in terms of $(\varphi_2(a^k, a^\ell))_{k,\ell}$, called the *fluctuation moments*, with the joint moments of multiple elements again being defined analogously. As a canonical example, we remark that $\mathcal{M}_{N \times N}(L^\infty(\Omega, \mathbb{P}))$ may be endowed with the functional $\varphi_2(\cdot, \cdot) = \text{Cov}(\text{Tr}(\cdot), \text{Tr}(\cdot))$, to make it a second-order probability space. In contrast to the first-order structure, the fluctuation moments are sensitive to the symmetry class of the underlying Wigner matrix and explicitly involve the fourth cumulant of the entry distribution (see [21, Thm. 6], as well as [7, 25, 26]). In particular, we observe a breaking of universality compared to the first-order problem of computing $\mathbb{E}\langle \cdot \rangle$. The joint fluctuation moments of Wigner and deterministic matrices are explicitly known (cf. [22, Thm. 13 of Ch. 5] for the GUE case and [21, Thm. 6] for general Wigner matrices).

A functional CLT for traces of products of the form (1.1) has recently been established in the companion paper [27] and the limiting covariance is derived using a recursion. In the present paper, we supply the combinatorial argument necessary to obtain the solution to the recursion and compute the limiting covariance explicitly. In particular, we show that if W is a GUE matrix, the leading order

term $m_2[F_{[1,k]}|F_{[k+1,k+\ell]}]$ of the covariance of $\text{Tr}(F_{[1,k]})$ and $\text{Tr}(F_{[k+1,k+\ell]})$ (with $F_{[k+1,k+\ell]} = f_{k+1}A_{k+1} \dots f_{k+\ell}A_{k+\ell}$ of the same build as (1.1)) is given by

$$\begin{aligned}
 m_2[F_{[1,k]}|F_{[k+1,k+\ell]}] &= \sum_{\pi \in \overline{NCP}(k,\ell)} \left(\prod_{B \in \pi} \left\langle \prod_{j \in B} A_j \right\rangle \right) \Phi_{\pi}^{(2)}(f_1, \dots, f_{k+\ell}) \\
 &+ \sum_{\pi_1 \times \pi_2 \in \overline{NCP}(k) \times \overline{NCP}(\ell)} \left(\prod_{B_1 \in \pi_1, B_2 \in \pi_2} \left\langle \prod_{j \in B_1} A_j \right\rangle \left\langle \prod_{j \in B_2} A_j \right\rangle \right) \\
 &\times \Phi_{\pi_1 \times \pi_2}^{(2)}(f_1, \dots, f_{k+\ell}). \tag{1.3}
 \end{aligned}$$

Here, $\overline{NCP}(k, \ell)$ denotes the non-crossing permutations of the (k, ℓ) -annulus and the functions $\Phi_{\pi}^{(2)}$ resp. $\Phi_{\pi_1 \times \pi_2}^{(2)}$ only depend on $f_1, \dots, f_{k+\ell}$ and the underlying permutation resp. partition. Similar to (1.2), we thus obtain a sum of terms that factorize into a contribution of the deterministic matrices resp. the test functions appearing in the product (1.1) with the underlying combinatorics again matching the results obtained for the case $f_j(x) = x$ in free probability theory (see [24]). Moreover, we show that the overall structure of (1.3) continues to hold if W is chosen to be a Wigner matrix with $W_{ij} \stackrel{d}{=} N^{-1/2} \chi_{od}$ for $i < j$ and $W_{jj} \stackrel{d}{=} N^{-1/2} \chi_d$ for general entry distributions χ_{od} and χ_d . In the general case, however, the sum in the first line of the right-hand side of (1.3) splits into four summands $\Phi_{\pi}^{(GUE)}$, $\kappa_4 \Phi_{\pi}^{(\kappa)}$, $\sigma \Phi_{\pi}^{(\sigma)}$, and $\tilde{\omega}_2 \Phi_{\pi}^{(\omega)}$ which have different prefactors in terms of the deterministic matrices $A_1, \dots, A_{k+\ell}$. Here, $\Phi_{\pi}^{(GUE)}$ corresponds to the GUE case in (1.3) and the remaining contributions are associated with the parameters

$$\kappa_4 = \mathbb{E}|\chi_{od}|^4 - 2, \quad \sigma = \mathbb{E}\chi_{od}^2, \quad \tilde{\omega}_2 = \mathbb{E}\chi_d^2 - 1 - \sigma \tag{1.4}$$

of the Wigner matrix W . A similar decomposition is also observed for $\Phi_{\pi_1 \times \pi_2}^{(2)}$ in (1.3). In particular, we find that the closed expression obtained from solving the recursion in [27] has the same overall structure as the formulas in [21, Thm. 6]. This shows that the analogies [6] established in the first-order setting have a counterpart for the second-order structures. Our combinatorial approach further allows us to give the functions in (1.3) in a closed form, thus yielding a fully explicit formula for the limiting covariance in the GUE case.

We remark that the main results of the present paper, i.e., combinatorial formulas for $m_2[F_{[1,k]}|F_{[k+1,k+\ell]}]$ such as (1.3), are applied to obtain an explicit limiting covariance structure for the multi-point functional CLT [27, Thm. 2.7]. Here, replacing the recursive definition of the limiting variance by a closed formula allows for an easier application of the theorem, e.g., to thermalization problems in physics. Although the form of the limiting variance in the $k = 2$ case considered in [27, Cor. 2.12] is sufficiently simple and could be derived without additional combinatorial tools, characterizing the fluctuation for general $k \geq 2$ does require them to obtain an explicit formula for the variance. We further emphasize that the main results in the companion paper [27] are of analytic nature and that their main technical difficulty lies in including functions with a mesoscopic scaling of the form

$$f_j(x) = g_j(N^\gamma(x - E)) \tag{1.5}$$

where g_j is a regular N -independent function, $E \in \mathbb{R}$ lies in the bulk of the limiting spectrum of W , and $N^{-\gamma}$ is larger than the typical eigenvalue spacing around E . In contrast, we assume all test functions to be N -independent in the present paper and focus on the combinatorial structures arising in the multi-point functional CLT. While an extension to the functions in (1.5) is possible using the techniques from [27], restricting to the macroscopic regime allows for a cleaner presentation of the results. It further facilitates working with more general assumptions on the Wigner matrix W . Note that [27, As. 1.1] corresponds to setting $\sigma = \tilde{\omega}_2 = 0$ in (1.4), while Assumption 1.1 matches the setting of [7, 21] with general $\sigma \in [-1, 1]$ and $\tilde{\omega}_2 \geq -2$, thus generalizing the formulas from [27].

We conclude the section with a brief overview of the paper. After introducing some commonly used notations, the assumptions on the Wigner matrix W are given in Assumption 1.1. We then give a brief overview of the combinatorics needed to identify the deterministic approximation of $\langle T_1 \dots T_k \rangle$ where $T_j := G(z_j)A_j$, and the multi-resolvent local laws needed for the analysis of the fluctuations (Sect. 1.2) as well as the definitions from free probability that are used to characterize the limiting covariance of $\langle T_1 \dots T_k \rangle - \mathbb{E}\langle T_1 \dots T_k \rangle$ and $\langle T_{k+1} \dots T_{k+\ell} \rangle - \mathbb{E}\langle T_{k+1} \dots T_{k+\ell} \rangle$ (Sect. 1.3). To prepare for the statements of our main results, we give a CLT for the case that all functions f_j are resolvents (Theorem 2.3). The role of the limiting covariance in the theorem is played by a recursively defined set function $m_2[\cdot|\cdot]$ (Definition 2.1), which is the main object of interest in the present paper. We study the recursion in detail in Sect. 2.2 and obtain the main results, explicit combinatorial formulas for its solution (Theorems 2.4, 2.6 - 2.8). In Sect. 2.3, we extend the CLT to more general test functions (Theorem 2.9 and Corollary 2.10) to discuss the connection to free probability theory in detail. In particular, we apply the results to the case $f_j(x) = x$ and show that the limiting covariance in the functional CLT reduces to the formula for the joint fluctuation moments of GUE and deterministic matrices (Corollary 2.11) as given in [21]. Lastly, the proofs are given in Sects. 3 and 4. To keep the presentation concise, some routine calculations are deferred to the appendix. This includes the proof of a multi-resolvent global law with transposes (Appendix A.1) and the proof of the resolvent CLT in Theorem 2.3 (Appendix A.2) which are similar to previous results in [6] and [27], respectively.

1.1 General Notation

We start by introducing some notation used throughout the paper. For two positive quantities f, g , we write $f \lesssim g$ and $f \sim g$ whenever there exist (deterministic, N -independent) constants $c, C > 0$ such that $f \leq Cg$ and $cg \leq f \leq Cg$, respectively. We denote the Hermitian conjugate of a matrix A by A^* and the complex conjugate of a scalar $z \in \mathbb{C}$ by \bar{z} . Moreover, $\|\cdot\|$ denotes the operator norm, $\text{Tr}(\cdot)$ is the usual trace and $\langle \cdot \rangle = N^{-1} \text{Tr}(\cdot)$. We further denote the covariance of two complex random variables X_1, X_2 by $\text{Cov}(X_1, X_2)$ and follow the convention

$$\text{Cov}(Y_1, Y_2) = \mathbb{E}(Y_1 - \mathbb{E}Y_1)\overline{(Y_2 - \mathbb{E}Y_2)},$$

i.e., the covariance is linear in the first and anti-linear in the second entry. For $k, a, b \in \mathbb{N}$ with $a \leq b$, we set $[k] = \{1, \dots, k\}$ and adopt the interval notation $[a, b] = \{a, a + 1, \dots, b\}$. We further write $\langle a, b \rangle$ or $[a, b)$ to indicate that a or b are excluded from the interval, respectively. Ordered sets are denoted by (\dots) instead of $\{\dots\}$. Sets of the form $[k]$ and intervals are treated as ordered sets.

Given a matrix $A \in \mathbb{C}^{N \times N}$, the traceless part of A is denoted by $\mathring{A} := A - \langle A \rangle \text{Id}$ where Id denotes the identity matrix. Further, $\mathbf{a} := \text{diag}(A)$ denotes the diagonal matrix obtained from extracting only the diagonal entries of A and $A_1 \odot A_2$ denotes the entry-wise (or Hadamard) product of two matrices A_1 and A_2 . For a Hermitian matrix W and spectral parameters $z_1, \dots, z_k \in \mathbb{C} \setminus \mathbb{R}$, we write the corresponding resolvents as $G_j = G(z_j) := (W - z_j)^{-1}$ and index products of resolvents using the interval notation

$$G_{[a,b]} := G_a G_{a+1} \dots G_b$$

for $a, b \in \mathbb{N}$ with $a \leq b$. Recalling that angled brackets indicate that an edge point of the interval is excluded, we write $G_{\langle a,b \rangle}$ and $G_{[a,b)}$ to exclude G_a or G_b from the product, respectively. Moreover, G_\emptyset is interpreted as zero. Note that this notation differs slightly from [5, 6]. As we often consider alternating products of resolvents with deterministic matrices A_1, \dots, A_k , define $T_j := G_j A_j$ and apply the same interval notation as above to write

$$T_{[k]} := T_1 \dots T_k = G_1 A_1 \dots G_k A_k, \quad T_{[a,b]} := T_a T_{a+1} \dots T_b. \tag{1.6}$$

Again, angled brackets are used to exclude T_a or T_b from the product, respectively, and T_\emptyset is interpreted as zero. We call a product of the type (1.6) *resolvent chain* of length k .

Throughout the paper, we assume W to be an $N \times N$ real or complex Wigner matrix satisfying the following assumptions.

Assumption 1.1 The matrix elements of W are independent up to Hermitian symmetry $W_{ij} = \overline{W_{ji}}$ and we assume identical distribution in the sense that there is a centered real random variable χ_d and a centered real or complex random variable χ_{od} such that $W_{ij} \stackrel{d}{=} N^{-1/2} \chi_{od}$ for $i < j$ and $W_{jj} \stackrel{d}{=} N^{-1/2} \chi_d$, respectively. We further assume that $\mathbb{E}|\chi_{od}|^2 = 1$ as well as the existence of all moments of χ_d and χ_{od} , i.e., there exist constants $C_p > 0$ for any $p \in \mathbb{N}$ such that

$$\mathbb{E}|\chi_d|^p + \mathbb{E}|\chi_{od}|^p \leq C_p.$$

We remark that Assumption 1.1 matches the model considered in [7] and [21]. Compared to the conditions $\mathbb{E}\chi_{od}^2 = 0$ and $\mathbb{E}\chi_d^2 = 1$ in [27], we allow for arbitrary values of the parameters $\sigma = \mathbb{E}\chi_{od}^2 \in [-1, 1]$ and $\omega_2 = \mathbb{E}\chi_d^2 \geq 0$. This description includes real symmetric Wigner ensembles such as GOE ($\sigma = 1$) as well as matrices of the form $W = D + iS$ where D is a diagonal matrix and S is skew-symmetric ($\sigma = -1$). We further introduce the notation

$$\kappa_4 := \mathbb{E}|\chi_{od}|^4 - 2 \tag{1.7}$$

for the normalized fourth cumulant of the off-diagonal entries as well as

$$\tilde{\omega}_2 := \omega_2 - 1 - \sigma. \tag{1.8}$$

The eigenvalue density profile of W is described by the semicircle law

$$\rho_{sc}(x) := \frac{\sqrt{4 - x^2}}{2\pi} \mathbb{1}_{[-2,2]}(x) \tag{1.9}$$

which mainly enters our analysis in the form of its Stieltjes transform

$$m(z) := \int \frac{\rho_{sc}(x)}{x - z} dx, \quad z \in \mathbb{C} \setminus \mathbb{R}. \tag{1.10}$$

We remind the reader that $m(z)$ is the unique solution of the Dyson equation

$$-\frac{1}{m(z)} = m(z) + z, \quad \Im z \Im(m(z)) > 0 \tag{1.11}$$

and that its derivative satisfies

$$m'(z) = \frac{m(z)^2}{1 - m(z)^2}. \tag{1.12}$$

Given fixed $z_1, \dots, z_k \in \mathbb{C} \setminus \mathbb{R}$, set $m_j = m(z_j)$ and $m'_j = m'(z_j)$, respectively. We further introduce

$$q_{i,j} = \frac{m_i m_j}{1 - m_i m_j}, \tag{1.13}$$

and remark that $q_{j,j} = m'_j$ whenever $i = j$.

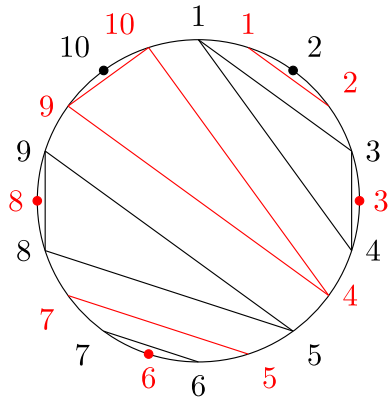
1.2 Preliminaries Part 1: First-Order Quantities

In this section, we briefly summarize the definitions and results from [5, 6] which are needed to characterize the deterministic approximation of $\langle T_{[1,k]} \rangle$.

Definition 1.2 (*Non-crossing partitions*) Let S be a finite (cyclically) ordered set of integers. We call a partition π of the set S *crossing* if there exist blocks $B \neq B'$ in π with $a, b \in B, c, d \in B'$, and $a < c < b < d$, otherwise we call it *non-crossing*. The set of non-crossing partitions is denoted by $NCP(S)$ and we abbreviate $NCP(k) := NCP([k])$. For each non-crossing partition $\pi = \{B_1, \dots, B_n\}$, set $|\pi| := n$ for the number of blocks in the partition.

Recall that non-crossing partitions have an alternative geometrical definition: Arrange the elements of S equidistantly in clockwise order on the circle and for each $\pi \in B$ consider the convex hull P_B of the points $s \in B$. Then π is non-crossing if and only if the polygons $\{P_B | B \in \pi\}$ are pair-wise disjoint. Because of this, we also call the elements of $NCP(k)$ *disk non-crossing* to distinguish them from their annulus analog defined below. We further recall the definition of the Kreweras complement (see Fig. 1 for an example).

Fig. 1 The non-crossing partition $\pi = \{\{1, 3, 4\}, \{2\}, \{5, 8, 9\}, \{6, 7\}, \{10\}\}$ (black) and its Kreweras complement $K(\pi) = \{\{1, 2\}, \{3\}, \{4, 9, 10\}, \{5, 7\}, \{6\}, \{8\}\}$ (red). Singleton sets are indicated by a black or red dot, respectively. (Color figure online)



Definition 1.3 (*Kreweras complement, disk case*) Let $S \subset \mathbb{N}$ be a finite set of integers equidistantly arranged in clockwise order on the circle and label the midpoints of the arcs between the points $s \in S$ also by the elements of S . We arrange the new labels such that the arc s follows the point s in clockwise order. Let $\pi \in NCP(S)$. Then the (*disk*) *Kreweras complement* of π , denoted by $K(\pi)$, is the element of $NCP(S)$ such that r, s belong to the same block of $K(\pi)$ if and only if the arcs labeled r, s are in the same connected component in the complement $D \setminus \cup_{B \in \pi} P_B$ of the polygons $\{P_B | B \in \pi\}$ in the labeled disk D .

Observe that $D \setminus \cup_{B \in \pi} P_B$ has $|S| - |\pi| + 1$ connected components, hence $|\pi| + |K(\pi)| = |S| + 1$. Further, $K^2 = K \circ K$ recovers π up to a rotation of D , i.e., $K^2(\pi)$ is the partition where for $S = \{s_1, \dots, s_k\}$ the elements in each block of π are shifted by $s_1 \mapsto s_2 \mapsto \dots \mapsto s_k \mapsto s_1$. In particular, taking the Kreweras complement is invertible as a map on $NCP(S)$.

Definition 1.4 (*Free cumulant function*) Fix $k \in \mathbb{N}$, denote the power set¹ of $[k]$ by $\mathcal{P}([k])$ and let $(f_n)_n$ be a family of functions such that f_n maps n -element sets in $\mathcal{P}([k])$ to \mathbb{C} . We define the (*first-order*) *free cumulants* $(f_{\circ, n})_n$ associated with the sequence $(f_n)_n$ through the relation

$$f_{|S|}[S] = \sum_{\pi \in NCP(S)} \prod_{B \in \pi} f_{\circ, |B|}[B] \tag{1.14}$$

for any $S \subseteq [k]$. For simplicity, we associate the sequences $(f_n)_n$ (resp. $(f_{\circ, n})_n$) with a single function f (resp. f_{\circ}) by setting $f[s_1, \dots, s_n] := f_n[\{s_1, \dots, s_n\}]$ (resp. $f_{\circ}[s_1, \dots, s_n] := f_{\circ, n}[\{s_1, \dots, s_n\}]$) and call f_{\circ} the (*first-order*) *free cumulant function* associated to f .

We emphasize that Definition 1.4 does not require the functions f_n to have any particular symmetries. However, in the free probability literature, f usually arises from

¹ Recall that $[k]$ is treated as an ordered set. Hence, the subsets in $\mathcal{P}([k])$ carry a natural ordering inherited from $[k]$.

a family of tracial functionals and is hence symmetric under the cyclic permutation of its entries (cf., e.g., [22, Ch. 2]). The implicit relation in (1.14) can be recursively turned into an explicit definition of f_\circ . Alternatively, we may also invert (1.14) explicitly using the Möbius function associated with the lattice of non-crossing partitions. Recall that $NCP(S)$ is a lattice with respect to the refinement order, i.e., the partial order in which $\pi \leq \nu$ if and only if for each $B \in \pi$ there exists $B' \in \nu$ with $B \subset B'$. Moreover, there are unique maximal and minimal elements given by $0_S := \{\{s\} | s \in S\}$ and $1_S := \{S\}$, respectively. The free cumulant function can then be written as

$$f_\circ[S] = \sum_{\pi \in NCP[S]} \mu(\pi, 1_S) \prod_{B \in \pi} f[B], \quad \mu(\pi, \nu) := \begin{cases} 1, & \pi = \nu, \\ -\sum_{\pi < \tau \leq \nu} \mu(\tau, \nu), & \pi < \nu, \end{cases} \tag{1.15}$$

using the Möbius function $\mu : \{(\pi, \nu) | \pi \leq \nu \in NCP(S)\} \rightarrow \mathbb{Z}$ that is recursively defined by (1.15). We remark that $\mu(\pi, 1_S)$ can be given in a closed form using the Catalan numbers (see, e.g., [6, Lem. 2.16]).

The following choice for the function f is of particular interest. Recall that m denotes the Stieltjes transform of the semicircle law (1.9).

Definition 1.5 (Divided differences) For finite sets $\{z_1, \dots, z_k\} \subset \mathbb{C} \setminus \mathbb{R}$ we recursively define

$$m[z_1, \dots, z_k] := \frac{m[z_2, \dots, z_k] - m[z_1, \dots, z_{k-1}]}{z_k - z_1}.$$

The definition readily extends to multi-sets $\{z_1, \dots, z_k\} \subset \mathbb{C} \setminus \mathbb{R}$ by setting

$$m[\underbrace{z_j, \dots, z_j}_{k \text{ times}}] := \frac{m^{(k-1)}(z_j)}{(k-1)!}$$

whenever an element z_j occurs with a multiplicity greater than one. Here, $m^{(k-1)}$ denotes the $(k-1)$ th derivative of the function m in (1.10). Note that $m[\cdot]$ is well-defined in the sense that $m[z_1, \dots, z_k]$ is independent of the ordering of the multi-set $\{z_1, \dots, z_k\}$. We abbreviate $m[1, \dots, k] := m[z_1, \dots, z_k]$.

We emphasize that $m[\cdot]$, and hence $m_\circ[\cdot]$, have full permutation symmetry, which is much more than what was assumed for f in Definition 1.4. The following example illustrates the combinatorial formulas (1.14) and (1.15) for $f = m[\cdot]$.

Example 1.6 (First-order free cumulants) In the case $k = 1$ we simply have $m[1] = m(z_1)$. For $k = 2$, the only non-crossing partitions are (12) and (1)(2) such that

$$m_\circ[1, 2] = m[1, 2] - m_1 m_2, \quad m_j := m[j] = m[z_j]$$

while for $k = 3$ we have

$$m_\circ[1, 2, 3] = m[1, 2, 3] - m_1 m[2, 3] - m_2 m[1, 3] - m_3 m[1, 2] + 2m_1 m_2 m_3.$$

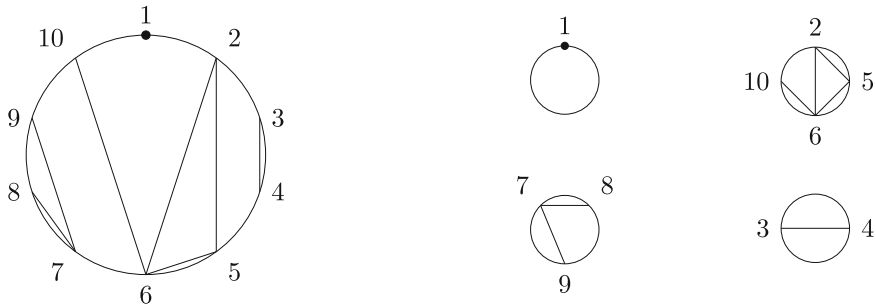


Fig. 2 An element of $NCG(10)$ and its connected subgraphs

The quantities m and m_\circ were studied in detail in [6], yielding a close connection to *non-crossing graphs*. We recall the definition and give an example in Fig. 2. These graphs are planar. For later convenience, we use a slightly more general notion of planar graphs throughout the paper than the standard literature by allowing for self-connections (loops) and multi-edges.

Definition 1.7 (*Disk non-crossing graphs*) Let $S \subset \mathbb{N}$ be a finite (cyclically) ordered set of integers equidistantly arranged in clockwise order on the circle. We call an undirected planar graph (S, E) on the vertex set S without loops or multi-edges (*disk crossing*) if there exist two edges $(a, b), (c, d) \in E$ with $a < c < b < d$, otherwise we call it (*disk non-crossing*).² The set of all (*disk non-crossing graphs*) with vertex set S is denoted by $NCG(S)$ and we denote the subset of connected graphs as $NCG_c(S)$. Whenever $S = [k]$, abbreviate $NCG(k) := NCG([k])$.

Emphasizing that Definition 1.7 lives on a disk is important, as we later introduce a non-crossing property on the annulus. Whenever both definitions are used together, we use the specifications *disk non-crossing* and *annular non-crossing* to distinguish the underlying geometry. By construction, every $\Gamma \in NCG(S)$ induces a non-crossing partition with blocks representing the vertices in the connected components of Γ . Further, any connected component of Γ is itself a (*disk non-crossing graph*).

Lemma 5.2 of [6] proves the representations

$$m[S] = \left(\prod_{s \in S} m_s \right) \sum_{\Gamma \in NCG(S)} \prod_{(i,j) \in E(\Gamma)} q_{i,j}, \tag{1.16}$$

$$m_\circ[S] = \left(\prod_{s \in S} m_s \right) \sum_{\Gamma \in NCG_c(S)} \prod_{(i,j) \in E(\Gamma)} q_{i,j} \tag{1.17}$$

in terms of the *weights* $q_{i,j}$ in (1.13). Here, $E(\Gamma)$ is the edge set of the graph Γ . Note that (1.16) and (1.17) are still well-defined if S is an ordered multi-set, i.e., if some elements are repeated. In this case, we consider $NCG(|S|)$ instead of $NCG(S)$ and

² The edges of a disk non-crossing graph Γ can be drawn in the interior of the disk without intersecting, i.e., Γ is a planar graph drawn inside a labeled disk.

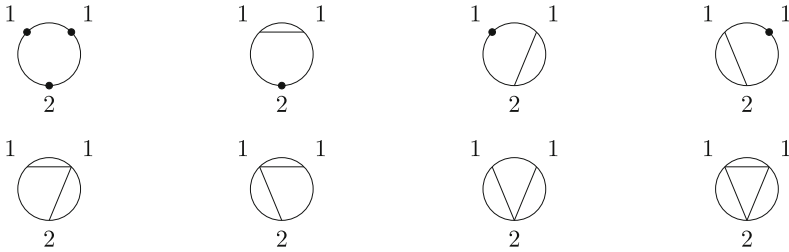


Fig. 3 The graphs contributing to the sum on the right-hand side of (1.16) (top and bottom row) and (1.17) (bottom row only) for the multi-set $S = (1, 1, 2)$. Note that the prefactor is given by $\prod_{s \in S} m_s = m_1^2 m_2$

use the one-to-one correspondence between the (possibly repeated) labels $\{s | s \in S\}$ and $\{1, \dots, |S|\}$ to obtain a uniquely defined right-hand side (see Fig. 3).

A key technical tool in the proof of our main results is the optimal multi-resolvent local law [5, Thm. 2.5]. As we only work on macroscopic scales, i.e., with N -independent spectral parameters, in the present paper, we state the result in the form of a global law and omit the dependence on $\eta_* = \min |\Im z_j|$. Recall the commonly used definition of stochastic domination.

Definition 1.8 (*Stochastic domination*) Let

$$X = \{X^{(N)}(u) | N \in \mathbb{N}, u \in U^{(N)}\} \text{ and } Y = \{Y^{(N)}(u) | N \in \mathbb{N}, u \in U^{(N)}\}$$

be two families of non-negative random variables that are indexed by N and possibly some other parameter u in some (possibly N -dependent) domain $U^{(N)}$. We say that X is *stochastically dominated* by Y , denoted by $X \prec Y$ or $X = \mathcal{O}_{\prec}(Y)$, if, for all $\varepsilon, C > 0$ we have

$$\sup_{u \in U^{(N)}} \mathbb{P}(X^{(N)}(u) > N^\varepsilon Y^{(N)}(u)) \leq N^{-C}$$

for large enough $N \geq N_0(\varepsilon, C)$.

Theorem 1.9 (Macroscopic version of [5, Thm. 2.5]) *Fix $k \in \mathbb{N}$ and pick spectral parameters $z_1, \dots, z_k \in \mathbb{C} \setminus \mathbb{R}$ with $|\Im z_j| \gtrsim 1$ and $\max_j |z_j| \leq N^{100}$ as well as deterministic matrices $A_1, \dots, A_k \in \mathbb{C}^{N \times N}$ with $\|A_i\| \lesssim 1$. Define³*

$$M_{[k]} := \sum_{\pi \in \mathcal{NCP}(k)} \left(\prod_{\substack{B \in K(\pi), \\ k \notin B}} \left\langle \prod_{j \in B} A_j \right\rangle \prod_{i \in B(k) \setminus \{k\}} A_i \right) \left(\prod_{B \in \pi} m_o[B] \right), \quad (1.18)$$

³ The noncommutative product $\prod_{j \in B} A_j$ for $B = (j_1, \dots, j_r)$ is defined as $A_{j_1} \dots A_{j_r}$ in the order inherited from $B \subseteq S$.

where $B(k)$ is the block in $K(\pi)$ that contains k . Recalling that $T_j = G_j A_j$, we have the averaged,⁴ local law

$$\langle T_{[1,k]} \rangle = \langle M_{[k]} A_k \rangle + \mathcal{O}_{\prec} \left(\frac{1}{N} \right), \tag{1.19}$$

and for $\mathbf{x}, \mathbf{y} \in \mathbb{C}^N$ with $\|\mathbf{x}\|, \|\mathbf{y}\| \lesssim 1$ we have the isotropic local law

$$\langle \mathbf{x}, T_{[1,k]} G_k \mathbf{y} \rangle = \langle \mathbf{x}, M_{[k]} \mathbf{y} \rangle + \mathcal{O}_{\prec} \left(\frac{1}{\sqrt{N}} \right). \tag{1.20}$$

As we frequently encounter $\langle M_{[k]} A_k \rangle$ in the following sections, we introduce the notation

$$m_1[T_1, \dots, T_k] = m_1[z_1, A_1, \dots, z_k, A_k] := \langle M_{[k]} A_k \rangle. \tag{1.21}$$

In particular,

$$m_1[T_1, \dots, T_k] = \sum_{\pi \in NCP(k)} \left(\prod_{B \in K(\pi)} \left\langle \prod_{j \in B} A_j \right\rangle \right) \left(\prod_{B \in \pi} m_{\circ}[B] \right) \tag{1.22}$$

and we have $m_1[G_1, \dots, G_k] = m[1, \dots, k]$ as a consequence of (1.14). Comparing (1.22) with the formulas in [22, Sect. 4.4], the set functions m and m_{\circ} may be interpreted as the moments resp. free cumulants characterizing the limiting non-commutative probability distribution of resolvents $(G_j)_j$ while $m_1[\cdot]$ describes the limiting joint distribution of resolvents $(G_j)_j$ and deterministic matrices $(A_j)_j$.

We further apply (1.19) or (1.20) for a product $T_{s_1} \dots T_{s_{k-1}} G_{s_k}$ that is indexed by a (cyclically) ordered set $S = (s_1, \dots, s_k)$ instead of an interval. In this case, the deterministic approximation is denoted as

$$M_S = M_{(s_1, \dots, s_k)}$$

with the same definition as in (1.18).

Remark The quantities $m[1, \dots, k]$, $m_{\circ}[1, \dots, k]$, $m_1[T_1, \dots, T_k]$, $(M_{[k]})_{ij}$, and $\|M_{[k]}\|$ are of order one for any $k \in \mathbb{N}$ and $i, j \in [N]$ in the macroscopic regime (cf. Lemma 2.4 and Appendix A of [5]). Theorem 1.9 asserts that the deterministic $M_{[k]}$ is the leading order approximation of $T_{[1,k]} G_k$. In particular, the error terms in (1.19) and (1.20) are smaller than the natural upper bound on their leading term by a factor of $1/N$ and $1/\sqrt{N}$, respectively.

We further need a generalization of the averaged local law (1.19) that includes transposes.

⁴ Recall that $\langle \cdot \rangle$ carries a N^{-1} normalization factor, but $\langle \cdot, \cdot \rangle$ does not. For $A := |y\rangle\langle x|$ with two unit vectors $\mathbf{x}, \mathbf{y} \in \mathbb{C}^N$, we hence have $\langle \mathbf{x} G(z) \mathbf{y} \rangle = \text{Tr}[G(z)A]$ and $\langle G(z)A \rangle = N^{-1} \langle \mathbf{x}, G(z) \mathbf{y} \rangle$.

Theorem 1.10 (Global law for resolvent chains with transposes) *Let $k \in \mathbb{N}$ and pick spectral parameters z_1, \dots, z_k with $|\Im(z_j)| \gtrsim 1$ and $\max_j |z_j| \leq N^{100}$ as well as deterministic matrices A_1, \dots, A_k with $\|A_j\| \lesssim 1$. Moreover, let $G_j^\#$ denote either the resolvent $G_j = G(z_j)$ or its transpose G_j^t and denote by $\#$ the binary vector that has a one in j th position if $G_j^\# = G_j^t$ and a zero otherwise. Then,*

$$\langle G_1^\# A_1 \dots G_k^\# A_k \rangle = \sum_{\pi \in NCP(k)} \left(\prod_{B \in K(\pi)} \left\langle \prod_{j \in B} A_j \right\rangle \right) \left(\prod_{B \in \pi} m_{\circ}^{\#, \sigma}[B] \right) + \mathcal{O}_{\prec} \left(\frac{1}{N} \right) \tag{1.23}$$

where $m_{\circ}^{\#, \sigma}[\cdot]$ denotes the free cumulants associated with the set function $m^{\#, \sigma}[\cdot]$ by Definition 1.4. Here, $m^{\#, \sigma}[\cdot]$ is defined to satisfy $m^{\#, \sigma}[\emptyset] = 0$ as well as the recursion

$$m^{\#, \sigma}[1, \dots, k] = m_1(1 + q_{1,k}^\#) \times \left(m^{\#, \sigma}[2, \dots, k] + \sum_{j=2}^k c_{1,j} m^{\#, \sigma}[1, \dots, j] m^{\#, \sigma}[j, \dots, k] \right) \tag{1.24}$$

with $q_{1,k}^\# = q_{1,k} = \frac{m_1 m_k}{1 - m_1 m_k}$ whenever $\#_1 = \#_k$, i.e., either both G_1 and G_k occur as transposes in the product $G_1^\# \dots G_k^\#$ or neither of them, and $q_{1,k}^\# = \frac{\sigma m_1 m_k}{1 - \sigma m_1 m_k}$ otherwise. Similarly, $c_{1,j} = 1$ whenever $\#_1 = \#_j$ and $c_{1,j} = \sigma$ otherwise. Recall that $\sigma = \mathbb{E} \chi_{od}^2$ where χ_{od} is the real or complex random variable that specifies the distribution of the off-diagonal entries of the Wigner matrix W .

The proof of Theorem 1.10 is, modulo careful bookkeeping of the transposes, similar to the proof of the averaged local law in [6, Thm. 3.4]. For the convenience of the reader, a brief sketch of the argument is included in Appendix A.1. We remark that the same result may be obtained on mesoscopic scales with optimal error bounds following the strategy of [5] (cf. [5, Rem. 2.2]) and that several examples in the cases $k \in \{2, 3\}$ are considered in Proposition 3.4 and Remark 3.5 of [4] as well as in Propositions 3.3 and 3.4 of [7].

Note that $\sigma = 1$ implies that the matrix W is real and its resolvent satisfies $G_j^t = G_j$. Hence, the statement of Theorem 1.10 reduces to that of an averaged global law for real symmetric Wigner matrices in this case. Due to the structural similarity between (1.22) and (1.23), we will slightly abuse notation and write the right-hand side of (1.23) as

$$m_1[G_1^\# A_1, \dots, G_k^\# A_k] := \sum_{\pi \in NCP(k)} \left(\prod_{B \in K(\pi)} \left\langle \prod_{j \in B} A_j \right\rangle \right) \left(\prod_{B \in \pi} m_{\circ}^{\#, \sigma}[B] \right). \tag{1.25}$$

Moreover, by Definition 1.4, we have $m_1[G_1^\#, \dots, G_k^\#] = m^{\#, \sigma}[1, \dots, k]$ and (1.24) reduces to the divided differences in Definition 1.5 whenever $\#$ is the zero vector.

1.3 Preliminaries Part 2: Second-Order Quantities

In this section, we give an overview of the definitions from free probability that are used in later sections (roughly following the notation of [22, Ch. 5]) as well as some related quantities appearing in the CLTs.

Recall that the key picture for describing the expectation of $\langle T_{[1,k]} \rangle$ is a disk with the labels $1, \dots, k$ organized in clockwise order along its boundary. In a very similar spirit, the key picture for describing the corresponding second-order object, i.e., the covariance of $\langle T_{[1,k]} \rangle$ and $\langle T_{[k+1,k+\ell]} \rangle$, consists of two concentric labeled circles. Let $k, \ell \in \mathbb{N}$ and arrange the numbers $1, \dots, k$ equidistantly in clockwise order on the outer circle and the numbers $k+1, \dots, k+\ell$ equidistantly in counter-clockwise order on the inner circle. We refer to the planar domain between these two circles together with the labeled points on its boundary as *the (k, ℓ) -annulus* (see Fig. 4). The labeled points will often serve as vertices of a graph. In this case, any edges connecting two points are drawn inside the annulus.

Definition 1.11 (*Annular non-crossing permutations*) Let $k, \ell \in \mathbb{N}$. We call a permutation of $[k+\ell]$ an *annular non-crossing permutation* if we can draw its cycles⁵ on the (k, ℓ) -annulus such that the following conditions (see [22, Def. 5 in Ch. 5]) are satisfied:

- (i) Non-crossing property: The cycles do not cross.
- (ii) Standardness: Each cycle encloses a region in the annulus that is homeomorphic to the disk with boundary oriented clockwise (in particular, the cycles follow the orientation of the numbering of the circles).
- (iii) Connectedness: At least one cycle connects both circles.

The set of annular non-crossing permutations is denoted by $\overrightarrow{NCP}(k, \ell)$. Any cycle that connects both circles is referred to as *connecting cycle*.

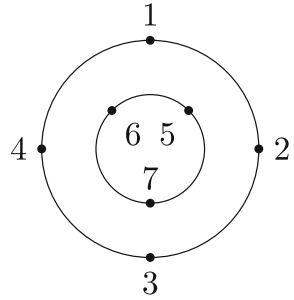
We remark that $\overrightarrow{NCP}(k, \ell)$ can be fully characterized by the avoidance of certain crossing patterns (cf. analogous geometric characterization of $NCP(k)$ below Definition 1.2) and an algebraic analog of the standardness condition. This equivalent definition is discussed, e.g., in [23, Sect. 3], but we will not use it here.

Definition 1.12 (*Annular non-crossing partitions*) Let $k, \ell \in \mathbb{N}$. We call the partitions induced by the cycles of $\overrightarrow{NCP}(k, \ell)$ *annular non-crossing partitions*. The set of annular non-crossing partitions is denoted by $NCP(k, \ell)$. A block that arises from a connecting cycle is referred to as *connecting block*.

While there is a one-to-one correspondence between the non-crossing partitions of the disk in Definition 1.2 and the permutations of $[k]$ avoiding the same crossing pattern, there is a crucial difference between non-crossing partitions and permutations on the (k, ℓ) -annulus. In particular, there is no bijective mapping between a permutation in $\overrightarrow{NCP}(k, \ell)$ and the partition of $[k+\ell]$ induced by its cycles, as, e.g., both permutations (123) and (132) correspond to the partition $\{\{1, 2, 3\}\}$, but give rise to

⁵ Recall that every permutation has a unique cycle decomposition. We represent a cycle $(abc\dots x)$ by an oriented graph with edge set $\{(a, b), (b, c), \dots, (x, a)\}$.

Fig. 4 The labels of the (4, 3)-annulus



different pictures due to the orientation induced by Definition 1.11(ii) (see Fig. 5). In general, a permutation always uniquely determines the underlying partition, but a partition can be obtained from more than one permutation. This happens if and only if there is exactly one connecting block (cf. [23, Sect. 4]).

Lastly, we consider partitions arising from permutations that respect the non-crossing property and standardness condition but do not have a connecting cycle. In this case, we may consider the permutation restricted to each circle separately, i.e., as an element of $NCP(k) \times NCP(\ell)$, and introduce an artificial connection by marking one block on each circle.

Definition 1.13 (*Marked non-crossing partition*) Consider $\pi \in NCP(k) \times NCP(\ell)$ that naturally splits into $\pi = \pi_1 \times \pi_2$ with $\pi_1 \in NCP(k), \pi_2 \in NCP(\ell)$. We pick one block of π_1 and one of π_2 , respectively, and mark them by underlining. The resulting object is referred to as a *marked non-crossing partition*.

Marking a block on each circle allows us to artificially introduce a connecting block by considering the union of the two marked blocks. As a consequence, any marked non-crossing partition can be associated with a unique element of $NCP(k, \ell)$. We further note that there are $|\pi_1| \cdot |\pi_2|$ possibilities to mark the blocks of $\pi = \pi_1 \times \pi_2 \in NCP(k) \times NCP(\ell)$. For example, $\{\{\underline{1}\}, \{2\}\} \times \{\{\underline{3}\}\}$ and $\{\{1\}, \{\underline{2}\}\} \times \{\{\underline{3}\}\}$ are considered different marked partitions although both arise from $\{\{1\}, \{2\}\} \times \{\{3\}\} \in NCP(2) \times NCP(1)$ (see Fig. 6).

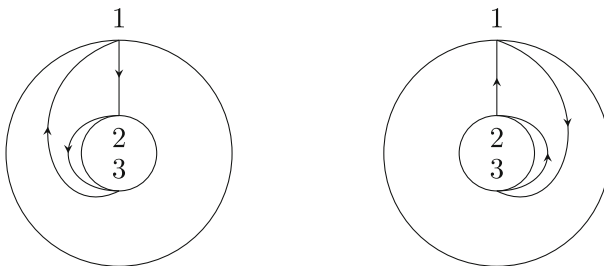


Fig. 5 The non-crossing permutations (123) and (132) on the (1, 2)-annulus. They are different as permutations, but their cycles induce the same non-crossing partition

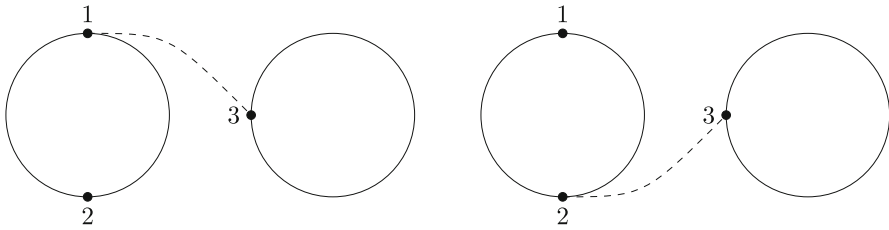


Fig. 6 A visualization of $\{\{1\}, \{2\}\} \times \{\{3\}\}$ and $\{\{1\}, \{2\}\} \times \{\{3\}\}$. Marking is indicated by a dashed line

We further recall from [22, Ch. 5] that there is a second-order analog of Definition 1.4.

Definition 1.14 (*Second-order free cumulant function*) Let $(f_{m,n})_{m,n}$ be a sequence of functions such that $f_{m,n}$ maps tuples (S_1, S_2) of two finite (cyclically) ordered sets of integers with $|S_1| = m$ and $|S_2| = n$ to \mathbb{C} . Assume further that the functions are symmetric under interchanging of the two arguments in the sense that $f_{m,n}[S_1|S_2] = f_{n,m}[S_2|S_1]$ and cyclic, i.e., $f_{m,n}[S_1\{s_1, \dots, s_n\}] = f_{m,n}[S_1\{s_2, \dots, s_n, s_1\}]$. Moreover, we assume that $f_{m,0}[S_1|\emptyset] = f_{0,n}[\emptyset|S_2] = 0$. Similar to Definition 1.4, we associate the sequence $(f_{m,n})_{m,n}$ with a single function f by setting

$$f[s_1, \dots, s_m | s_{m+1}, \dots, s_{m+n}] := f_{m,n}[\{s_1, \dots, s_m\} | \{s_{m+1}, \dots, s_{m+n}\}].$$

We implicitly define the *second-order free cumulant function* of f as the unique map $f_{\circ\circ}$ defined on pairs of finite (cyclically) ordered sets (U_1, U_2) that satisfies

$$f[S_1|S_2] = \sum_{\pi \in \text{NCP}(|S_1|, |S_2|)} \prod_{B \in \pi} f_{\circ}[B] + \sum_{\substack{\pi_1 \times \pi_2 \in \text{NCP}(|S_1|) \times \text{NCP}(|S_2|), \\ U_1 \in \pi_1, U_2 \in \pi_2 \text{ marked}}} f_{\circ\circ}[U_1|U_2] \prod_{\substack{B \in \pi_1 \setminus U_1 \\ \cup \pi_2 \setminus U_2}} f_{\circ}[B] \quad (1.26)$$

for any finite S_1, S_2 . Here, f_{\circ} is the first-order free cumulant function introduced in Definition 1.4.

Note that we use a set function f that is symmetric under the interchanging of its arguments instead of its skew-symmetric version $f[S_1|S_2] = \overline{f[S_2|S_1]}$ typically used in the free probability literature to mimic the covariance functional (cf. [22, Ch. 5]). This choice will simplify the computations by reducing the number of complex conjugates arising in the intermediate steps.

Similar to (1.14), the implicit relation (1.26) may be turned into an explicit definition of $f_{\circ\circ}$ by recursion. Note that the term $f_{\circ\circ}[[k] | [k + 1, k + \ell]]$ in formula (1.26) with $f[[k] | [k + 1, k + \ell]]$ on the left-hand side only occurs for the marked partition $\{\{1, \dots, k\}\} \times \{\{k + 1, \dots, k + \ell\}\}$ and hence always has coefficient one, so we can express it in terms of f, f_{\circ} , and the previously identified values of $f_{\circ\circ}$. This shows that $f_{\circ\circ}$ is well-defined. Although we will not rely on Möbius inversion to express

f_{∞} , we remark that it is possible to include both $\overrightarrow{NCP}(k, \ell)$ and the marked elements of $NCP(k) \times NCP(\ell)$ into one common definition, the *non-crossing partitioned permutations*, which can be endowed with a partial ordering and hence render it suitable for Möbius inversion. This would allow to rewrite the right-hand side of (1.26) to a structure similar to (1.14) and obtain a closed formula similar to (1.15). We refer to Sects. 4 and 5 of [8] for the full construction.

Similar to (1.14) above, the relation (1.26) is applied for one particular choice of f built up from the Stieltjes transform $m(z)$. We will later see that the set function $\tilde{m}[\cdot|\cdot]$ defined below arises as the deterministic approximation of the (appropriately scaled) covariance of $\langle G_{[1,k]} \rangle$ and $\langle G_{[k+1,k+\ell]} \rangle$ in a similar way that the divided differences $m[\cdot]$ arise for the expectation of $\langle G_{[1,k]} \rangle$. In particular, $\tilde{m}[\cdot|\cdot]$ satisfies the symmetry and cyclicity assumption in Definition 1.14. We give a recursive definition of $\tilde{m}[\cdot|\cdot]$ for now, however, closed formulas are later obtained in Sect. 2.2.

Definition 1.15 Let $S_1 = (z_1, \dots, z_{k'}) \subset \mathbb{C} \setminus \mathbb{R}$ and $S_2 = (z_{k'+1}, \dots, z_{k'+\ell'}) \subset \mathbb{C} \setminus \mathbb{R}$ be two finite ordered multi-sets. We define $\tilde{m}[\cdot|\cdot]$ to be the set function taking values in \mathbb{C} with the properties (i)–(iii) listed below. Similar to $m[\cdot]$ in Definition 1.5, we interpret $\tilde{m}[\cdot|\cdot]$ as a function of the indices of the spectral parameters.

(i) Symmetry: $\tilde{m}[\cdot|\cdot]$ is symmetric under the interchanging of its arguments, i.e., for any sets $B_1 \subseteq S_1, B_2 \subseteq S_2$ we have

$$\tilde{m}[(i, z_i \in B_1)|(j, z_j \in B_2)] = \tilde{m}[(j, z_j \in B_2)|(i, z_i \in B_1)].$$

(ii) Initial condition: For any sets $B_1 \subseteq S_1, B_2 \subseteq S_2$ we have

$$\tilde{m}[(i, z_i \in B_1)|\emptyset] = \tilde{m}[\emptyset|(j, j \in B_2)] = 0. \tag{1.27}$$

(iii) Recursion: Let $B_1 \subseteq S_1$ and $B_2 \subseteq S_2$ be ordered subsets with $|B_1| = k \leq k'$ and $|B_2| = \ell \leq \ell'$ elements, respectively. For simplicity, we index them by $[k]$ and $[k + 1, k + \ell]$. The function $\tilde{m}[\cdot|\cdot]$ satisfies the following linear recursion

$$\begin{aligned} &\tilde{m}[1, \dots, k|k + 1, \dots, k + \ell] \\ &= \frac{m_1}{1 - m_1 m_k} \left(\tilde{m}[2, \dots, k|k + 1, \dots, k + \ell] \right. \\ &\quad + \sum_{j=1}^{k-1} \tilde{m}[1, \dots, j|k + 1, \dots, k + \ell] m[j, \dots, k] \\ &\quad \left. + \sum_{j=2}^k m[1, \dots, j] \tilde{m}[j, \dots, k|k + 1, \dots, k + \ell] + s_{GUE} + s_k + s_\sigma + s_\omega \right) \end{aligned} \tag{1.28}$$

where the source terms in the last line are given by

$$s_{GUE} := \sum_{j=1}^{\ell} m[1, \dots, k, k + j, \dots, k + \ell, k + 1, \dots, k + j]$$

$$\begin{aligned}
 s_\kappa &:= \kappa_4 \sum_{r=1}^k \sum_{k+1 \leq s \leq t \leq k+\ell} m[1, \dots, r] \\
 &\quad \times m[r, \dots, k] m[s, \dots, t] m[t, \dots, k+\ell, k+1, \dots, s] \\
 s_\sigma &:= \sigma \sum_{j=1}^\ell m^{\#\sigma} [1, \dots, k, k+j, \dots, k+\ell, k+1, \dots, k+j] \\
 s_\omega &:= \tilde{\omega}_2 \sum_{j=1}^\ell m[1, \dots, k] m[k+j, \dots, k+\ell, k+1, \dots, k+j].
 \end{aligned}$$

Here, we wrote out the underlying multi-set in the definition of s_κ to indicate that it evaluates to $m[s, s]$ instead of m_s if $t = s$ and the vector $\# \in \{0, 1\}^{k+\ell+1}$ is given by $\#_1 = \dots = \#_k = 0$ and $\#_{k+1} = \dots = \#_{k+\ell+1} = 1$. Recall that $m[\cdot]$ denotes the divided differences as introduced in Definition 1.5 and that $m^{\#\sigma}[\cdot]$ was introduced in Theorem 1.10.

Note that the recursion for $\tilde{m}[\cdot|\cdot]$ is linear with different types of source terms in the last line of (1.28). Therefore, we may introduce the decomposition

$$\tilde{m}[\cdot|\cdot] = \tilde{m}_{GUE}[\cdot|\cdot] + \kappa_4 \tilde{m}_\kappa[\cdot|\cdot] + \sigma \tilde{m}_\sigma[\cdot|\cdot] + \tilde{\omega}_2 \tilde{m}_\omega[\cdot|\cdot] \tag{1.29}$$

where $\tilde{m}_{GUE}[\cdot|\cdot]$ satisfies (1.28) for $\kappa_4 = \sigma = \tilde{\omega}_2 = 0$ and $\kappa_4 \tilde{m}_\kappa[\cdot|\cdot]$, $\sigma \tilde{m}_\sigma[\cdot|\cdot]$, and $\tilde{\omega}_2 \tilde{m}_\omega[\cdot|\cdot]$ satisfy (1.28) with s_κ , s_σ , and s_ω as the only source term, respectively.

Note that the right-hand side of (1.28) only contains divided differences and $\tilde{m}[B_1|B_2]$ for $|B_1| + |B_2| < k + \ell$, so (1.28) indeed defines $\tilde{m}[\cdot|\cdot]$ recursively. The symmetry assumption in (i) then extends (1.28) to the second entry of $\tilde{m}[\cdot|\cdot]$. Moreover, all source terms in the last line of (1.28) are fully expressible as a function of $m_1, \dots, m_{k+\ell}$ by (1.16), making $\tilde{m}[\cdot|\cdot]$ eventually a function of $m_1, \dots, m_{k+\ell}$ as well.

As an example, setting $\sigma = \tilde{\omega}_2 = 0$ and applying the recursion once gives

$$\begin{aligned}
 \tilde{m}[1|2] &= \frac{m_1^2 m_2^2}{(1 - m_1^2)(1 - m_2^2)(1 - m_1 m_2)^2} + \kappa_4 \frac{m_1^3 m_2^3}{(1 - m_1^2)(1 - m_2^2)} \\
 &= \frac{m_1' m_2'}{(1 - m_1 m_2)^2} + \kappa_4 m_1 m_1' m_2 m_2' \tag{1.30}
 \end{aligned}$$

with $m_i' = m'(z_i)$. We remark that $\tilde{m}_{GUE}[1|2]$, seen as a function of (z_1, z_2) , is sometimes referred to as the second-order Cauchy transform of the GUE ensemble in the free probability literature (cf. [10]). The corresponding first-order object is $-m(z)$, which is obtained by applying the usual Cauchy transform to the semicircle law.⁶

We consider another special case in the following example.

⁶ The Cauchy transform of a probability measure μ is given by $c(z) = \int_{\mathbb{R}} \frac{\mu(dx)}{z-x}$ and hence only differs from the Stieltjes transform by a sign.

Example 1.16 Whenever $\kappa_4 = \sigma = \tilde{\omega}_2 = 0$ and one argument of $\tilde{m}_{GUE}[\cdot]$ is a singleton set, only the fourth line of (1.28) gives a non-zero contribution. Rewriting this term using (1.16) (cf. Lemma 1.17) yields the closed formula

$$\begin{aligned} \tilde{m}_{GUE}[1|2, \dots, \ell + 1] &= \frac{m_1}{1 - m_1^2} \sum_{j=2}^{\ell+1} m[1, \dots, j, \dots, \ell + 1, j] \\ &= \left(\prod_{s=1}^{\ell+1} m_s \right) \left(\sum_{\Gamma \in NCP(\ell+1)} \prod_{(a,b) \in E(\Gamma)} q_{a,b} \right) \\ &\quad \times \sum_{j=2}^{\ell+1} \frac{m'_1}{m_1} \frac{m'_j}{m_j} \left(1 + \sum_{i \neq j} q_{i,j} \right). \end{aligned} \tag{1.31}$$

For $\ell = 2$, we hence obtain

$$\tilde{m}_{GUE}[1|2, 3] = \frac{m'_1(m'_2 m_3(1 - m_1 m_3) + m_2 m'_3(1 - m_1 m_2))}{(1 - m_1 m_2)^2(1 - m_1 m_3)^2(1 - m_2 m_3)}.$$

Note that the right-hand side of (1.31) is fully expressed terms of non-crossing graphs on a labeled disk. This is because $NCP(1, \ell)$ and marked elements of $NCP(1) \times NCP(\ell)$ can be reduced to disk non-crossing partitions in this special case. In particular, the orientation of the circles is not relevant for this example.

The proof of (1.31) is immediate from the following combinatorial lemma which may be of independent interest. We give its proof in Appendix B.1.

Lemma 1.17 For $j \in \{1, \dots, k\}$, $k \geq 1$, we have

$$m[1, \dots, j, \dots, k, j] = m[1, \dots, k] \left(1 + \sum_{l \in [k] \setminus \{j\}} q_{j,l} \right) \frac{m'_j}{m_j} \tag{1.32}$$

with $q_{i,j}$ as in (1.13).

We also give some examples to illustrate the combinatorial formula (1.26) for the choice $f[\cdot] = \tilde{m}_{GUE}[\cdot]$.

Example 1.18 (Second-order free cumulants) Let $\kappa_4 = \sigma = \tilde{\omega}_2 = 0$. In the case $k = \ell = 1$, the only non-crossing annular permutation is (12) and there is also only one option for marking $\{\{1\}\} \times \{\{2\}\}$, namely $\{\{1\}\} \times \{\{2\}\}$. Rearranging (1.26), we thus get

$$m_{\circ\circ}[1|2] = \tilde{m}[1|2] - m[1, 2] + m_1 m_2.$$

Similarly, considering $k = 1$ and $\ell = 2$ yields

$$\begin{aligned} m_{\circ\circ}[1|2, 3] &= \tilde{m}[1|2, 3] - \tilde{m}[1|2]m_3 - \tilde{m}[1|3]m_2 - 2m[1, 2, 3] \\ &\quad + 2m_1 m[2, 3] + 2m_2 m[1, 3] + 2m_3 m[1, 2] - 4m_1 m_2 m_3. \end{aligned}$$

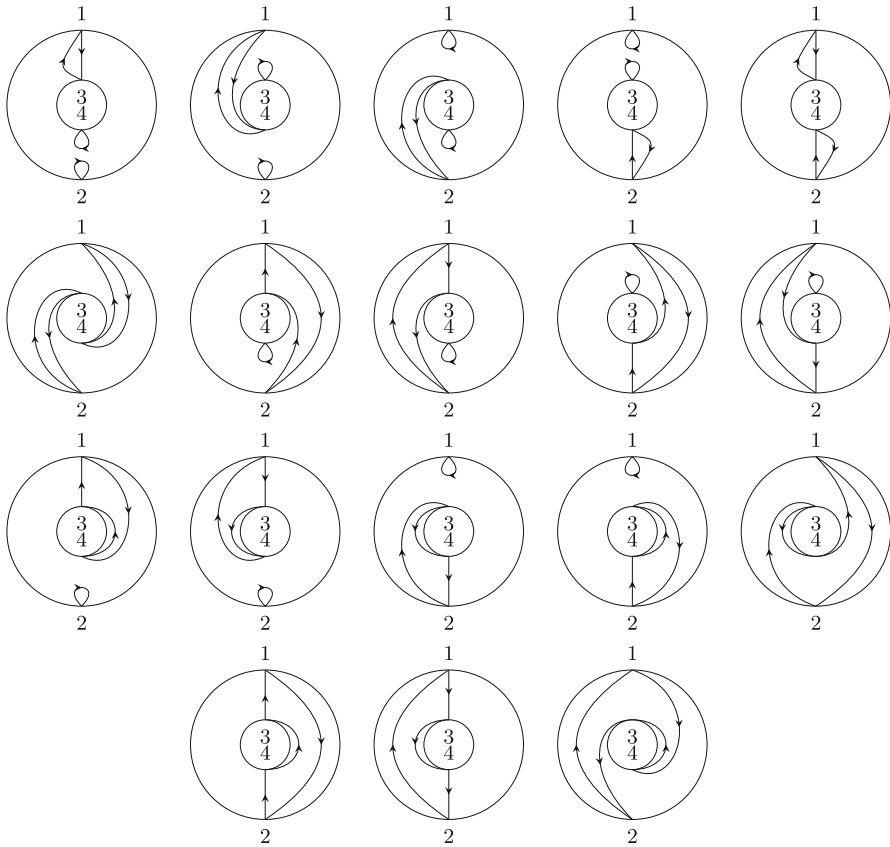


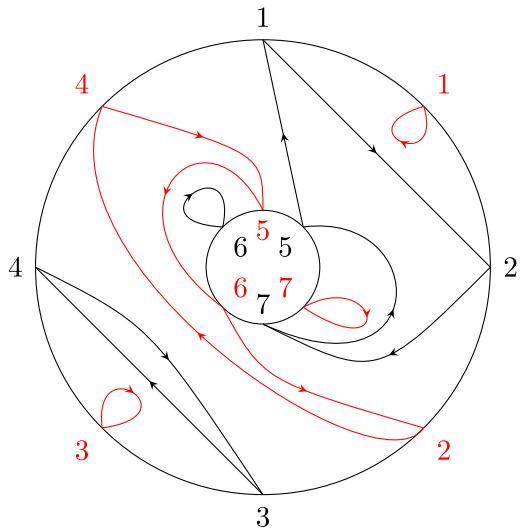
Fig. 7 The 18 elements of $\overrightarrow{NCP}(2, 2)$

In the case $k = \ell = 2$, there are 18 distinct non-crossing annular permutations (see Fig. 7) and 4 elements in $NCP(2) \times NCP(2)$. However, the second sum in (1.26) consists of 9 terms in total due to the marking of the blocks as, e.g., $\{\{1\}, \{2\}\} \times \{\{3, 4\}\}$ and $\{\{1\}, \{2\}\} \times \{\{3, 4\}\}$ correspond to $m_{\circ\circ}[1|3, 4]m_2$ and $m_{\circ\circ}[2|3, 4]m_1$, respectively, which do not need to coincide. In total, the formula defining $m_{\circ\circ}[1, 2|3, 4]$ has 27 terms on the right-hand side of (1.26).

We conclude this section by recalling the Kreweras complement for annular non-crossing permutations from [23] (see Fig. 8 for an example). Similarly to the disk case, taking the Kreweras complement is an invertible map on $\overrightarrow{NCP}(k, \ell)$.

Definition 1.19 (*Kreweras complement, annulus case*) Consider the (k, ℓ) -annulus and label the midpoints of the arcs between the points $1, \dots, k + \ell$ (black in Fig. 8) also by $1, \dots, k + \ell$ (red in Fig. 8). Respecting the orientation of the two circles, we arrange the new labels such that the arc s follows the point s . Let $\pi \in \overrightarrow{NCP}(k, \ell)$ be visualized on the (k, ℓ) -annulus as in Definition 1.11. The (annular) Kreweras complement $K(\pi) \in \overrightarrow{NCP}(k, \ell)$ is defined as the maximal annular non-crossing

Fig. 8 The annular non-crossing permutation $\pi = (1275)(34)(6)$ in black and its Kreweras complement $K(\pi) = (1)(2456)(3)(7)$ in red. (Color figure online)



permutation on $[k + \ell]$ that can be drawn using only the labels at the midpoints of the arcs and without intersecting the cycles of π . In particular, each cycle of $K(\pi)$ again encloses a region in the annulus that is homeomorphic to the disk with boundary oriented clockwise. In this context, we consider an annular non-crossing permutation maximal if none of its cycles can be extended (by merging cycles) without inducing a crossing.

Note that $|\pi| + |K(\pi)| = k + \ell$ for any $\pi \in \overrightarrow{NCP}(k, \ell)$ (see, e.g., [23, Sect. 6]). We remark that while defining the annular Kreweras complement on the level of partitions would also be possible, the resulting map does not have the same properties as in the disk case (see, e.g., [23, Sect. 1] for a discussion). Therefore, we will only consider the annular Kreweras complement for permutations. Note that one can further assign a unique Kreweras complement to any marked non-crossing partition π arising from some element $\pi_1 \times \pi_2 \in NCP(k) \times NCP(\ell)$ by applying Definition 1.3 circle-wise. In this case, we write $K(\pi) = K(\pi_1) \times K(\pi_2)$.

2 Main Results

The main focus of the present paper lies in determining the limiting covariance structure arising in the CLT for the centered statistics

$$X_\alpha := \langle T_{[1,k]} \rangle - \mathbb{E} \langle T_{[1,k]} \rangle = \langle G_1 A_1 \dots G_k A_k \rangle - \mathbb{E} \langle G_1 A_1 \dots G_k A_k \rangle, \quad (2.1)$$

$$Y_\alpha := \langle f_1(W) A_1 \dots f_k(W) A_k \rangle - \mathbb{E} \langle f_1(W) A_1 \dots f_k(W) A_k \rangle. \quad (2.2)$$

Here, $\alpha = ((z_1, A_1), \dots, (z_k, A_k))$ resp. $\alpha := ((f_1, A_1), \dots, (f_k, A_k))$ is a multi-index containing bounded deterministic matrices A_1, \dots, A_k and either the spectral

parameters $z_1, \dots, z_k \in \mathbb{C} \setminus \mathbb{R}$ with $|\Im z_j| \gtrsim 1$ appearing in the resolvents or the test functions $f_1, \dots, f_k \in H^{k+1}(\mathbb{R})$ with $\|f_j\| \lesssim 1$. Whenever we need to refer to the number k of resolvents (resp. test functions) in the product X_α (resp. Y_α), we carry the parameter k as a superscript and write $X_\alpha^{(k)}$ (resp. $Y_\alpha^{(k)}$). Recall that we set $T_j = G_j A_j = G(z_j) A_j$ as well as $T_{[i,j]} = T_i T_{i+1} \dots T_j$. Similarly, we introduce $F_j := f_j(W) A_j$ and use the interval notation

$$F_{[i,j]} := f_i(W) A_i \dots f_j(W) A_j$$

for $i < j$ as well as $F_\emptyset = 0$.

2.1 Resolvent Central Limit Theorem and Recursion

We start by identifying the joint distribution of multiple $X_{\alpha_i}^{(k_i)}$ with different k_i and α_i . To state the limiting covariance structure, we introduce a recursively defined set function $m_2[\cdot|\cdot]$, which we later identify as the deterministic approximation of the (appropriately scaled) covariance of $\langle T_{[1,k]} \rangle$ and $\langle T_{[k+1,k+\ell]} \rangle$ similar to $M_{[k]}$ and $m_1[\cdot]$ arising for the expectation of $T_{[1,k]} G_k$ (see Theorem 1.9 as well as (1.21) and (1.22)).⁷ Note that $\alpha = ((z_1, A_1), \dots, (z_k, A_k))$ contains the same information on the spectral parameters and deterministic matrices involved as the set of matrices $(T_j, j \in [k])$. We will, therefore, occasionally abuse notation and use (z_j, A_j) and $T_j = G_j A_j$ interchangeably. In particular, we write

$$m_2[\alpha|\beta] = m_2[T_1, \dots, T_k | T_{k+1}, \dots, T_{k+\ell}]$$

where the two multi-indices α and β index the spectral parameters and deterministic matrices in T_1, \dots, T_k and $T_{k+1}, \dots, T_{k+\ell}$, respectively. At this point, we only give a recursive definition for $m_2[\cdot|\cdot]$, however, explicit formulas are later obtained in Sect. 2.2. Note that the case $\sigma = \tilde{\omega}_2 = 0$ of Definition 2.1 was already given in [27].

Definition 2.1 Let $S_1 = (T_1, \dots, T_{k'})$ and $S_2 = (T_{k'+1}, \dots, T_{k'+\ell'})$ be two (ordered) finite sets of complex $N \times N$ -matrices of the form $T_j = G_j A_j$. We define $m_2[\cdot|\cdot]$ as the (deterministic) function of pairs of sets S_1, S_2 with values in \mathbb{C} and the following properties:

- (i) Symmetry: $m_2[\cdot|\cdot]$ is symmetric under the interchanging of its arguments, i.e., for any sets $B_1 \subseteq S_1, B_2 \subseteq S_2$ we have

$$m_2[(T_i, i \in B_1) | (T_j, j \in B_2)] = m_2[(T_j, j \in B_2) | (T_i, i \in B_1)].$$

- (ii) Initial condition: For any sets $B_1 \subseteq S_1, B_2 \subseteq S_2$ we have

$$m_2[(T_i, i \in B_1) | \emptyset] = m_2[\emptyset | (T_j, j \in B_2)] = 0. \tag{2.3}$$

⁷ Note the similarity between the notations $m_1[\cdot]$ and $m_2[\cdot|\cdot]$, which take one and two resolvent chains as arguments, respectively.

(iii) Recursion: Let $B_1 \subseteq S_1$ and $B_2 \subseteq S_2$ be ordered subsets with $|B_1| = k \leq k'$ and $|B_2| = \ell \leq \ell'$ elements, respectively. We index the matrices in B_1 by $[k]$ and the matrices in B_2 by $[k + 1, k + \ell]$. The function $m_2[\cdot|\cdot]$ satisfies the following linear recursion

$$\begin{aligned}
 & m_2[T_1, \dots, T_k|T_{k+1}, \dots, T_{k+\ell}] \\
 &= m_1 \left(m_2[T_2, \dots, T_{k-1}, G_k A_k A_1|T_{k+1}, \dots, T_{k+\ell}] \right. \\
 &\quad + q_{1,k} m_2[T_2, \dots, T_{k-1}, G_k A_1|T_{k+1}, \dots, T_{k+\ell}] \langle A_k \rangle \tag{2.4} \\
 &\quad + \sum_{j=1}^{k-1} m_2[T_1, \dots, T_{j-1}, G_j|T_{k+1}, \dots, T_{k+\ell}] \\
 &\quad + (m_1[T_j, \dots, T_k] + q_{1,k} m_1[T_j, \dots, T_{k-1}, G_k] \langle A_k \rangle) \\
 &\quad + \sum_{j=2}^k m_1[T_1, \dots, T_{j-1}, G_j] (m_2[T_j, \dots, T_k|T_{k+1}, \dots, T_{k+\ell}] \\
 &\quad \left. + q_{1,k} m_2[T_j, \dots, T_{k-1}, G_k|T_{k+1}, \dots, T_{k+\ell}] \langle A_k \rangle) + \mathfrak{s}_{GUE} + \mathfrak{s}_\kappa + \mathfrak{s}_\sigma + \mathfrak{s}_\omega \right)
 \end{aligned}$$

where the source terms \mathfrak{s}_{GUE} , \mathfrak{s}_κ , \mathfrak{s}_σ , and \mathfrak{s}_ω are given by

$$\begin{aligned}
 \mathfrak{s}_{GUE} &:= \sum_{j=1}^{\ell} (m_1[T_1, \dots, T_k, T_{k+j}, \dots, T_{k+j-1}, G_{k+j}] \\
 &\quad + q_{1,k} m_1[T_1, \dots, T_{k-1}, G_k, T_{k+j}, \dots, T_{k+j-1}, G_{k+j}] \langle A_k \rangle) \tag{2.5}
 \end{aligned}$$

$$\begin{aligned}
 \mathfrak{s}_\kappa &:= \kappa_4 \sum_{r=1}^k \sum_{s=k+1}^{k+\ell} \left(\sum_{t=k+1}^s \langle M_{[r]} \odot M_{(s, \dots, k+\ell, k+1, \dots, t)} \rangle \langle (M_{[r,k]} A_k) \odot M_{[t,s]} \rangle \right. \\
 &\quad \left. + \sum_{t=s}^{k+\ell} \langle M_{[r]} \odot M_{[s,t]} \rangle \langle (M_{[r,k]} A_k) \odot M_{(t, \dots, k+\ell, k+1, \dots, s)} \rangle \right) \\
 &\quad + \kappa_4 q_{1,k} \sum_{r=1}^k \sum_{s=k+1}^{k+\ell} \left(\sum_{t=k+1}^s \langle M_{[r]} \odot M_{(s, \dots, k+\ell, k+1, \dots, t)} \rangle \langle M_{[r,k]} \odot M_{[t,s]} \rangle \right. \\
 &\quad \left. + \sum_{t=s}^{k+\ell} \langle M_{[r]} \odot M_{[s,t]} \rangle \langle M_{[r,k]} \odot M_{(t, \dots, k+\ell, k+1, \dots, s)} \rangle \right) \langle A_k \rangle. \tag{2.6}
 \end{aligned}$$

$$\begin{aligned}
 \mathfrak{s}_\sigma &:= \sigma \sum_{j=1}^{\ell} m_1[T_1, \dots, T_k, G_{k+j}^t A_{k+j-1}^t, \dots, G_{k+\ell}^t A_{k+1}^t, \dots, G_{k+j-1}^t A_{k+j}^t, G_{k+j}] \\
 &\quad + q_{1,k} \sigma \sum_{j=1}^{\ell} m_1[T_1, \dots, T_{k-1}, G_k, G_{k+j}^t A_{k+j-1}^t, \dots, G_{k+\ell}^t A_{k+1}^t, \\
 &\quad \dots, G_{k+j-1}^t A_{k+j}^t, G_{k+j}] \langle A_k \rangle \tag{2.7}
 \end{aligned}$$

$$\begin{aligned} \mathfrak{s}_\omega &:= \widetilde{\omega}_2 \sum_{j=1}^{\ell} \langle (M_{[k]} A_k) \odot M_{(k+j, \dots, k+\ell, k+1, \dots, k+j)} \rangle \\ &\quad + q_{1,k} \widetilde{\omega}_2 \sum_{j=1}^{\ell} \langle M_{[k]} \odot M_{(k+j, \dots, k+\ell, k+1, \dots, k+j)} \rangle \langle A_k \rangle \end{aligned} \tag{2.8}$$

Recall that \odot denotes the Hadamard product, $q_{1,k}$ was defined in (1.13), and $M_{(\dots)}$ was defined in Theorem 1.9. Moreover, recall that $m_1[\cdot]$ was defined in (1.21) and the notation with transposes was introduced in (1.25).

Note that setting $A_1 = \dots = A_{k+\ell} = \text{Id}$ reduces (2.4) to (1.28), showing that

$$m_2[G_1, \dots, G_k | G_{k+1}, \dots, G_{k+\ell}] = m[1, \dots, k | k+1, \dots, k+\ell].$$

We use the linearity of the recursion and the different types of source terms to introduce the decomposition

$$m_2[\cdot|\cdot] = m_{GUE}[\cdot|\cdot] + \kappa_4 m_\kappa[\cdot|\cdot] + \sigma m_\sigma[\cdot|\cdot] + \widetilde{\omega}_2 m_\omega[\cdot|\cdot], \tag{2.9}$$

where $m_{GUE}[\cdot|\cdot]$ satisfies (2.4) for $\kappa_4 = \sigma = \widetilde{\omega}_2 = 0$, and $\kappa_4 m_\kappa[\cdot|\cdot]$, $\sigma m_\sigma[\cdot|\cdot]$ resp. $\widetilde{\omega}_2 m_\omega[\cdot|\cdot]$ satisfy (2.4) with \mathfrak{s}_κ , \mathfrak{s}_σ resp. \mathfrak{s}_ω as only source term. Note that $\mathfrak{s}_{GUE} + \mathfrak{s}_\kappa + \mathfrak{s}_\sigma + \mathfrak{s}_\omega$ in (2.4) is fully expressible as a function of $A_1, \dots, A_{k+\ell}$ and $m_1, \dots, m_{k+\ell}$ by (1.22), (1.17) and Lemma A.2, eventually making $m_2[\cdot|\cdot]$ a function of the same quantities.

Recall that we set $X_\alpha = \langle T_1 \dots T_k \rangle - \mathbb{E} \langle T_1 \dots T_k \rangle$ with $\alpha = ((z_1, A_1), \dots, (z_k, A_k))$. Before stating the CLT for X_α , we note the following definition.

Definition 2.2 Consider two functions of the Wigner matrix W in Assumption 1.1, which we denote as N -dependent random variables $X^{(N)}$ and $Y^{(N)}$. We say that $X^{(N)} = Y^{(N)} + \mathcal{O}(N^{-c})$ with $c > 0$ in the sense of moments if for any polynomial \mathcal{P} it holds that

$$\mathbb{E} \mathcal{P}(X^{(N)}) = \mathbb{E} \mathcal{P}(Y^{(N)}) + \mathcal{O}(N^{-c+\varepsilon}),$$

for any small $\varepsilon > 0$, where the implicit constant in $\mathcal{O}(\cdot)$ only depends on the polynomial \mathcal{P} and the constants in Assumption 1.1.

We now give a CLT for X_α in (2.1). As the main interest of the present paper is the deterministic approximation $m_2[\cdot|\cdot]$, we restrict the discussion of the CLT to the macroscopic regime ($|\Im z_i| \gtrsim 1$) for technical simplicity. Note that this implies in particular that $m_2[\cdot|\cdot]$ as well as its components $m_{GUE}[\cdot|\cdot]$, $m_\kappa[\cdot|\cdot]$, $m_\sigma[\cdot|\cdot]$, and $m_\omega[\cdot|\cdot]$ are of order one. The proof of Theorem 2.3 is analogous to that of [27, Thm. 3.6]. For the convenience of the reader, we include the necessary modifications for adapting the proof in [27] to the generalized model in Assumption 1.1 in Appendix A.2.

Theorem 2.3 (Macroscopic CLT for resolvents) Fix $p \in \mathbb{N}$, let $\alpha_1, \dots, \alpha_p$ be multi-indices, and let W be a Wigner matrix satisfying Assumption 1.1. For each

$j = 1, \dots, p$ pick a set of spectral parameters $z_1^{(j)}, \dots, z_{k_j}^{(j)}$ with $|\Re z_i^{(j)}| \gtrsim 1$ and $\max_j |z_j| \leq N^{100}$ as well as deterministic matrices $A_1^{(j)}, \dots, A_{k_j}^{(j)}$ with $\|A_i^{(j)}\| \lesssim 1$. Then,

$$N^p \mathbb{E} \left(\prod_{j=1}^p X_{\alpha_j} \right) = \sum_{Q \in \text{Pair}(\{p\})} \prod_{\{i,j\} \in Q} m_2[\alpha_i | \alpha_j] + \mathcal{O} \left(\frac{N^\varepsilon}{\sqrt{N}} \right) \tag{2.10}$$

for any $\varepsilon > 0$. Here, $\text{Pair}(S)$ denotes the pairings of a set S and $m_2[\cdot | \cdot]$ is a set function that satisfies Definition 2.1. Equation (2.10) establishes an asymptotic version of Wick’s rule and hence identifies the joint limiting distribution of the random variables $(X_{\alpha_j})_j$ as asymptotically complex Gaussian in the sense of moments in the limit $N \rightarrow \infty$.

By (2.9), Theorem 2.3 implies in particular that

$$\lim_{N \rightarrow \infty} N^2 \mathbb{E}(X_\alpha X_\beta) = m_{GUE}[\alpha | \beta] + \kappa_4 m_\kappa[\alpha | \beta] + \sigma m_\sigma[\alpha | \beta] + \tilde{\omega}_2 m_\omega[\alpha | \beta].$$

We remark that $m_2[\cdot | \cdot]$ is cyclic in the sense that

$$m_2[(T_j, j \in S_1) | T_1, \dots, T_k] = m_2[(T_j, j \in S_1) | T_2, \dots, T_k, T_1]$$

and that further

$$\begin{aligned} & m_2[(T_j, j \in S_1) | T_1, \dots, T_{k-1}, G_k] \\ &= \frac{m_2[(T_j, j \in S_1) | T_2, \dots, T_{k-1}, G_k A_1] - m_2[(T_j, j \in S_1) | T_1, \dots, T_{k-1}]}{z_k - z_1}. \end{aligned}$$

whenever $z_1 \neq z_k$, $A_k = \text{Id}$, and $\sigma = 0$. These identities can be obtained from the “meta argument” below [5, Lem. 4.1] (see also [27, Cor. 3.7]) using that the analogous formulas for the original resolvent chains are trivially true by resolvent identities. However, any additional information on $m_2[\cdot | \cdot]$ has to be obtained from the recursion (2.4) directly.

2.2 Solution of the Recursion

After identifying $m_2[\alpha | \beta]$ as the deterministic approximation of $\mathbb{E}[X_\alpha X_\beta]$, we consider Definition 2.1 in detail. In this section, we derive the main result, which is a solution to the (deterministic) recursion (2.4). This characterizes the overall structure of the function $m_2[\cdot | \cdot]$ and yields explicit combinatorial formulas to replace the recursive definition in applications. Making use of the linearity of the recursion and the decomposition (2.9), it is sufficient to consider the components $m_{GUE}[\cdot | \cdot]$, $m_\kappa[\cdot | \cdot]$, $m_\sigma[\cdot | \cdot]$, and $m_\omega[\cdot | \cdot]$ separately. We start by studying $m_{GUE}[\cdot | \cdot]$. The proof consists of two steps that are carried out in Sect. 3.

Theorem 2.4 Let $\alpha = ((z_1, A_1), \dots, (z_k, A_k))$ and $\beta = ((z_{k+1}, A_{k+1}), \dots, (z_{k+\ell}, A_{k+\ell}))$ for some $k, \ell \in \mathbb{N}$. Then,

$$\begin{aligned}
 m_{GUE}[\alpha|\beta] &= \sum_{\pi \in \underline{NCP}(k, \ell)} \left(\prod_{B \in K(\pi)} \left\langle \prod_{j \in B} A_j \right\rangle \right) \prod_{B \in \pi} m_o[B] \\
 &+ \sum_{\substack{\pi_1 \times \pi_2 \in \underline{NCP}(k) \times \underline{NCP}(\ell), \\ U_1 \in \pi_1, U_2 \in \pi_2 \text{ marked}}} \left(\prod_{\substack{B_1 \in K(\pi_1), \\ B_2 \in K(\pi_2)}} \left\langle \prod_{j \in B_1} A_j \right\rangle \left\langle \prod_{j \in B_2} A_j \right\rangle \right) m_{oo}[U_1|U_2] \\
 &\times \prod_{\substack{B_1 \in \pi_1 \setminus U_1, \\ B_2 \in \pi_2 \setminus U_2}} m_o[B_1]m_o[B_2] \tag{2.11}
 \end{aligned}$$

with m_o and m_{oo} being the first and second-order free cumulant functions as defined in (1.14) and (1.26), respectively.

Observe that the right-hand side of (2.11) reduces to the combinatorial expression in (1.26) if $A_1 = \dots = A_{k+\ell} = \text{Id}$.

Remark Note that the right-hand side of (2.11) is symmetric with respect to interchanging of $((z_1, A_1), \dots, (z_k, A_k))$ and $((z_{k+1}, A_{k+1}), \dots, (z_{k+\ell}, A_{k+\ell}))$, which is consistent with the symmetry of $m_2[\cdot|\cdot]$ in Definition 2.1(i). We can check this directly from (2.11) by observing that there is a one-to-one correspondence between non-crossing permutations of the (k, ℓ) -annulus and those of the (ℓ, k) -annulus. This follows from drawing the cycles of the permutation as curves on the respective annuli and observing that interchanging the inner and outer circle with a conformal map (e.g., by inversion to a concentric circle between the outer and inner circle) preserves the standardness and non-crossing property of the picture (cf. Definition 1.15). Moreover, this symmetry of $\tilde{m}_{GUE}[\cdot|\cdot]$ implies that $m_{oo}[\cdot|\cdot]$ is also invariant under interchanging $((z_1, A_1), \dots, (z_k, A_k))$ and $((z_{k+1}, A_{k+1}), \dots, (z_{k+\ell}, A_{k+\ell}))$ since $\tilde{m}_{GUE}[\cdot|\cdot]$ determines $m_{oo}[\cdot|\cdot]$ uniquely by Definition 1.14.

Example 2.5 (Asymptotics of covariances for GUE) We consider a special case of Theorem 2.3. Let $p = 2, k_1 = k_2 = 1$, and assume that W is a GUE matrix.⁸ By decomposing A_1 and A_2 into a tracial and a traceless part, the deterministic approximation for the covariance follows directly from [7, Thm. 4.1], giving

$$\begin{aligned}
 &N^2 \mathbb{E}(\langle T_1 \rangle - \mathbb{E}\langle T_1 \rangle)(\langle T_2 \rangle - \mathbb{E}\langle T_2 \rangle) \\
 &= \langle A_1 A_2 \rangle \frac{m_1^2 m_2^2}{(1 - m_1 m_2)} + \langle A_1 \rangle \langle A_2 \rangle \left(\frac{m'_1 m'_2}{(1 - m_1 m_2)^2} - \frac{m_1^2 m_2^2}{(1 - m_1 m_2)} \right) + \mathcal{O}\left(\frac{N^\varepsilon}{\sqrt{N}}\right) \\
 &= \langle A_1 A_2 \rangle m_o[1, 2] + \langle A_1 \rangle \langle A_2 \rangle m_{oo}[1|2] + \mathcal{O}\left(\frac{N^\varepsilon}{\sqrt{N}}\right),
 \end{aligned}$$

where the last equation follows from the formulas in Examples 1.6 and 1.18. Note that the error bound Ψ/\sqrt{L} in [7, Eq. (91)] evaluates to $\mathcal{O}(1/\sqrt{N})$ on macroscopic scales.

⁸ Note that we are only using that W satisfies Assumption 1.1 and $\kappa_4 = \sigma = \tilde{\omega}_2 = 0$.

We remark that the deterministic leading term matches the formula for $m_{GUE}[T_1|T_2]$ obtained from applying (2.4) to the initial condition $m_{GUE}[T_1|\emptyset] = 0$.

Next, we consider the recursion for $m_\kappa[\cdot|\cdot]$. We obtain a closed solution similar to Theorem 2.4, i.e., a sum of terms that factorizes into two parts depending only on the deterministic matrices $A_1, \dots, A_{k+\ell}$ and the spectral parameters $z_1, \dots, z_{k+\ell}$, respectively. The proof of Theorem 2.6 is given in Sect. 4.1.

Theorem 2.6 *Let $\alpha = ((z_1, A_1), \dots, (z_k, A_k))$ and $\beta = ((z_{k+1}, A_{k+1}), \dots, (z_{k+\ell}, A_{k+\ell}))$ for some $k, \ell \in \mathbb{N}$. Then there exist*

- (i) a family $(\psi_{\pi,B})_{B \in \pi}$ of functions $\psi_{\pi,B} : \mathbb{C}^{|B|} \rightarrow \mathbb{C}$ for every $\pi \in \overrightarrow{NCP}(k, \ell)$ and
- (ii) a family $(\psi_{\pi,U_1,U_2})_{U_1 \subset [k], U_2 \subset [k+1, k+\ell]}$ of functions $\psi_{\pi,U_1,U_2} : \mathbb{C}^{|U_1|} \times \mathbb{C}^{|U_2|} \rightarrow \mathbb{C}$ that are invariant under interchanging of the two arguments as well as functions $(\psi_{\pi_1,B_1})_{B_1 \in \pi_1 \setminus U_1}$ and $(\psi_{\pi_2,B_2})_{B_2 \in \pi_2 \setminus U_2}$ with $\psi_{\pi_i,B_i} : \mathbb{C}^{|B_i|} \rightarrow \mathbb{C}$ for every $\pi = \pi_1 \times \pi_2 \in NCP(k) \times NCP(\ell)$ with marked blocks $U_1 \in \pi_1$ and $U_2 \in \pi_2$

such that

$$\begin{aligned}
 m_\kappa[\alpha|\beta] &= \sum_{\pi \in \overrightarrow{NCP}(k, \ell)} \prod_{B \in K(\pi)} \left\langle \left(\prod_{j \in B \cap [k]} A_j \right) \odot \left(\prod_{j \in B \cap [k+1, k+\ell]} A_j \right) \right\rangle \\
 &\quad \times \prod_{B \in \pi} \psi_{\pi,B}(z_j | j \in B) \\
 &\quad + \sum_{\substack{\pi = \pi_1 \times \pi_2 \in NCP(k) \times NCP(\ell), \\ U_1 \in \pi_1, U_2 \in \pi_2 \text{ marked}}} \left(\prod_{\substack{B_1 \in K(\pi_1), \\ B_2 \in K(\pi_2)}} \left\langle \prod_{j \in B_1} A_j \right\rangle \left\langle \prod_{j \in B_2} A_j \right\rangle \right) \\
 &\quad \times \psi_{\pi,U_1,U_2}(z_j | j \in U_1 \cup U_2) \prod_{\substack{B_1 \in \pi_1 \setminus U_1, \\ B_2 \in \pi_2 \setminus U_2}} \psi_{\pi_1,B_1}(z_j | j \in B_1) \psi_{\pi_2,B_2}(z_j | j \in B_2),
 \end{aligned} \tag{2.12}$$

where \odot denotes the Hadamard product.

Theorem 2.6 is a purely structural result which shows that $m_\kappa[\cdot|\cdot]$ can be written in terms of non-crossing permutations and partitions similar to $m_{GUE}[\cdot|\cdot]$. Despite the obvious similarities between (2.11) and (2.12), considering the minimal example

$$\begin{aligned}
 m_2[T_1|T_2] &= m_{GUE}[T_1|T_2] + \kappa_4 m_\kappa[T_1|T_2] \\
 &= \langle A_1 A_2 \rangle \frac{m_1^2 m_2^2}{(1 - m_1 m_2)} + \langle A_1 \rangle \langle A_2 \rangle \left(\frac{m_1' m_2'}{(1 - m_1 m_2)^2} - \frac{m_1^2 m_2^2}{(1 - m_1 m_2)} \right) \\
 &\quad + \kappa_4 (\langle \mathbf{a}_1 \mathbf{a}_2 \rangle m_1^3 m_2^3 + \langle A_1 \rangle \langle A_2 \rangle (2m_1 m_1' m_2 m_2' - m_1^3 m_2^3))
 \end{aligned}$$

with $\sigma = \tilde{\omega}_2 = 0$ already shows that the functions ψ_i describing the dependence on the spectral parameters do not coincide with the free cumulant functions $m_\circ[\cdot]$ and $m_{\circ\circ}[\cdot|\cdot]$ in general. However, (2.4) implies that the functions ψ_i themselves satisfy a recursion. Theorem 2.6 hence reduces the computation of $m_\kappa[\cdot|\cdot]$ from iterating (2.4) to an inductive computation of the part that only depends on the spectral parameters.

We continue by deriving an explicit formula for $m_\sigma[\cdot|\cdot]$. As the source term s_σ in the corresponding recursion is, up to transposes, identical to s_{GUE} , the solution of the recursion is analogous to Theorem 2.4 but uses $m^{\#, \sigma}[\cdot]$ instead of the iterated divided differences $m[\cdot]$. We give the proof in Sect. 4.2.

Theorem 2.7 *Let $\alpha = ((z_1, A_1), \dots, (z_k, A_k))$, $\beta = ((z_{k+1}, A_{k+1}), \dots, (z_{k+\ell}, A_{k+\ell}))$ for some $k, \ell \in \mathbb{N}$ and abbreviate*

$$\tilde{m}_\sigma[1, \dots, k|k + 1, \dots, k + \ell] := m_\sigma[G_1, \dots, G_k|G_{k+1}, \dots, G_{k+\ell}]$$

in the special case $A_1 = \dots = A_{k+\ell}$. If $B \in \pi$ is a connecting cycle of $\pi \in \overline{NCP}(k, \ell)$ decomposed as $B = (i_1, \dots, i_r) \circ (j_1, \dots, j_s)$ with $i_1, \dots, i_r \subset [k]$ and $j_1, \dots, j_s \subset [k + 1, k + \ell]$, we introduce the notation

$$B_\sigma := (i_1, \dots, i_r) \circ (j_s, \dots, j_1).$$

Then,

$$\begin{aligned} m_\sigma[\alpha|\beta] &= \sum_{\pi \in \overline{NCP}(k, \ell)} \left(\prod_{B \in K(\pi)} \left\langle \prod_{j \in B \cap [k]} A_j \left(\prod_{j \in B \cap [k+1, k+j]} A_j \right)^t \right\rangle \right) \prod_{B \in \pi} m_\sigma^{\#, \sigma}[B_\sigma] \\ &+ \sum_{\substack{\pi_1 \times \pi_2 \in \overline{NCP}(k) \times \overline{NCP}(\ell), \\ U_1 \in \pi_1, U_2 \in \pi_2 \text{ marked}}} \left(\prod_{\substack{B_1 \in K(\pi_1), \\ B_2 \in K(\pi_2)}} \left\langle \prod_{j \in B_1} A_j \right\rangle \left\langle \prod_{j \in B_2} A_j \right\rangle \right) \\ &\times (m_\sigma)_{\circ\circ}[U_1|U_2] \prod_{\substack{B_1 \in \pi_1 \setminus U_1, \\ B_2 \in \pi_2 \setminus U_2}} m_\sigma^{\#, \sigma}[B_1] m_\sigma^{\#, \sigma}[B_2] \end{aligned} \tag{2.13}$$

where $\# = (0, \dots, 0, 1, \dots, 1)$ with the number of zeros and ones matching the number of labels on the inner and outer circle involved in B , respectively. Moreover, $m_\sigma^{\#, \sigma}[\cdot]$ denotes the free cumulant function associated with $m^{\#, \sigma}[\cdot]$ via (1.14) and $(m_\sigma)_{\circ\circ}[\cdot|\cdot]$ denotes the second-order free cumulant function associated to $\tilde{m}_\sigma[\cdot|\cdot]$ and $m^{\#, \sigma}[\cdot]$ via (1.26), respectively.

Note that the set function $m_\sigma[\cdot|\cdot]$ satisfies the same factorization property as $m_{GUE}[\cdot|\cdot]$ and $m_\kappa[\cdot|\cdot]$. For $k = \ell = 1$, it readily follows that

$$m_\sigma[T_1|T_2] = \langle A_1 A_2^t \rangle \frac{m_1^2 m_2^2}{1 - \sigma m_1 m_2} + \langle A_1 \rangle \langle A_2 \rangle \left(\frac{m_1' m_2'}{(1 - \sigma m_1 m_2)^2} - \frac{m_1^2 m_2^2}{1 - \sigma m_1 m_2} \right).$$

We further remark that (2.9) evaluates to $m_2[\cdot|\cdot] = m_{GUE}[\cdot|\cdot] + m_\sigma[\cdot|\cdot]$ in the case of GOE matrices ($\sigma = 1$ and $\kappa_4 = \tilde{\omega}_2 = 0$).

It remains to consider $m_\omega[\cdot|\cdot]$. The proof of Theorem 2.8 is given in Sect. 4.2.

Theorem 2.8 *Let $\alpha = ((z_1, A_1), \dots, (z_k, A_k))$ and $\beta = ((z_{k+1}, A_{k+1}), \dots, (z_{k+\ell}, A_{k+\ell}))$ for some $k, \ell \in \mathbb{N}$. Then there exist*

- (i) a family $(\Psi_{\pi,B})_{B \in \pi}$ of functions $\Psi_{B,\pi} : \mathbb{C}^{|B|} \rightarrow \mathbb{C}$ for every $\pi \in \overrightarrow{NCP}(k, \ell)$ and
- (ii) a family $(\Psi_{\pi,U_1,U_2})_{U_1 \subset [k], U_2 \subset [k+1, k+\ell]}$ of functions $\Psi_{\pi,U_1,U_2} : \mathbb{C}^{|U_1|} \times \mathbb{C}^{|U_2|} \rightarrow \mathbb{C}$ that are invariant under interchanging of the two arguments as well as functions $(\Psi_{\pi_1,B_1})_{B_1 \in \pi_1 \setminus U_1}$ and $(\Psi_{\pi_2,B_2})_{B_2 \in \pi_2 \setminus U_2}$ with $\Psi_{\pi_i,B_i} : \mathbb{C}^{|B_i|} \rightarrow \mathbb{C}$ for every $\pi = \pi_1 \times \pi_2 \in NCP(k) \times NCP(\ell)$ with marked blocks $U_1 \in \pi_1$ and $U_2 \in \pi_2$

such that

$$\begin{aligned}
 m_\omega[\alpha|\beta] &= \sum_{\pi \in \overrightarrow{NCP}(k, \ell)} \prod_{B \in K(\pi)} \left\langle \left(\prod_{j \in B \cap [k]} A_j \right) \odot \left(\prod_{j \in B \cap [k+1, k+\ell]} A_j \right) \right\rangle \\
 &\times \prod_{B \in \pi} \Psi_{B,\pi}(z_j | j \in B) \\
 &+ \sum_{\substack{\pi = \pi_1 \times \pi_2 \in NCP(k) \times NCP(\ell), \\ U_1 \in \pi_1, U_2 \in \pi_2 \text{ marked}}} \left(\prod_{\substack{B_1 \in K(\pi_1), \\ B_2 \in K(\pi_2)}} \left\langle \prod_{j \in B_1} A_j \right\rangle \left\langle \prod_{j \in B_2} A_j \right\rangle \right) \\
 &\times \Psi_{\pi,U_1,U_2}(z_j | j \in U_1 \cup U_2) \prod_{\substack{B_1 \in \pi_1 \setminus U_1, \\ B_2 \in \pi_2 \setminus U_2}} \Psi_{\pi_1,B_1}(z_j | j \in B_1) \Psi_{\pi_2,B_2}(z_j | j \in B_2).
 \end{aligned} \tag{2.14}$$

Similar to Theorem 2.6, Theorem 2.8 is a pure structural result. It shows that the last contribution $m_\omega[\cdot|\cdot]$ satisfies the same factorization property as $m_{GUE}[\cdot|\cdot]$, $m_\kappa[\cdot|\cdot]$, and $m_\sigma[\cdot|\cdot]$. In the case $k = \ell = 1$, we have the formula

$$m_\omega[T_1|T_2] = \langle \mathbf{a}_1 \mathbf{a}_2 \rangle m_1^2 m_2^2 + \langle A_1 \rangle \langle A_2 \rangle (m'_1 m'_2 - m_1^2 m_2^2).$$

It further follows from (2.4) that the functions Ψ_i themselves satisfy a recursion. Theorem 2.8 thus reduces the computation of $m_\omega[\cdot|\cdot]$ to the inductive computation of the part that only depends on the spectral parameters.

2.3 General Test Functions and Applications to Free Probability

We conclude the discussion by comparing the explicit formulas from Sect. 2.2 to the free probability results in [21]. To do so, we generalize the CLT for resolvents in Theorem 2.3 to a full multi-point functional CLT for (N -independent) test functions f_1, \dots, f_k , i.e., a CLT for the statistics Y_α in (2.2). Given Theorem 2.3 and the formulas from Sect. 2.2, the proof follows from a classical application of the Helffer–Sjöstrand formula (analogous to the proof of [27, Thm. 2.7]) and is hence omitted. Note that we restrict Theorem 2.9 to real-valued test functions only for simplicity. Extending the results in this section to complex-valued test functions only requires minor modifications to the argument.

Theorem 2.9 (Macroscopic multi-point functional CLT) *Let $k \in \mathbb{N}$ and pick deterministic matrices $A_1, \dots, A_k \in \mathbb{C}^{N \times N}$ with $\|A_j\| \lesssim 1$. Let further W be a Wigner matrix satisfying Assumption 1.1 and let $f_1, \dots, f_k \in H^{k+1}(\mathbb{R})$ be real-valued compactly supported test functions with $\|f_j\| \lesssim 1$. Then, for any $\varepsilon > 0$, the centered*

statistics (2.2) are approximately distributed (in the sense of moments) as

$$NY_\alpha^{(k,a)} = \xi(\alpha) + \mathcal{O}\left(\frac{N^\varepsilon}{\sqrt{N}}\right) \tag{2.15}$$

with a centered (N -dependent) Gaussian process $\xi(\alpha)$ satisfying

$$\begin{aligned} \mathbb{E}\xi(\alpha)\xi(\beta) = & \sum_{\pi \in \text{NCP}(k,\ell)} \left(\prod_{B \in K(\pi)} \left\langle \prod_{j \in B} A_j \right\rangle \right) \Phi_\pi^{(GUE)}(f_1, \dots, f_{k+\ell}) \\ & + \kappa_4 \sum_{\pi \in \text{NCP}(k,\ell)} \prod_{B \in K(\pi)} \left\langle \left(\prod_{j \in B \cap [k]} A_j \right) \odot \left(\prod_{j \in B \cap [k+1, k+\ell]} A_j \right) \right\rangle \\ & \times \Phi_\pi^{(\kappa)}(f_1, \dots, f_{k+\ell}) \\ & + \sigma \sum_{\pi \in \text{NCP}(k,\ell)} \left(\prod_{B \in K(\pi)} \left\langle \prod_{j \in B \cap [k]} A_j \left(\prod_{j \in B \cap [k+1, k+j]} A_j \right)^t \right\rangle \right) \\ & \times \Phi_\pi^{(\sigma)}(f_1, \dots, f_{k+\ell}) \\ & + \tilde{\omega}_2 \sum_{\pi \in \text{NCP}(k,\ell)} \prod_{B \in K(\pi)} \left\langle \left(\prod_{j \in B \cap [k]} A_j \right) \odot \left(\prod_{j \in B \cap [k+1, k+\ell]} A_j \right) \right\rangle \\ & \times \Phi_\pi^{(\omega)}(f_1, \dots, f_{k+\ell}) \\ & + \sum_{\substack{\pi_1 \times \pi_2 \in \text{NCP}(k) \times \text{NCP}(\ell), \\ U_1 \in \pi_1, U_2 \in \pi_2 \text{ marked}}} \left(\prod_{\substack{B_1 \in K(\pi_1), \\ B_2 \in K(\pi_2)}} \left\langle \prod_{j \in B_1} A_j \right\rangle \left\langle \prod_{j \in B_2} A_j \right\rangle \right) \\ & \times \Phi_{\pi_1 \times \pi_2, U_1 \times U_2}(f_1, \dots, f_{k+\ell}). \end{aligned} \tag{2.16}$$

Here, β denotes another multi-index of length ℓ containing the deterministic matrices $A_{k+1}, \dots, A_{k+\ell}$ satisfying $\|A_j\| \lesssim 1$ and the test functions $f_{k+1}, \dots, f_{k+\ell} \in H^{\ell+1}(\mathbb{R})$. The functions $\Phi_\pi^{(\cdot)}$ and $\Phi_{\pi_1 \times \pi_2, U_1 \times U_2}$ in (2.16) can be computed recursively and only depend on the underlying permutation resp. marked partition, the functions $f_1, \dots, f_{k+\ell}$ and the model parameters κ_4, σ , and $\tilde{\omega}_2$.

For the later applications, we note the following formulas for the case $\kappa_4 = \sigma = \tilde{\omega}_2 = 0$. Corollary 2.10 is proven in [27] by explicitly evaluating the integrals that are obtained from the Helffer–Sjörstrand formula in the proof of the functional CLT.

Corollary 2.10 (Cor. 2.9 in [27]) *Consider Theorem 2.9 for a GUE matrix⁹ W . In this case, we have*

$$\Phi_\pi^{(GUE)}(f_1, \dots, f_{k+\ell}) := \prod_{B \in \pi} \text{sc}_\circ[B], \tag{2.17}$$

⁹ We only use that W satisfies Assumption 1.1 and $\kappa_4 = \sigma = \tilde{\omega}_2 = 0$.

where $sc_o[\cdot]$ denotes the free cumulant function associated with

$$sc[i_1, \dots, i_n] := \int_{-2}^2 \left[\prod_{j=1}^n f_{i_j}(x) \right] \rho_{sc}(x) dx, \tag{2.18}$$

with ρ_{sc} as in (1.9), and

$$\Phi_{\pi_1 \times \pi_2, U_1 \times U_2}(f_1, \dots, f_{k+\ell}) := sc_{oo}[U_1|U_2] \prod_{\substack{B_1 \in \pi_1 \setminus U_1, \\ B_2 \in \pi_2 \setminus U_2}} sc_o[B_1]sc_o[B_2], \tag{2.19}$$

where $sc_{oo}[\cdot|\cdot]$ denotes the second-order free cumulants associated with $sc[\cdot]$ in (2.18) and

$$sc[i_1, \dots, i_n | i_{n+1}, \dots, i_{n+m}] := \frac{1}{2} \int_{-2}^2 \int_{-2}^2 \left(\prod_{j=1}^n f_{i_j}(x) \right)' \left(\prod_{j=1}^m f_{i_{n+j}}(y) \right)' u(x, y) dx dy \tag{2.20}$$

by Definition 1.14. The kernel $u : [-2, 2] \times [-2, 2] \rightarrow \mathbb{R}$ is given by

$$u(x, y) := \frac{1}{4\pi^2} \ln \left[\frac{(\sqrt{4-x^2} + \sqrt{4-y^2})^2(xy + 4 - \sqrt{4-x^2}\sqrt{4-y^2})}{(\sqrt{4-x^2} - \sqrt{4-y^2})^2(xy + 4 + \sqrt{4-x^2}\sqrt{4-y^2})} \right]. \tag{2.21}$$

We remark that the formula (2.21) also appears in [9] and [28] (see also [27, Cor. 3.8]).

Whenever $f_j(x) = x$ for all $j = 1, \dots, k + \ell$ or, more generally, f_j is an (N -independent) polynomial,¹⁰ the $N \rightarrow \infty$ limit of (2.16) describes the second-order limiting distribution of GUE and deterministic matrices in free probability. It is readily checked that Theorem 2.9 indeed coincides with the free probability literature in this case. The computations to obtain Corollary 2.11 are included in Appendix B.2.

Corollary 2.11 *Under the assumptions of Corollary 2.10 let $f_1(x) = \dots = f_{k+\ell}(x) = x$, i.e., $Y_\alpha^{(k,a)} = \langle WA_1 \dots WA_k \rangle - \mathbb{E}\langle \dots \rangle$ and $Y_\beta^{(\ell,b)} = \langle WA_{k+1} \dots WA_{k+\ell} \rangle - \mathbb{E}\langle \dots \rangle$. Then,*

$$\lim_{N \rightarrow \infty} N^2 \mathbb{E} \left(Y_\alpha^{(k,a)} Y_\beta^{(\ell,b)} \right) = \sum_{\pi \in \overrightarrow{NC\hat{P}}_2(k,\ell)} \left(\prod_{B \in K(\pi)} \left\langle \prod_{j \in B} A_j \right\rangle \right) \tag{2.22}$$

where $\overrightarrow{NC\hat{P}}_2(k, \ell)$ denotes the pairings in $\overrightarrow{NC\hat{P}}(k, \ell)$.

¹⁰ We implicitly assume f_j to be compactly supported by setting $\tilde{f}_j(x) = f_j(x)\chi(x)$, where f_j is a polynomial supported on all of \mathbb{R} and χ is a smooth cutoff function that is equal to one on $[-5/2, 5/2]$ and equal to zero on $[-3, 3]^c$. Since $f_j(W) = \tilde{f}_j(W)$ with high probability by eigenvalue rigidity, we may use f_j and \tilde{f}_j interchangeably here.

We remark that the limit in (2.22) reproduces the well-known result of second-order freeness of GUE and deterministic matrices from [24]. Moreover, computations similar to Corollaries 2.10 and 2.11 for the GOE ($\sigma = 1$ and $\kappa_4 = \tilde{\omega}_2 = 0$) yield

$$\begin{aligned} & \lim_{N \rightarrow \infty} N^2 \mathbb{E} \left(Y_\alpha^{(k,a)} Y_\beta^{(\ell,b)} \right) \\ &= \sum_{\pi \in \underline{NCP}_2(k,\ell)} \left(\prod_{B \in K(\pi)} \left\langle \prod_{j \in B} A_j \right\rangle \right) \\ &+ \sum_{\pi \in \underline{NCP}_2(k,\ell)} \left(\prod_{B \in K(\pi)} \left\langle \prod_{j \in B \cap [k]} A_j \left(\prod_{j \in B \cap [k+1, k+j]} A_j \right)^t \right\rangle \right) \end{aligned}$$

which is the (real) second-order freeness of GOE and deterministic matrices established in [26]. Apart from these two special cases, the deterministic approximation in Theorem 2.9 mirrors the overall structure of the joint second-order distribution of Wigner and deterministic matrices described in [21, Eq. (3)]. We remark that resolvents and functions with an N -dependent mesoscopic scaling (1.5) as considered in [27] are usually not accessible in free probability theory as many of the standard techniques rely on explicit moment computations. Theorem 2.9 and its mesoscopic analog [27, Thm. 2.7] thus show that the underlying combinatorics of non-crossing annular permutations and marked partitions are, in fact, more general.

We further remark that the parallels between Theorem 2.9 and [21] continue to hold if we consider multiple independent Wigner matrices instead of one matrix W . More precisely, for n independent GUE (or GOE) matrices, the underlying combinatorial structure is given by the so-called *non-mixing* annular non-crossing permutations resp. *non-mixing* marked partitions for n colors (cf. Remark [27, Cor. 2.11]).

3 Proof of Theorem 2.4 (Formula for $m_{GUE}[\cdot|\cdot]$)

The proof of Theorem 2.4 consists of two main steps:

- (1) Solve the recursion for $m_{GUE}[\cdot|\cdot]$ in the case $A_1 = \dots = A_{k+\ell} = \text{Id}$. As $m[\cdot]$ can be expressed in terms of non-crossing graphs on the disk, we can rewrite the coefficients and the source term of the recursion in terms of graphs. By defining a suitable family of graphs on the (k, ℓ) -annulus, we extend this representation to obtain an ansatz for $\tilde{m}_{GUE}[\cdot|\cdot]$ and, by Definition 1.14, also an expression for the second-order free cumulant function.
- (2) Using the formula for $\tilde{m}_{GUE}[\cdot|\cdot]$ obtained in Step 1, write out the right-hand side of (2.11) and show that it satisfies the recursion for m_{GUE} . Checking that this ansatz also satisfies the remaining properties in Definition 2.1 yields equality between the two sides of (2.11).

We address the two steps in Sects. 3.1 and 3.2, respectively.

3.1 Step 1: Graphs

As we only consider \tilde{m}_{GUE} throughout this section, let $\kappa_4 = \sigma = \tilde{\omega} = 0$ and thus $\tilde{m}_{GUE}[\cdot|\cdot] = \tilde{m}[\cdot|\cdot]$. Recall from (1.16) and (1.17) that both $m[\cdot|\cdot]$ and $m_{\circ}[\cdot|\cdot]$ are expressible in terms of disk non-crossing graphs. In this section, we give analogous combinatorial formulas for $\tilde{m}[\cdot|\cdot]$ and $m_{\circ\circ}[\cdot|\cdot]$. For this task, we define a new, albeit closely related, multi-set of graphs on the (k, ℓ) -annulus. We start by introducing a transformation that translates between the disk and the annulus picture (cf. [22, Thm. 8 in Ch. 5]).

Definition 3.1 Fix $k, \ell \in \mathbb{N}$, $1 \leq j \leq \ell$, and consider a disk with the $k + \ell + 1$ labels $1, \dots, k, k + j, \dots, k + \ell, k + 1, \dots, k + j$ equidistantly placed around its boundary in clockwise order. We define a map τ , referred to as *mediating map*, that takes this picture to the (k, ℓ) -annulus as follows:

- (1) Use a homeomorphic continuous deformation, e.g., a conformal map, to map the disk and its labels to the $(k, \ell + 1)$ -annulus with a slit located between 1 and k on the outer circle and the two copies of $k + j$ on the inner circle.
- (2) Remove the slit to obtain an annulus.
- (3) Merge the two copies of the label $k + j$.

We visualize τ for an example in Fig. 9. The two labels $k + j$ are denoted as 6 and 6' to distinguish between them more easily.

The map τ induces a transformation of any graph Γ defined on the labeled disk to a graph defined on the (k, ℓ) -annulus. We denote the resulting annulus graph as $\tau(\Gamma)$. By construction, $\tau(\Gamma)$ is planar whenever Γ is a disk non-crossing graph. Recall that we use a slightly more general notion of planar graphs than the standard literature by allowing for loops and multi-edges. We give an example in Fig. 10. For better visibility, the loop arising in the last step is moved from between the two $(1, 3)$ edges to the right.

We can now introduce the family of graphs $\mathcal{G}(k, \ell)$ that constitute the key tool in the proof of Theorem 2.4. In analogy to the disk non-crossing graphs in Definition 1.7, we require the elements of $\mathcal{G}(k, \ell)$ to be drawn on the (k, ℓ) -annulus with the vertices placed around the boundary and the edges drawn in the interior of the annulus (see Fig. 10 and Example 3.4).

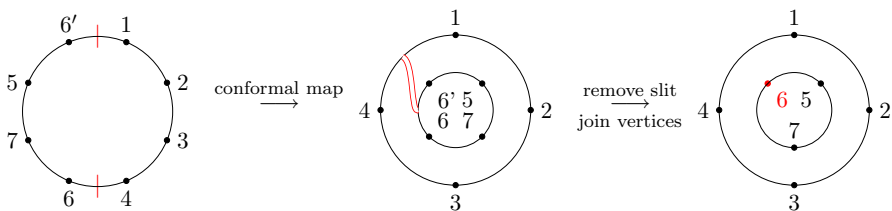


Fig. 9 The geometry of the transformation τ for $k = 4, \ell = 3$, and $j = 2$. (Color figure online)

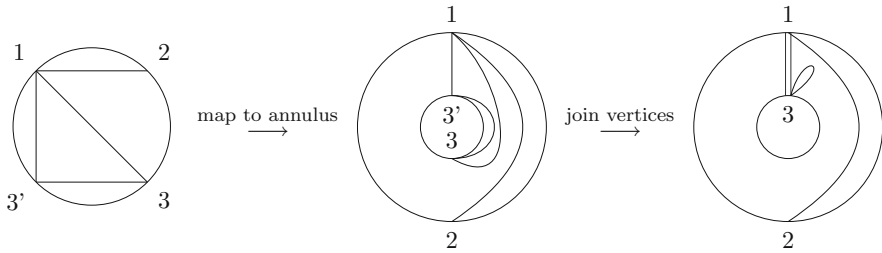


Fig. 10 Construction of $\tau(\Gamma)$ for a graph $\Gamma \in NCG(\{1, 2, 3, 3'\})$

Definition 3.2 For $k, \ell \in \mathbb{N}$ we define $\mathcal{G}([k], [k + 1, k + \ell])$ to be the multi-set¹¹ of undirected, planar graphs on the (k, ℓ) -annulus with vertex set $\{1, \dots, k + \ell\}$ and possible loops or double edges that is obtained from the following recursive construction:

(i) For any $S_1 \subset [k]$ and $S_2 \subset [k + 1, k + \ell]$ we have

$$\mathcal{G}(S_1, \emptyset) = \mathcal{G}(\emptyset, S_2) = \emptyset. \tag{3.1}$$

(ii) The multi-set $\mathcal{G}([k], [k + 1, k + \ell])$ can be constructed from the multi-sets $\mathcal{G}(S_1, [k + 1, k + \ell])$ with $S_1 \subsetneq [k]$ as follows: We define

$$\mathcal{G}_{-(1,k)} := \mathcal{G}_1 \cup \mathcal{G}_2 \cup \mathcal{G}_3 \cup \mathcal{G}_4, \tag{3.2}$$

as the disjoint union of the sets

$$\begin{aligned} \mathcal{G}_1 &:= \{\Gamma \cup \{1\} \mid \Gamma \in \mathcal{G}([2, k], [k + 1, k + \ell])\} \\ \mathcal{G}_2 &:= \bigcup_{j=2}^k \{\Gamma = \Gamma_1 \cup \Gamma_2 \mid \Gamma_1 \in NCG([1, j]) \\ &\quad \text{with edge } (1, j), \Gamma_2 \in \mathcal{G}([j, k], [k + 1, k + \ell])\}, \\ \mathcal{G}_3 &:= \bigcup_{j=1}^{k-1} \{\Gamma = \Gamma_1 \cup \Gamma_2 \mid \Gamma_1 \in \mathcal{G}([1, j], [k + 1, k + \ell]) \\ &\quad \text{with edge } (1, j), \Gamma_2 \in NCG([j, k])\}, \\ \mathcal{G}_4 &:= \bigcup_{j=1}^{\ell} \tau\left(\left\{\Gamma \in NCG(\{1, \dots, k, k + j, \dots, k + \ell, k + 1, \dots, k + j - 1, k + j\}) \mid \right. \right. \\ &\quad \left. \left. \Gamma \text{ has edge } (1, k + j)\right\}\right). \end{aligned}$$

Here, the union $\Gamma \cup \{1\}$ in \mathcal{G}_1 is to be understood as adding a separated vertex 1 to Γ while the union $\Gamma_1 \cup \Gamma_2$ in \mathcal{G}_2 and \mathcal{G}_3 refers to the graph with the vertex set $\{1, \dots, k + \ell\}$ and the edge set given by the union of the edge sets of Γ_1 and Γ_2 ,

¹¹ The graph Γ may be obtained in several different ways from the recursive construction. This is reflected by the multiplicity of Γ in the multi-set $\mathcal{G}([k], [k + 1, k + \ell])$.

respectively. Further, recall τ from Definition 3.1. We remark that all elements of $\mathcal{G}_1, \dots, \mathcal{G}_4$ are planar graphs. Next, let

$$\mathcal{G}_{(1,k)} := \{\Gamma \cup \{(1, k)\} | \Gamma \in \mathcal{G}_{-(1,k)}\},$$

where the union is to be understood as adding an edge $(1, k)$ to each graph in $\mathcal{G}_{-(1,k)}$. Observe that the resulting graphs are again planar. Finally, we define

$$\mathcal{G}([k], [k + 1, k + \ell]) := \mathcal{G}_{(1,k)} \cup \mathcal{G}_{-(1,k)}, \tag{3.3}$$

where any graphs that occur more than once are counted with multiplicity. In particular, whenever the same graph occurs in both $\mathcal{G}_{(1,k)}$ and $\mathcal{G}_{-(1,k)}$, the multiplicity of $\Gamma \in \mathcal{G}([k], [k + 1, k + \ell])$ is the total number of occurrences of Γ in both subsets.

- (iii) $\mathcal{G}(S_1, S_2) = \mathcal{G}(S_2, S_1)$ for any $S_1 \subset [k]$ and $S_2 \subset [k + 1, k + \ell]$ (in the sense that there is a well-defined bijective mapping that takes each element of $\mathcal{G}(S_1, S_2)$ to its counterpart).

We abbreviate $\mathcal{G}(k, \ell) := \mathcal{G}([k], [k + 1, k + \ell])$ and refer to its elements as *good graphs*. The subset of connected good graphs is denoted by $\mathcal{G}_c(k, \ell)$. Any edge (i, j) with $i \in [k]$ and $j \in [k + 1, k + \ell]$ is referred to as *connecting edge*.

We give a schematic overview of the elements in $\mathcal{G}(k, \ell)$ in Fig. 11. To avoid overcrowding, labels and edges that are not prescribed by the definition are left out.

Remark The occurrence of multi-edges or loops is inherent to the construction of $\mathcal{G}(k, \ell)$, which can be seen from a simple counting argument. First, note that an element of $\mathcal{G}(k, \ell)$ can have at most $2(k + \ell)$ edges by construction. To see this, observe that the elements of $\mathcal{G}_{-(1,k)}$ with the highest number of edges lie in \mathcal{G}_4 and that $\Gamma \in \mathcal{G}_4$ has the same number of edges as the underlying disk non-crossing graph. As any disk non-crossing graph on n vertices has at most $2n - 3$ edges (realized by a triangulation), the maximal number of edges for $\Gamma \in \mathcal{G}_4$ is $2(k + \ell + 1) - 3 = 2(k + \ell) - 1$. As (3.3) may add another edge to Γ , the maximum for $\mathcal{G}(k, \ell)$ is $2(k + \ell)$ edges. On the other hand, the maximal number of edges in a planar graph on the (k, ℓ) -annulus without multi-edges or loops is only $2(k + \ell)$ if $k, \ell \geq 3$ (again realized by a triangulation). It is readily seen that such a graph has strictly less than $2(k + \ell)$ edges if either k or ℓ is one or two. As the construction in Definition 3.2 does not introduce crossings

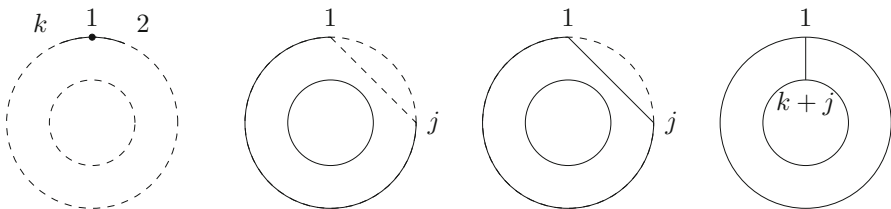


Fig. 11 Schematic visualization of the elements of $\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3,$ and \mathcal{G}_4 (left to right). Different subgraphs are indicated by different line styles

and begins with the cases $k, \ell \leq 2$, the difference between the two maxima must be reflected as multi-edges or loops.

We further remark that $\mathcal{G}(k, \ell)$ is a genuine multi-set, i.e., some graphs appear with multiplicity larger than one, unless $k = \ell = 1$ (cf. Lemma 3.5(d) and Fig. 13). The key property shared by all elements of $\mathcal{G}(k, \ell)$ is that each graph arises from some disk non-crossing graph along the recursive construction. Therefore, we may interpret \mathcal{G}_4 as a kind of source term. In view of Lemma 3.3, the multi-set $\mathcal{G}(k, \ell)$ gives an annulus analog to the disk non-crossing graphs, as it plays the same role in the combinatorial description of $\tilde{m}[\cdot|\cdot]$ as $NCG(S)$ does for $m[\cdot]$.

Lemma 3.3 *For fixed $k, \ell \in \mathbb{N}$, we have*

$$\tilde{m}[1, \dots, k|k + 1, \dots, k + \ell] = \left(\prod_{s=1}^{k+\ell} m_s \right) \sum_{\Gamma \in \mathcal{G}(k, \ell)} \prod_{(i, j) \in E(\Gamma)} q_{i, j}. \tag{3.4}$$

Proof As $q_{i, j} = q_{j, i}$ by (1.13) and $\mathcal{G}(k, \ell) = \mathcal{G}(\ell, k)$ by Definition 3.2(iii), it readily follows that the right-hand side of (3.4) is symmetric under the interchanging $[k]$ and $[k + 1, k + \ell]$. Similar to the proof of [6, Lem. 5.2], we use the combinatorial formula (3.4) as an ansatz to solve the recursion given in (1.27) and (1.28). First, observe that

$$\left(\prod_{s=1}^k m_s \right) \sum_{\Gamma \in \mathcal{G}(S_1, \emptyset)} \prod_{(i, j) \in E(\Gamma)} q_{i, j} = \left(\prod_{s=k+1}^{k+\ell} m_s \right) \sum_{\Gamma \in \mathcal{G}(\emptyset, S_2)} \prod_{(i, j) \in E(\Gamma)} q_{i, j} = 0$$

for any $S_1 \subset [k]$ and $S_2 \subset [k + 1, k + \ell]$ due to the sums being empty. Hence, the initial condition (1.27) is satisfied.

It remains to check (1.28). We introduce the notation

$$q_\Gamma := \prod_{(i, j) \in E(\Gamma)} q_{i, j}$$

and conclude from the decompositions (3.3) and (3.2) that

$$\begin{aligned} \sum_{\Gamma \in \mathcal{G}(k, \ell)} q_\Gamma &= (1 + q_{1, k}) \sum_{\Gamma \in \mathcal{G}_{\rightarrow(1, k)}} q_\Gamma \\ &= (1 + q_{1, k}) \left(\sum_{\Gamma \in \mathcal{G}_1} q_\Gamma + \sum_{\Gamma \in \mathcal{G}_2} q_\Gamma + \sum_{\Gamma \in \mathcal{G}_3} q_\Gamma + \sum_{\Gamma \in \mathcal{G}_4} q_\Gamma \right). \end{aligned} \tag{3.5}$$

Noting that the vertex 1 in $\Gamma \in \mathcal{G}_1$ has no adjacent edges, we may write

$$\sum_{\Gamma \in \mathcal{G}_1} q_\Gamma = \sum_{\Gamma \in \mathcal{G}([2, k], [k+1, k+\ell])} q_\Gamma.$$

Further, the transformation τ only changes the geometry underlying a graph Γ , but does not influence its edge set. By the definition of \mathcal{G}_4 , any $\Gamma \in NCG([1, \dots, k + \ell, k + j])$

used in the construction must have at least one edge $(1, k + j)$. As a consequence, the product q_Γ always includes the factor $q_{1,k+j} = m_1 m_{k+j} (1 + q_{1,k+j})$. This yields

$$\sum_{\Gamma \in \mathcal{G}_4} q_\Gamma = \sum_{j=1}^{\ell} (m_1 m_{k+j} \sum_{\Gamma \in NCG(\{1, \dots, k+\ell, k+j\})} q_\Gamma).$$

Note that the identity $q_{1,k+j} = m_1 m_j (1 + q_{1,k+j})$ allows writing summations restricted to graphs with an edge $(1, k + j)$ on the left-hand side into an unrestricted sum over all graphs on the right-hand side. We use the same trick for \mathcal{G}_2 and \mathcal{G}_3 as $q_{\Gamma_1 \cup \Gamma_2} = q_{\Gamma_1} q_{\Gamma_2}$ for the union of graphs introduced in Definition 3.2. With these replacements, (3.5) can be written as

$$\begin{aligned} & \frac{1}{1 + q_{1,k}} \sum_{\Gamma \in \mathcal{G}(k, \ell)} q_\Gamma \\ &= \sum_{\Gamma \in \mathcal{G}(\{2, k, [k+1, k+\ell]\})} q_\Gamma + \sum_{j=2}^k m_1 m_j \left(\sum_{\Gamma \in NCG(\{1, j\})} q_\Gamma \right) \left(\sum_{\Gamma \in \mathcal{G}(\{j, k, [k+1, k+\ell]\})} q_\Gamma \right) \\ &+ \sum_{j=1}^{k-1} m_1 m_j \left(\sum_{\Gamma \in \mathcal{G}(\{1, j, [k+1, k+\ell]\})} q_\Gamma \right) \left(\sum_{\Gamma \in NCG(\{j, k\})} q_\Gamma \right) \\ &+ \sum_{j=1}^{\ell} m_1 m_{k+j} \left(\sum_{\Gamma \in NCG(\{1, \dots, k+\ell, k+j\})} q_\Gamma \right). \end{aligned}$$

Multiplying both sides with $\prod_{s=1}^{k+\ell} m_s$ and noting that $1 + q_{1,j} = (1 - m_1 m_j)^{-1}$, we see that the right-hand side of (3.4) satisfies (1.28) as claimed.

Example 3.4 We have $|\mathcal{G}(1, 1)| = 8$. The graphs are visualized in Fig. 12.

In particular, we readily reobtain (1.30) by evaluating q_Γ for every graph in the above list. Note that decomposing (1.30) into the form $m_1 m_2 \sum_{\Gamma \in \mathcal{G}(1, 1)} q_\Gamma$ is possible in multiple ways. However, picking graphs that contain a connecting edge yields the set $\mathcal{G}(1, 1)$ in Fig. 12 as the smallest possible set.

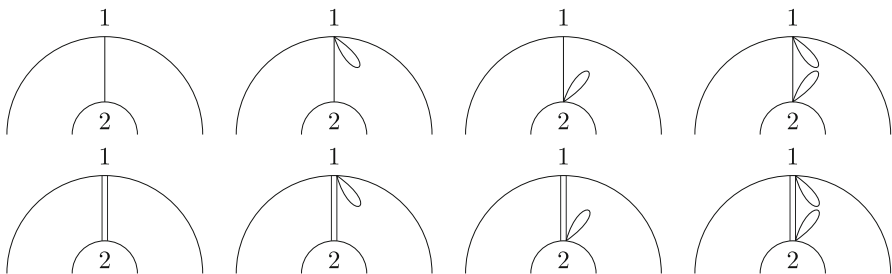


Fig. 12 The elements of $\mathcal{G}(1, 1)$

We state a few properties of the elements of $\mathcal{G}(k, \ell)$. In Lemma 3.5, we focus on general characteristics of $\Gamma \in \mathcal{G}(k, \ell)$ as a planar graph drawn on the (k, ℓ) -annulus. The properties of $\mathcal{G}(k, \ell)$ that are needed for the proof of the combinatorial formula for m_{∞} are given in the separate Lemma 3.6.

Lemma 3.5 *Let $k, \ell \in \mathbb{N}$.*

- (a) *The connected components of any $\Gamma \in \mathcal{G}(k, \ell)$ give rise to a non-crossing partition of the (k, ℓ) -annulus. In particular, Γ contains a connecting edge if $k, \ell \geq 1$.*
- (b) *$\Gamma \in \mathcal{G}(k, \ell)$ may have at most two loops. Loops only occur at vertices that are adjacent to a connecting edge. Two loops on vertices on the same circle do not occur.*
- (c) *$\Gamma \in \mathcal{G}(k, \ell)$ may have up to $k + \ell - 1$ double edges. Double edges are either connecting edges or adjacent to a connecting edge. Edges with a multiplicity higher than two do not occur.*
- (d) *$\mathcal{G}(k, \ell)$ is a genuine multi-set unless $k = \ell = 1$.*

Proof (a) We use proof by induction. First, note that the elements of $\mathcal{G}(S_1, \emptyset) = \emptyset$ and $\mathcal{G}(\emptyset, S_2) = \emptyset$ with $S_1 \subseteq [k]$ and $S_2 \subseteq [k + 1, k + \ell]$ clearly give rise to an annular non-crossing partition. Moreover, Example 3.4 establishes the claim in the case $k = \ell = 1$ and shows that any $\Gamma \in \mathcal{G}(1, 1)$ contains at least one connecting edge.

Assume next that the elements of $\mathcal{G}(S_1, [k + 1, k + \ell])$ give rise to an annular non-crossing partition for any $S_1 \subset [k]$ with $|S_1| \leq k - 1$ and a fixed $\ell \geq 1$. We aim to show that the connected components of any $\Gamma \in \mathcal{G}(k, \ell)$ also correspond to the blocks of some $\pi \in NCP(k, \ell)$. Due to the symmetry induced by Definition 3.2(iii), this is enough to establish the induction step.

By definition, the vertices 1 and k lie next to each other on the outer circle. Hence, adding an edge $(1, k)$ may connect two connected components, but cannot introduce a crossing in the partition obtained from them. It is, therefore, sufficient to check the claim for elements of $\mathcal{G}_{-(1,k)}$ or, equivalently, for the sets $\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3$, and \mathcal{G}_4 in (3.2). First, note that the transformation τ indeed takes the disk partition induced by the disk non-crossing graph to an annular non-crossing partition. This is due to the continuous homeomorphism used in the definition of τ . Moreover, the edge $(1, k + j)$ prescribed for $\Gamma \in NCG(\{1, \dots, k, k + j, \dots, k + j - 1, k + j\})$ by the definition is mapped to a connecting edge, ensuring that the resulting partition has a connecting block.

By construction, the graphs in $\mathcal{G}_1, \mathcal{G}_2$, and \mathcal{G}_3 contain an element of $\mathcal{G}(S_1, [k + 1, k + \ell])$ with some $S_1 \subset [k]$ as a subgraph. Applying the induction hypothesis for $\mathcal{G}([2, k], k + 1, k + \ell)$ and noting that a separate vertex 1 only adds a singleton set to the underlying partition, we can conclude that \mathcal{G}_1 , too, behaves as claimed. The argument for \mathcal{G}_2 and \mathcal{G}_3 is similar. Here, the key observation is that the connected components of the added disk non-crossing graph induce a non-crossing partition of an interval placed along the outer circle. Recalling that all elements of $NCP(k, \ell)$ have at least one connecting block, the corresponding connected component of $\Gamma \in \mathcal{G}(k, \ell)$ must contain a connecting edge.

(b) It is readily seen that the elements of $\mathcal{G}(S_1, \emptyset) = \emptyset$ and $\mathcal{G}(\emptyset, S_2) = \emptyset$ with $S_1 \subseteq [k]$ and $S_2 \subseteq [k + 1, k + \ell]$, as well as all $\Gamma \in \mathcal{G}(1, 1)$ have the claimed structure. Moreover, adding an edge $(1, k)$ to a graph cannot introduce a loop unless $k = 1$, so it is again sufficient to establish the induction step for $\mathcal{G}_{-(1,k)}$.

By Definition 1.7, a disk non-crossing graph does not contain any loops. Hence, any loops of $\Gamma \in \mathcal{G}_1 \cup \mathcal{G}_2 \cup \mathcal{G}_3$ must occur in the subgraph in $\mathcal{G}(S_1, [k + 1, k + \ell])$ with $S_1 \subset [k]$ used to construct Γ . As the induction hypothesis applies to this subgraph, we conclude that Γ has at most two loops that satisfy the claimed placement rules, respectively. For $\Gamma \in \mathcal{G}_4$, note that applying τ to an element of $NCG(\{1, \dots, k, k + j, \dots, k + j - 1, k + j\})$ yields an annulus graph with a loop if and only if the disk non-crossing graph contains an edge $(k + j, k + j)$. As the latter must also contain an edge $(1, k + j)$ by definition, a loop in Γ is always adjacent to a connecting edge.

We remark that the loops in $\Gamma \in \mathcal{G}(k, \ell)$ are a consequence of the definition of τ rather than an artifact of the recursive construction of $\mathcal{G}(k, \ell)$, i.e., they can be created only once. In particular, Γ can only contain a loop if it arises from \mathcal{G}_4 or its analog in a previous iteration. Hence, the construction cannot yield a graph with more than two loops or more than one loop per circle, respectively. In particular, any graph in $\mathcal{G}(k, \ell)$ containing more than one loop is necessarily obtained from an element in $\mathcal{G}(1, 1)$ with two loops.

(c) Again, the elements of $\mathcal{G}(S_1, \emptyset) = \emptyset$ and $\mathcal{G}(\emptyset, S_2) = \emptyset$ with $S_1 \subseteq [k]$ and $S_2 \subseteq [k + 1, k + \ell]$, as well as all $\Gamma \in \mathcal{G}(1, 1)$ have the claimed structure. We further note that any double edges of $\Gamma \in \mathcal{G}_1 \cup \mathcal{G}_2 \cup \mathcal{G}_3$ must occur in the subgraph in $\mathcal{G}(S_1, [k + 1, k + \ell])$ with $S_1 \subset [k]$ used to construct Γ and that the induction hypothesis applies to the latter. For $\Gamma \in \mathcal{G}_4$, a double edge occurs if and only if the corresponding element of $NCG(\{1, \dots, k, k + j, \dots, k + j - 1, k + j\})$ has a vertex i that shares an edge with both copies of $k + j$. However, there is at most one such vertex in $[k]$ and $[k + 1, k + \ell]$, respectively, as having two vertices i, i' connected to both copies of $k + j$ in either set induces a crossing. In particular, any double edge shares a vertex with the edge $(1, k + j) \in \Gamma$ prescribed by the definition. Hence, all graphs in $\mathcal{G}_{-(1,k)}$ have the claimed structure.

It remains to consider $\mathcal{G}_{(1,k)}$, i.e., to add an edge $(1, k)$ to the graphs considered previously. In the case $j = k$ of \mathcal{G}_2 , this doubles an existing $(1, k)$ edge. By part (a), any such graph must also have an edge connecting k with a vertex in $[k + 1, k + \ell]$. This shows that the placement rule for double edges is satisfied. Further, at most one doubled edge can be added with each application of the recursion, i.e., there are at most $k + \ell - 1$ double edges in total. It is readily checked that two is indeed the highest edge multiplicity possible.

(d) As the case $k = \ell = 1$ has already been discussed in Example 3.4, let $k = 2, \ell = 1$, and consider $\tau(NCG(\{1, 2, 3, 3\}))$. We relabel the vertices as $1, 2, 3, 3'$ to distinguish between the two copies of the doubled vertex 3 more easily. Observe that the graphs Γ_1 and Γ_2 with edge sets $\{(1, 3'), (2, 3')\}$ and $\{(1, 3'), (2, 3)\}$, respectively, give rise to the same element of $\mathcal{G}(2, 1)$, namely the annulus graph with edge set $\{(1, 3), (2, 3)\}$ (see Fig. 13).

Hence, $\mathcal{G}(1, 2)$ is indeed a multi-set. By Definition 3.2, either $\mathcal{G}(1, 2)$ or $\mathcal{G}(2, 1)$ is used in the construction of $\mathcal{G}(k, \ell)$ for $k, \ell > 2$, i.e., the construction yields again a multi-set.

For the following discussion, we introduce the disjoint decomposition

$$\mathcal{G}(k, \ell) = \mathcal{G}_{dec}(k, \ell) \cup \mathcal{G}_{-dec}(k, \ell), \tag{3.6}$$

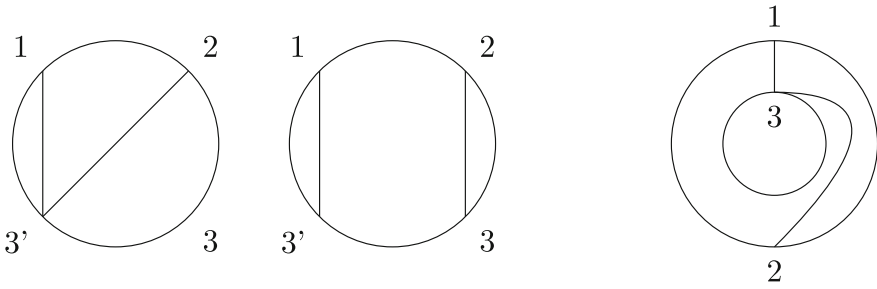


Fig. 13 Two disk non-crossing graphs (left) that give rise to the same annulus graph (right)

where $\mathcal{G}_{dec}(k, \ell)$ denotes the graphs in $\mathcal{G}(k, \ell)$ that have double edges or loops and $\mathcal{G}_{-dec}(k, \ell)$ denotes the graphs that do not have either. We refer to the elements of $\mathcal{G}_{dec}(k, \ell)$ as decorated graphs.

Lemma 3.6 *Let $k, \ell \in \mathbb{N}$.*

- (a) *Whenever $\Gamma \in \mathcal{G}(k, \ell)$ has more than one connected component that contains a connecting edge, every connected component of Γ can be uniquely identified with a disk non-crossing graph.*
- (b) *Whenever $\Gamma \in \mathcal{G}(k, \ell)$ has exactly one connected component with a connecting edge and Γ is not decorated, the same identification as in (a) holds, but it is no longer unique. If Γ_1 denotes the connected component of Γ that contains a connecting edge and $U_1 \cup U_2$ with $U_1 \subset [k]$ and $U_2 \subset [k + 1, k + \ell]$ is the vertex set of Γ_1 , there are $|U_1| \cdot |U_2|$ different ways to identify the connected components of Γ with a disk non-crossing graph.*

Lemma 3.6 translates between a graph $\Gamma \in \mathcal{G}(k, \ell)$, the partition induced by its connected components and the cycle structure arising in (2.11). We give a schematic of the construction of the disk graphs in Fig. 14. Recall from Definition 1.11 that any cycle of an annular non-crossing permutation encloses a region homeomorphic to the unit disk with the boundary oriented clockwise.

Note that the assignment of the orientation in the second step on the left of Fig. 14 is not unique if π_Γ has only one connecting block (cf. [23, Prop. 4.6], see Fig. 5 for an example). We remark that (a) and (b) are almost complementary cases and that only decorated graphs with exactly one connected component containing a connecting edge

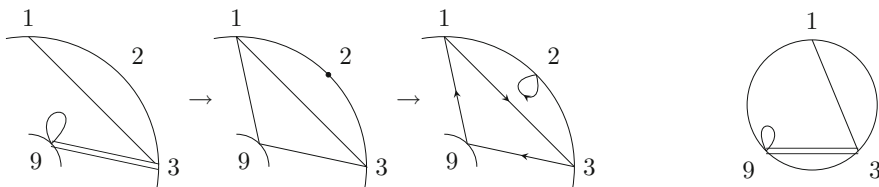


Fig. 14 A subgraph of $\Gamma \in \mathcal{G}(k, \ell)$ with the induced partition π_Γ and a possible permutation π'_Γ (left) as well as the disk non-crossing graph obtained for a connected component of Γ (right)

are not covered by Lemma 3.6. The proof of (a) further shows that any $\Gamma \in \mathcal{G}(k, \ell)$ with at least two connected components containing a connecting edge cannot be decorated.

We give the full construction behind the schematics in Fig. 14.

Proof of Lemma 3.6 (b) It follows from Lemma 3.5(a) that the connected components of Γ give rise to a partition $\pi_\Gamma \in NCP(k, \ell)$. By Propositions 4.4. and 4.6 of [23], any element of $NCP(k, \ell)$ may be identified with an annular non-crossing permutation, i.e., its blocks can be given an orientation (see the left of Fig. 14). This orientation of the individual cycles is naturally induced by the orientation of the inner and outer circle. However, the identification between $NCP(k, \ell)$ and $\overline{NCP}(k, \ell)$ is only unique whenever the underlying partition has more than one connecting block. If there is only one connecting block $U_1 \cup U_2$ with $U_1 \subset [k], U_2 \subset [k + 1, k + \ell]$, there are $|U_1| \cdot |U_2|$ possibilities to identify π_Γ with an annular non-crossing permutation (cf. Fig. 5). We fix one $\pi'_\Gamma \in \overline{NCP}(k, \ell)$ that is associated with π_Γ in this way.

By definition, the cycles of π'_Γ can be drawn on the (k, ℓ) -annulus such that each cycle encloses a region between the circles homeomorphic to the disk with boundary oriented clockwise. Adding the elements of the cycle around the boundary yields a labeled disk as in Definition 1.7. We may use the same transformation to map a connected component of Γ to a disk non-crossing graph (see Fig. 15). Note that this transformation cannot induce any crossings of the edges of Γ , however, any loops or double edges of the original graph are kept. Hence, we only obtain a disk non-crossing graph in the sense of Definition 1.7 if Γ does not contain any loops or double edges to begin with.

(a) Assume next that $\Gamma \in \mathcal{G}(k, \ell)$ has at least two connected components that contain a connecting edge. It follows from Definition 3.2 that this structure can only arise from \mathcal{G}_4 in (3.2) or its analog in a previous iteration of the recursive definition. Since adding subgraphs that only live on one circle of the (k, ℓ) -annulus does not interfere with the following argument, assume w.l.o.g. that Γ arises from \mathcal{G}_4 directly. To have two connecting blocks, Γ must have at least two connecting edges (i_1, i_2) and (i'_1, i'_2) where $i_1, i'_1 \in [k], i_2, i'_2 \in [k + 1, k + \ell]$ and $i_1 \neq i'_1, i_2 \neq i'_2$. Note that either (i_1, i_2) or (i'_1, i'_2) may coincide with the edge $(1, k + j)$ prescribed by the definition.

Let $\tilde{\Gamma} \in NCG(\{1, \dots, k, k + j, \dots, k + j - 1, k + j\})$ denote a disk non-crossing graph such that $\tau(\tilde{\Gamma}) = \Gamma$. Here, τ denotes the map introduced in Definition 3.1. By construction, $\tilde{\Gamma}$ also has two edges (i_1, i_2) and (i'_1, i'_2) with $i_1, i'_1 \in [k], i_2, i'_2 \in [k + 1, k + \ell]$ and $i_1 \neq i'_1, i_2 \neq i'_2$. This structure imposes several restrictions on $\tilde{\Gamma}$, as can be seen from Fig. 16.

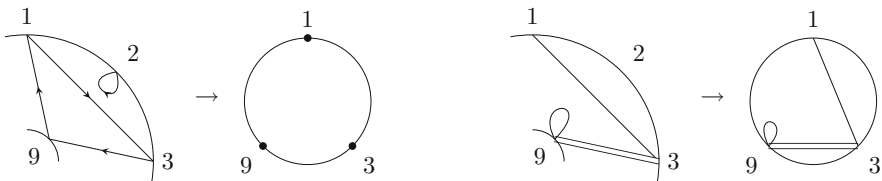


Fig. 15 Transformation of a cycle of $\pi'_\Gamma \in \overline{NCP}(k, \ell)$ to a circle (left) and the induced transformation of a connected component of $\Gamma \in \mathcal{G}(k, \ell)$ into a disk graph (right)

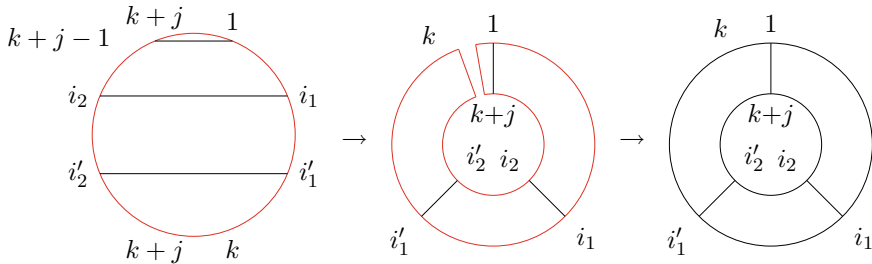


Fig. 16 A schematic visualization of $\tilde{\Gamma}$ (left) and its mapping to the original graph $\Gamma = \tau(\tilde{\Gamma})$ (right). (Color figure online)

First, there cannot be an edge $(1, k) \in \tilde{\Gamma}$ without violating the non-crossing condition. Together with (3.3), this implies that $\Gamma = \tau(\tilde{\Gamma})$ has at most a single edge $(1, k)$. Further, $\tilde{\Gamma}$ cannot contain a vertex that connects to both copies of $k + j$. This implies that $\Gamma = \tau(\tilde{\Gamma})$ cannot have any double edges. Lastly, Γ cannot have any loops, as $\tilde{\Gamma}$ containing an edge $(k + j, k + j)$ would induce a crossing, too. Hence, any $\Gamma \in \mathcal{G}(k, \ell)$ with more than one connected component containing a connecting edge has only single edges and no loops.

So far, we have only considered the non-crossing annular partition induced by an element of $\mathcal{G}(k, \ell)$. The following lemma allows us to partially reverse this relation and explicitly construct a graph that is associated with a given $\pi \in NCP(k, \ell)$.

Lemma 3.7 *For every $\pi \in NCP(k, \ell)$ there is at least one $\Gamma \in \mathcal{G}(k, \ell)$ for which the vertex sets of the connected components coincide with the blocks of π . If π has exactly one connecting block, then there is at least one such graph in $\mathcal{G}_{dec}(k, \ell)$ and one in $\mathcal{G}_{-dec}(k, \ell)$, respectively.*

We briefly sketch the construction of an element on $\mathcal{G}(k, \ell)$ from a given annular non-crossing partition. For the example, we assume that the sketched connecting block is the only one in the partition. After completing the steps sketched in Fig. 17, the remaining connected components of the graph are readily added using steps corresponding to $\mathcal{G}_1, \mathcal{G}_2$, and \mathcal{G}_3 in Definition 3.2 and the symmetry under interchanging of the inner and outer circle.

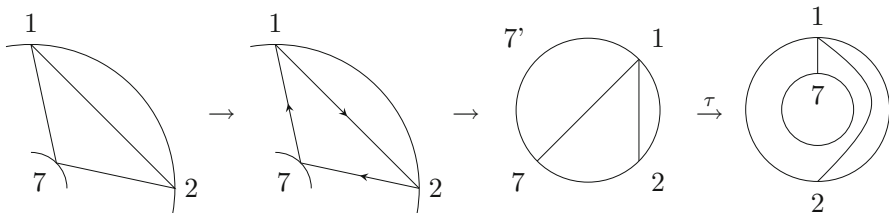


Fig. 17 The construction of an element of $\mathcal{G}(2, 1)$ (right) from the connecting block of an annular non-crossing partition (left)

Proof of Lemma 3.7 Fix $\pi \in NCP(k, \ell)$ and assume first that π has at least two connecting blocks. W.l.o.g. let 1 and k be assigned to different blocks of π . Indeed, if 1 and k occurred in the same block, we may split it into two disjoint parts that contain 1 and k , respectively, and later add an edge $(1, k)$ to the graph obtained from this modified partition. Further, assume w.l.o.g. that 1 is contained in a connecting block. If $i > 1$ were the smallest element that occurs in a connecting block, we may further remove any blocks containing $\{1, \dots, i - 1\}$ from the partition and later add a suitable subgraph. Note that the latter is possible by only using steps corresponding to \mathcal{G}_1 and \mathcal{G}_2 in Definition 3.2.

Under these assumptions, there exists $\tilde{\pi} \in NCP(\{1, \dots, k, k + j, \dots, k + j - 1, k + j\})$ such that the transformation τ in Definition 3.1 maps the blocks of $\tilde{\pi}$ to the blocks of π . This is readily seen from Fig. 9, as 1 and k being in different blocks implies that none of the blocks of π intersects the slit between 1 and k . We pick j such that 1 and $k + j$ are in the same block of π and may further choose $\tilde{\pi}$ such that one copy of the doubled label $k + 1$ occurs as a singleton set. Finally, define $\tilde{\Gamma} \in NCG(\{1, \dots, k, k + 1, \dots, k + \ell, k + 1\})$ by considering each block $B = \{i_1, \dots, i_n\} \in \tilde{\pi}$ separately and adding the edges $(i_1, i_2), \dots, (i_1, i_n)$ to the graph. By construction, the vertex sets of the connected components of $\tilde{\Gamma}$ coincide with the blocks of $\tilde{\pi}$. Considering $\Gamma_\pi = \tau(\tilde{\Gamma})$ now yields the element of $\mathcal{G}(k, \ell)$ with the claimed properties (cf. Fig. 17, where the procedure is sketched for a single block).

Next, consider $\pi \in NCP(k, \ell)$ that has only one connecting block $U = U_1 \cup U_2$ with $U_1 \subseteq [k], U_2 \subseteq [k + 1, k + \ell]$. Similar to the first case, we can construct a disk non-crossing graph Γ_U for which $\tau(\Gamma_U) \in \mathcal{G}_c(|U_1|, |U_2|)$. As $\tau(\Gamma_U)$ is only required to have one connecting edge, there is also a choice for Γ_U containing a $(k + j, k + j)$ edge (cf. Fig. 15). In particular, we obtain at least one graph in $\mathcal{G}_{c,-dec}(|U_1|, |U_2|)$ and one graph in $\mathcal{G}_{c,dec}(|U_1|, |U_2|)$, respectively. The remaining connected subgraphs of Γ_π are then added recursively by alternating between adding an isolated vertex (cf. \mathcal{G}_1) and a disk non-crossing graph (cf. \mathcal{G}_2 and \mathcal{G}_3) to $\tau(\Gamma_U)$. Note that starting from $\tau(\Gamma_U) \in \mathcal{G}_{c,-dec}(|U_1|, |U_2|)$ yields a graph that is not decorated while starting from $\tau(\Gamma_U) \in \mathcal{G}_{c,dec}(|U_1|, |U_2|)$ yields a decorated graph.

Using the properties of $\Gamma \in \mathcal{G}(k, \ell)$ from Lemmas 3.6 and 3.7, we obtain an explicit non-recursive combinatorial formula for m_{oo} .

Lemma 3.8 *Let $k, \ell \in \mathbb{N}$. Then,*

$$m_{oo}[1, \dots, k|k + 1, \dots, k + \ell] = \left(\prod_{s=1}^{k+\ell} m_s \right) \sum_{\Gamma \in \mathcal{G}_c(k, \ell)} c_\Gamma q_\Gamma, \tag{3.7}$$

with suitable constants $c_\Gamma \in \mathbb{Z}$. In particular, $c_\Gamma = 1$ for $\Gamma \in \mathcal{G}_{dec}(k, \ell)$.

Note that the constants c_Γ for $\Gamma \in \mathcal{G}_{-dec}(k, \ell)$ are readily obtained from the multiplicity of the corresponding graph in the multi-set, however, their exact values are not needed for the proof of Theorem 2.4.

Proof We start by observing that any connected component of $\Gamma \in \mathcal{G}(k, \ell)$ with a connecting edge is itself an annular non-crossing graph. Further, a connected component of Γ that only involves vertices from either $[k]$ or $[k + \ell]$ cannot contain loops or double edges (cf. (b) and (c) of Lemma 3.5) and we may identify it with a disk non-crossing graph.

To simplify notation, decompose any $B \subseteq S_1 \cup S_2$ as a union $B_1 \cup B_2$ with $B_1 \subseteq [k]$, $B_2 \subseteq [k + 1, k + \ell]$ and associate it with a tuple (B_1, B_2) if neither of the subsets is empty. This allows us to use the common notation $NCG(B)$ for both disk and annular graphs by setting $NCG(B) = \mathcal{G}(B_1, B_2)$ whenever B contains elements from both $[k]$ and $[k + 1, k + \ell]$, and $NCG(B) = NCG(B_1)$ resp. $NCG(B) = NCG(B_2)$ if it does not. Splitting the sum in (3.4) according to the underlying partition, we can write

$$\begin{aligned} \tilde{m}[1, \dots, k|k + 1, \dots, k + \ell] &= \left(\prod_{s=1}^{k+\ell} m_s \right) \sum_{\Gamma \in \mathcal{G}(k, \ell)} q_\Gamma \\ &= \sum_{\pi \in NCP(k, \ell)} \prod_{B \in \pi} \left[\left(\prod_{s \in B} m_s \right) \sum_{\Gamma \in NCG_c(B)} q_\Gamma \right] \end{aligned} \tag{3.8}$$

where $NCG_c(B)$ denotes the connected graphs in $NCG(B)$. By Lemma 3.7, none of the sums on the right-hand side are empty. However, there may be multiple permutations in $NCP(k, \ell)$ as well as a marked element of $NCP(k) \times NCP(\ell)$ associated with a given annular non-crossing partition $\pi \in NCP(k, \ell)$. To obtain the same structure as in (1.26), we thus need to decompose the sum over $\pi \in NCP(k, \ell)$ on the right-hand side of (3.8) further.

Distinguishing by the number of connecting blocks of π yields

$$\begin{aligned} &\sum_{\pi \in NCP(k, \ell)} \prod_{B \in \pi} \left[\left(\prod_{s \in B} m_s \right) \sum_{\Gamma \in NCG_c(B)} q_\Gamma \right] \tag{3.9} \\ &= \sum_{\substack{\pi \in NCP(k, \ell), \\ 1 \text{ conn. block}}} \prod_{B \in \pi} \left[\left(\prod_{s \in B} m_s \right) \sum_{\Gamma \in NCG_c(B)} q_\Gamma \right] \\ &+ \sum_{\substack{\pi \in NCP(k, \ell), \\ \geq 2 \text{ conn. blocks}}} \prod_{B \in \pi} \left[\left(\prod_{s \in B} m_s \right) \sum_{\Gamma \in NCG_c(B)} q_\Gamma \right]. \end{aligned}$$

As an annular non-crossing partition with at least two connecting blocks uniquely corresponds to an annular non-crossing permutation (cf. [23, Prop. 4.4]), we may replace $NCP(k, \ell)$ by $NCP(k, \ell)$ in the summation if we interpret a cycle as an ordered set. Together with Lemma 3.6 and (1.17), it follows that

$$\sum_{\substack{\pi \in NCP(k, \ell), \\ \geq 2 \text{ conn. blocks}}} \prod_{B \in \pi} \left[\left(\prod_{s \in B} m_s \right) \sum_{\Gamma \in NCG_c(B)} q_\Gamma \right]$$

$$\begin{aligned}
 &= \sum_{\substack{\pi \in \overline{NCP}(k, \ell), \\ \geq 2 \text{ conn. cycles}}} \prod_{B \in \pi} \left[\left(\prod_{s \in B_1 \cup B_2} m_s \right) \sum_{\Gamma \in \overline{NCG}_c(B_1 \cup B_2)} q_\Gamma \right] \\
 &= \sum_{\substack{\pi \in \overline{NCP}(k, \ell), \\ \geq 2 \text{ conn. cycles}}} \prod m_o[B_1 \cup B_2].
 \end{aligned} \tag{3.10}$$

where we used that m_o is invariant under permutation of the spectral parameters $(z_j)_j$.

Applying a similar argument for the partitions with one connecting block U , interpreted as (U_1, U_2) with $U_1 = U \cap [k], U_2 = U \cap [k + 1, k + \ell]$, yields

$$\begin{aligned}
 &\sum_{\substack{\pi \in \overline{NCP}(k, \ell), \\ 1 \text{ conn. block}}} \prod_{B \in \pi} \left[\left(\prod_{s \in B} m_s \right) \sum_{\Gamma \in \overline{NCG}_c(B)} q_\Gamma \right] \\
 &= \sum_{\substack{\pi \in \overline{NCP}(k, \ell), \\ 1 \text{ conn. block}}} \left[\left(\prod_{s \in U} m_s \right) \sum_{\Gamma \in \mathcal{G}_c(U)} q_\Gamma \right] \prod_{B \in \pi \setminus \{U\}} \left[\left(\prod_{s \in B} m_s \right) \sum_{\Gamma \in \overline{NCG}_c(B)} q_\Gamma \right] \\
 &= \sum_{\substack{\pi \in \overline{NCP}(k, \ell), \\ 1 \text{ conn. block}}} \left[\left(\prod_{s \in U} m_s \right) \sum_{\Gamma \in \mathcal{G}_c(U)} q_\Gamma \right] \prod_{B \in \pi \setminus \{U\}} m_o[B]
 \end{aligned} \tag{3.11}$$

by (1.17). Note that $B \neq U$ cannot involve both sets $[k]$ and $[k + 1, k + \ell]$, as U is the only connecting block. Using (3.6), we decompose

$$\sum_{\Gamma \in \mathcal{G}_c(U)} q_\Gamma = \sum_{\Gamma \in \mathcal{G}_{c,dec}(U)} q_\Gamma + \sum_{\Gamma \in \mathcal{G}_{c,-dec}(U)} q_\Gamma$$

and recall from Lemma 3.7 that neither sum on the right-hand side is empty. Note that the split induced by (3.6) also decomposes the right-hand side of (3.11) into two terms.

Next, consider the term corresponding to $\mathcal{G}_{c,-dec}(U)$ and recall that any $\Gamma \in \mathcal{G}_{c,-dec}(U)$ can be identified with a disk non-crossing graph by Lemma 3.6(b). However, this identification is not unique. Decompose U into $U_1 = U \cap [k], U_2 = U \cap [k + 1, k + \ell]$. Then there are $|U_1| \cdot |U_2|$ different disk graphs that can be obtained from a given $\Gamma \in \mathcal{G}_{c,-dec}(U)$. As the resulting graphs only differ in the labeling of the vertices and m_o is invariant under permutation of its arguments, the contribution of each graph to the sum is the same. Recall that the number of annular non-crossing permutations arising from $\pi \in \overline{NCP}(k, \ell)$ is equal to $|U_1| \cdot |U_2|$ by [23, Prop. 4.6]. We thus write

$$\sum_{\Gamma \in \mathcal{G}_{c,-dec}(U)} q_\Gamma = |U_1| \cdot |U_2| \sum_{\Gamma \in \overline{NCG}_c(U_1 \cup U_2)} q_\Gamma + \sum_{\Gamma \in \mathcal{G}_{c,-dec}(U)} c_\Gamma q_\Gamma.$$

with suitable constants $c_\Gamma \in \mathbb{Z}$. In particular, we do not necessarily have $c_\Gamma = 1$. This can be seen from considering the element of $\mathcal{G}(2, 1)$ that has the edge set

$\{(1, 2), (1, 3), (2, 3)\}$. Hence,

$$\sum_{\Gamma \in \mathcal{G}_c(U)} q_\Gamma = |U_1| \cdot |U_2| \sum_{\Gamma \in NCG_c(U_1 \cup U_2)} q_\Gamma + \sum_{\Gamma \in \mathcal{G}_c(U)} c_\Gamma q_\Gamma. \tag{3.12}$$

where we set $c_\Gamma = 1$ for $\Gamma \in \mathcal{G}_{c,dec}(U)$. Note that the contribution of the first term of (3.12) to (3.11) is

$$\begin{aligned} & \sum_{\substack{\pi \in NCP(k, \ell), \\ 1 \text{ conn. block } U}} \left[\left(\prod_{s \in U} m_s \right) \cdot |U_1| \cdot |U_2| \sum_{\Gamma \in NCG_c(U_1 \cup U_2)} q_\Gamma \right] \prod_{B \in \pi \setminus \{U\}} m_o[B] \\ &= \sum_{\substack{\pi \in NCP(k, \ell), \\ 1 \text{ conn. block}}} \prod_{B \in \pi} m_o[B_1 \cup B_2]. \end{aligned} \tag{3.13}$$

Putting everything together, the right-hand side of (3.9) reads

$$\begin{aligned} \tilde{m}[1, \dots, k | k + 1, \dots, k + \ell] &= \sum_{\pi \in NCP(k, \ell)} \prod_{B \in \pi} m_o[B] \\ &+ \sum_{\substack{\pi \in NCP(k) \times NCP(\ell), \\ U_1, U_2 \text{ marked}}} \prod_{B \in \pi \setminus \{U_1, U_2\}} m_o[B_1 \cup B_2] \\ &\times \left[\left(\prod_{s \in U} m_s \right) \sum_{\Gamma \in \mathcal{G}_c(U)} c_\Gamma q_\Gamma \right]. \end{aligned} \tag{3.14}$$

with $c_\Gamma \in \mathbb{Z}$ as in (3.12) and $U = U_1 \cup U_2$. The first term on the right-hand side of (3.14) is the sum of (3.10) and (3.13). The second term is obtained from the second term in (3.12) by noting that U is the only connecting block of the partition, i.e., any block of $\pi \setminus \{U\}$ only lives on one of the circles. Observing that this matches the structure in (1.26), we obtain (3.7) by comparing term-by-term.

We finally have all the necessary tools to prove Theorem 2.4.

3.2 Step 2: Conclusion

Proof of Theorem 2.4 Let \mathfrak{f} denote the right-hand side of (2.11). We recall that the initial condition (2.3) is immediate from Definition 3.2(ii) and that the symmetry in Definition 2.1(i) follows from the remark on Theorem 2.4 above. Hence, it only remains to check that \mathfrak{f} satisfies the recursion (2.4). To simplify notation, we interpret \mathfrak{f} as a function of the multi-indices α and β .

Similar to the proof of [6, Lem. 4.4], we use (1.17) and (3.7) to write out the first and second-order free cumulant functions in terms of suitable graphs. This yields

$$\frac{\mathfrak{f}[\alpha | \beta]}{m_1 \dots m_{k+\ell}} = \sum_{\pi \in NCP(k, \ell)} \left(\prod_{B \in K(\pi)} \left\langle \prod_{j \in B} A_j \right\rangle \right) \prod_{B \in \pi} \left(\sum_{\Gamma \in NCG_c(B_1 \cup B_2)} q_\Gamma \right)$$

$$\begin{aligned}
 &+ \sum_{\substack{\pi \in NCP(k) \times NCP(\ell) \\ U_1, U_2 \text{ marked}}} \left(\prod_{B \in K(\pi)} \left\langle \prod_{j \in B} A_j \right\rangle \right) \left(\sum_{\Gamma \in NCG_c(U_1, U_2)} c_\Gamma q_\Gamma \right) \\
 &\times \prod_{B \in \pi \setminus \{U_1, U_2\}} \left(\sum_{\Gamma \in NCG_c(B_1 \cup B_2)} q_\Gamma \right). \tag{3.15}
 \end{aligned}$$

To collect the terms involving the deterministic matrices A_1, \dots, A_k , we define

$$\mathcal{F}(\pi) := \prod_{B \in K(\pi')} \left\langle \prod_{j \in B} A_j \right\rangle \tag{3.16}$$

for any $\pi \in NCP(k, \ell)$ that has more than one connecting block. Here, $\pi' \in \overline{NCP}(k, \ell)$ denotes the unique permutation for which the blocks of π' coincide with π . Further, we define

$$\mathcal{F}(\pi) := \sum_{\substack{\pi' \in \overline{NCP}(k, \ell) \\ \text{blocks}(\pi') = \pi}} \left(\prod_{B \in K(\pi')} \left\langle \prod_{j \in B} A_j \right\rangle \right) + \prod_{B \in K(\pi'')} \left\langle \prod_{j \in B} A_j \right\rangle \tag{3.17}$$

for any $\pi \in NCP(k, \ell)$ that has exactly one connecting block U . Here, π'' is the marked partition obtained from π by splitting the connecting block into $U_1 = U \cap [k]$ and $U_2 = U \cap [k + 1, k + \ell]$, and marking U_1, U_2 on the respective circles.

Next, we decompose the sum over $\overline{NCP}(k, \ell)$ in (3.15) according to the number of connecting cycles in the permutation and rewrite the right-hand side as

$$\frac{f[\alpha|\beta]}{m_1 \cdots m_{k+\ell}} = \sum_{\pi \in NCP(k, \ell)} \mathcal{F}(\pi) \prod_{B \in \pi} \left(\sum_{\Gamma \in NCG_c(B)} q_\Gamma \right)$$

using (3.16) and (3.17). Recall that any connected component of $\Gamma \in \mathcal{G}(k, \ell)$ can either be identified with a disk non-crossing graph or itself satisfies Definition 3.2. We can thus interpret $NCG_c(B)$ as a connected component of a bigger graph and rewrite

$$\sum_{\pi \in NCP(k, \ell)} \mathcal{F}(\pi) \prod_{B \in \pi} \left(\sum_{\Gamma \in NCG_c(B)} q_\Gamma \right) = \sum_{\Gamma \in \mathcal{G}(k, \ell)} \mathcal{F}(\pi_\Gamma) q_\Gamma. \tag{3.18}$$

Here, π_Γ denotes the partition arising from the vertex sets of the connected components of Γ .

Using (3.3) from Definition 3.2, decompose the right-hand side of (3.18) as

$$\sum_{\Gamma \in \mathcal{G}(k, \ell)} \mathcal{F}(\pi_\Gamma) q_\Gamma = \sum_{\Gamma \in \mathcal{G}_{-(1, k)}} q_\Gamma (\mathcal{F}(\pi_\Gamma) + q_{1, k} \mathcal{F}(\pi_{\Gamma \cup \{(1, k)\}})). \tag{3.19}$$

Recall that $\Gamma \cup \{(1, k)\}$ denotes the graph obtained from adding an edge $(1, k)$ to Γ and that, therefore, $q_{\Gamma \cup \{(1, k)\}} = q_{1, k} q_\Gamma$. Next, apply (3.2) to split the right-hand side

of (3.19) into contributions corresponding to $\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3,$ and \mathcal{G}_4 . This yields

$$\begin{aligned} \sum_{\Gamma \in \mathcal{G}(k, \ell)} \mathcal{F}(\pi_\Gamma) q_\Gamma &= \sum_{\Gamma \in \mathcal{G}([2, k], [k+1, k+\ell])} q_\Gamma (\mathcal{F}(\pi_\Gamma) + q_{1, k} \mathcal{F}(\pi_{\Gamma \cup \{(1, k)\}})) \\ &+ \sum_{j=2}^k \sum_{\Gamma \in NCG_{(1, j)}([1, j] \times \mathcal{G}([j, k], [k+1, k+\ell]))} q_\Gamma (\mathcal{F}(\pi_\Gamma) \\ &+ q_{1, k} \mathcal{F}(\pi_{\Gamma \cup \{(1, k)\}})) \\ &+ \sum_{j=1}^{k-1} \sum_{\Gamma \in \mathcal{G}_{(1, j)}([1, j], [k+1, k+\ell]) \times NCG([j, k])} q_\Gamma (\mathcal{F}(\pi_\Gamma) \\ &+ q_{1, k} \mathcal{F}(\pi_{\Gamma \cup \{(1, k)\}})) \\ &+ \sum_{j=1}^\ell \sum_{\Gamma \in NCG_{(1, k+j)}(\{1, k, k+j, \dots, k+\ell, \dots, k+j\})} q_\Gamma (\mathcal{F}(\pi_\Gamma) \\ &+ q_{1, k} \mathcal{F}(\pi_{\Gamma \cup \{(1, k)\}})), \end{aligned} \tag{3.20}$$

where the edges prescribed in the definitions of $\mathcal{G}_2, \mathcal{G}_3,$ and \mathcal{G}_4 are added as a subscript to NCG and \mathcal{G} , respectively. It remains to compare (3.20) with (2.4).

Computation of the first line of (3.20): Recall that $\sum_{\Gamma \in \mathcal{G}_1} q_\Gamma = \sum_{\Gamma \in \mathcal{G}([2, k], [k+1, k+\ell])} q_\Gamma$. As any $\Gamma \in \mathcal{G}_1$ has an isolated vertex 1, the set $\{1\}$ appears as a singleton block of the underlying partition π_Γ . Let π'_Γ denote an annular non-crossing permutation with blocks given by π_Γ . Then $(\dots k 1 \dots) \in K(\pi'_\Gamma)$, i.e., 1 and k appear in the same cycle of the Kreweras complement. This gives

$$\prod_{B \in K(\pi'_\Gamma)} \left\langle \prod_{j \in B} A_j \right\rangle = \prod_{B \in K(\pi'_\Gamma)_{\{(1, k)\} \cup [k+1, k+\ell]}} \left\langle \prod_{j \in B} A'_j \right\rangle \tag{3.21}$$

where $A'_j = A_j$ for $j = 2, \dots, k - 1$ and $j \in [k + 1, k + \ell]$, but $A'_k = A_k A_1$. Considering $\mathcal{F}(\pi_{\Gamma \cup \{(1, k)\}})$, note that adding an edge $(1, k)$ to the graph Γ implies that $(k) \in K(\pi_{\Gamma \cup \{(1, k)\}})$, i.e., $\langle A_k \rangle$ always occurs as a separate factor in \mathcal{F} . Hence, whenever π_Γ has more than one connecting block we can use (3.16) to evaluate

$$\begin{aligned} &\mathcal{F}(\pi_\Gamma) + q_{1, k} \mathcal{F}(\pi_{\Gamma \cup \{(1, k)\}}) \\ &= \prod_{B \in K(\pi'_\Gamma)_{\{(1, k)\} \cup [k+1, k+\ell]}} \left\langle \prod_{j \in B} A'_j \right\rangle + q_{1, k} \langle A_k \rangle \prod_{B \in K(\pi'_\Gamma)_{\{(1, k)\} \cup [k+1, k+\ell]}} \left\langle \prod_{j \in B} A_j \right\rangle \end{aligned} \tag{3.22}$$

with A'_j as in (3.21). In the remaining cases, π_Γ has only one connecting block U , and \mathcal{F} is a sum of two terms. As the first term of (3.17) can be evaluated similarly to (3.22), we only consider the second term. Let π''_Γ denote the element of $NCP(k) \times NCP(\ell)$ in which the blocks $U_1 = U \cap [k]$ and $U_2 = U \cap [k + 1, k + \ell]$ are marked. Recall

that the marking does not influence the Kreweras complement, which is taken for both circles separately here. As 1 and k lie on the same circle, we can argue as in the permutation case. It follows that

$$\begin{aligned} & \left(\prod_{s=1}^{k+\ell} m_s \right) \sum_{\Gamma \in \mathcal{G}([2,k],[k+1,k+\ell])} q_\Gamma \left(\mathcal{F}(\pi_\Gamma) + q_{1,k} \mathcal{F}(\pi_{\Gamma \cup \{(1,k)\}}) \right) \\ &= m_1 (F[T_2, \dots, T_{k-1}, G_k A_k A_1 | T_{k+1}, \dots, T_{k+\ell}] \\ & \quad + q_{1,k} F[T_2, \dots, T_{k-1} G_k A_1 | T_{k+1}, \dots, T_{k+\ell}] \langle A_k \rangle). \end{aligned} \tag{3.23}$$

Computation of the second line of (3.20): For the contribution arising from \mathcal{G}_2 , recall that $\Gamma = \Gamma_1 \cup \Gamma_2$ with $\Gamma_1 \in NCG_{(1,j)}([1, j])$ and $\Gamma_2 \in \mathcal{G}([j, k], [k+1, k+\ell])$ implies $q_\Gamma = q_{\Gamma_1} q_{\Gamma_2}$, i.e., the weights factorize. Further, the term involving Γ_1 evaluates to

$$\sum_{\Gamma_1 \in NCG_{(1,j)}([1, j])} q_{\Gamma_1} = \frac{q_{1,j}}{1 + q_{1,j}} \sum_{\Gamma_1 \in NCG[1, j]} q_{\Gamma_1} = m_1 m_j \sum_{\Gamma_1 \in NCG[1, j]} q_{\Gamma_1} \tag{3.24}$$

as q_{Γ_1} always includes a factor $q_{1,j}$ if $\Gamma_1 \in NCG_{(1,j)}([1, j])$.

Let $\pi_\Gamma \in NCP(j) \times NCP(k - j + 1, \ell)$ denote the partition associated with a graph $\Gamma \in NCG_{(1,j)}([1, j]) \times \mathcal{G}([j, k], [k + 1, k + \ell])$ in the second term of (3.20) and let π'_Γ is an annular non-crossing permutation with blocks given by π_Γ . Since the edge $(1, j)$ must occur in Γ , the vertices 1 and j must be associated with same cycle of π'_Γ . Hence, the elements in $[1, j]$ and $[j, k]$ are in different cycles of $K(\pi'_\Gamma)$. Moreover, none of $1, \dots, j - 1$ can be part of a connecting cycle in $K(\pi'_\Gamma)$. This implies the decomposition

$$\prod_{B \in K(\pi'_\Gamma)} \left\langle \prod_{j \in B} A_j \right\rangle = \left(\prod_{B \in K(\pi'_\Gamma|_{[1, j]})} \left\langle \prod_{i \in B \setminus \{j\}} A_i \right\rangle \right) \left(\prod_{B \in K(\pi'_\Gamma|_{[j, k] \cup [k+1, k+\ell]})} \left\langle \prod_{i \in B} A_i \right\rangle \right).$$

Again, the only difference between $\mathcal{F}(\pi_\Gamma)$ and $\mathcal{F}(\pi_{\Gamma \cup \{(1,k)\}})$ is the fact that $\langle A_k \rangle$ must occur as a separate factor in the second case. Thus, we can argue as in (3.22) and evaluate

$$\begin{aligned} & \mathcal{F}(\pi_\Gamma) + q_{1,k} \mathcal{F}(\pi_{\Gamma \cup \{(1,k)\}}) \\ &= \left(\prod_{B \in K(\pi_\Gamma|_{[1, j]})} \left\langle \prod_{i \in B \setminus \{j\}} A_i \right\rangle \right) \left(\prod_{B \in K(\pi'_\Gamma|_{[j, k] \cup [k+1, k+\ell]})} \left\langle \prod_{i \in B} A_i \right\rangle \right) \\ & \quad + q_{1,k} \langle A_k \rangle \left(\prod_{B \in K(\pi_\Gamma|_{[1, j]})} \left\langle \prod_{i \in B \setminus \{j\}} A_i \right\rangle \right) \\ & \quad \times \left(\prod_{B \in K(\pi'_\Gamma|_{[j, k] \cup [k+1, k+\ell]})} \left\langle \prod_{i \in B \setminus \{k\}} A_i \right\rangle \right) \end{aligned} \tag{3.25}$$

whenever π_Γ has more than one connecting block and (3.16) applies.

In the remaining cases, $\mathcal{F}(\pi_\Gamma)$ is evaluated using (3.17). Here, the first term can be treated similarly to (3.25), leaving only the contribution of the marked partition π''_Γ . Recalling that $[1, j]$ and $[j, k]$ lie on the same circle and that $K(\pi''_\Gamma)$ is evaluated circle-wise, we can argue as in the permutation case. In particular,

$$\begin{aligned} \mathcal{F}(\pi_\Gamma) &= \sum_{\substack{\pi'_\Gamma \in \overrightarrow{NCP}(k, \ell) \\ \text{blocks}(\pi'_\Gamma) = \pi_\Gamma}} \left(\prod_{B \in K(\pi'_\Gamma|_{[1, j]})} \left\langle \prod_{i \in B \setminus \{j\}} A_i \right\rangle \right) \left(\prod_{B \in K(\pi'_\Gamma|_{[j, k] \cup [k+1, k+\ell]})} \left\langle \prod_{i \in B} A_i \right\rangle \right) \\ &+ \left(\prod_{B \in K(\pi''_\Gamma|_{[1, j]})} \left\langle \prod_{i \in B \setminus \{j\}} A_i \right\rangle \right) \left(\prod_{B \in K(\pi''_\Gamma|_{[j, k] \cup [k+1, k+\ell]})} \left\langle \prod_{i \in B} A_i \right\rangle \right) \\ &= \left(\prod_{B \in K(\pi_\Gamma|_{[1, j]})} \left\langle \prod_{i \in B \setminus \{j\}} A_i \right\rangle \right) \mathcal{F}(\pi_{\Gamma_2}) \end{aligned}$$

i.e., \mathcal{F} factorizes similar to (3.25). Here, Γ_2 denotes the subgraph of $\Gamma \in \mathcal{G}_2$ that lies in $\mathcal{G}([j, k], [k + 1, k + \ell])$. Using (1.22) and (1.17), we evaluate

$$\begin{aligned} &\left(\prod_{s=1}^{k+\ell} m_s \right) \sum_{\Gamma \in \text{NCG}_{(1, j)}([1, j]) \times \mathcal{G}([j, k], [k+1, k+\ell])} q_\Gamma \left(\mathcal{F}(\pi_\Gamma) + q_{1, k} \mathcal{F}(\pi_{\Gamma \cup \{(1, k)\}}) \right) \\ &= m_1 m_1 [T_1, \dots, T_{j-1}, G_j] \left(F[T_j, \dots, T_k | T_{k+1}, \dots, T_{k+\ell}] \right. \\ &\quad \left. + q_{1, k} F[T_j, \dots, T_{k-1}, G_k | T_{k+1}, \dots, T_{k+\ell}] \langle A_k \rangle \right). \end{aligned} \tag{3.26}$$

Computation of the third line of (3.20): The contribution from \mathcal{G}_3 can be treated similarly to the second line of (3.20).

Computation of the fourth line of (3.20): For the term that arises from \mathcal{G}_4 , recall that τ only influences the geometry of the graph, but not its edge set. Hence, the summation reduces to the underlying disk non-crossing graphs and we can evaluate it using (1.22) and (1.17). As the only difference between $\mathcal{F}(\pi_\Gamma)$ and $\mathcal{F}(\pi_{\Gamma \cup \{(1, k)\}})$ is again the fact that $\langle A_k \rangle$ must occur as a separate factor in the second case, we obtain

$$\begin{aligned} &\left(\prod_{s=1}^{k+\ell} m_s \right) \sum_{\Gamma \in \text{NCG}_{(1, k+j)}(\{1, k, k+j, \dots, k+\ell, \dots, k+j\})} q_\Gamma \left(\mathcal{F}(\pi_\Gamma) + q_{1, k} \mathcal{F}(\pi_{\Gamma \cup \{(1, k)\}}) \right) \\ &= m_1 (m_1 [T_1, \dots, T_k, T_{k+j}, \dots, T_{k+j-1}, G_{k+j}] \\ &\quad + q_{1, k} m_1 [T_1, \dots, T_{k-1}, G_k, T_{k+j}, \dots, T_{k+j-1}, G_{k+j}] \langle A_k \rangle). \end{aligned} \tag{3.27}$$

Similar to (3.24), Γ containing an edge $(1, k + j)$ ensures that the contribution has the prefactor m_1 .

Putting (3.23), (3.26), and (3.27) together, we see that (3.20) is equivalent to (2.4) with $f[\cdot]$ in place of $m_2[\cdot]$. We conclude that $f[\cdot]$ satisfies the same symmetry

and initial condition as $m_2[\cdot|\cdot]$, as well as the same recursion. Recall that these three properties uniquely identify $f[\alpha|\beta]$ and $m_2[\alpha|\beta]$ for any multi-indices α, β . It now readily follows by induction that $f[\alpha|\beta]$ and $m_2[\alpha|\beta]$ indeed coincide for all α, β , i.e., that $f[\cdot|\cdot] = m_2[\cdot|\cdot]$, as claimed.

4 Proof of the Formulas for $m_\kappa[\cdot|\cdot]$, $m_\sigma[\cdot|\cdot]$, and $m_\omega[\cdot|\cdot]$

4.1 Proof of Theorem 2.6

We use proof by induction over the length of the multi-indices and the recursion for $m_\kappa[\cdot|\cdot]$ from (2.4). First, recall that by definition of $NCP(k)$, $NCP(\ell)$, and $\overline{NCP}(k, \ell)$ both sums on the right-hand side of (2.12) are empty whenever either α or β is empty. Observing that Definition 2.1(i) together with Theorem 2.4 implies that $m_\kappa[S_1|\emptyset] = m_\kappa[\emptyset|S_2] = 0$, the base case is established. As the formula on the right-hand side of (2.12) is further symmetric under the interchanging α and β (cf. Remark Theorem 2.4), it is sufficient to only carry out the induction step for one of the arguments in $m_\kappa[\cdot|\cdot]$.

Fix $k, \ell \in \mathbb{N}$ and assume that (2.12) holds for multi-indices α of length $1, \dots, k - 1$ and β of length ℓ . Recalling that $\kappa_4 m_\kappa[\cdot|\cdot]$ satisfies (2.4) with s_κ in (2.6) as the only source term, we can use the recursion to express $m_\kappa[\alpha|\beta]$ in terms of $m_1[\cdot]$ and $m_\kappa[\alpha'|\beta]$ with multi-indices α' of length $1, \dots, k - 1$. The claim thus follows by applying the induction hypothesis and showing that the expression obtained from the recursion can be rewritten to match the structure on the right-hand side of (2.12). We start by considering the summands on the right-hand side of (2.4) separately and then check that their sum, i.e., $m_\kappa[\cdot|\cdot]$, is of the same form. To facilitate keeping track of the individual contributions, we start with the terms on the right-hand side of (2.4) that do not contain the prefactor $q_{1,k}$ and abbreviate

$$\begin{aligned}
 K_1 &:= m_1 m_\kappa[T_2, \dots, T_{k-1}, G_k A_k A_1 | T_{k+1}, \dots, T_{k+\ell}] \\
 K_2^{(j)} &:= m_1 m_\kappa[T_1, \dots, T_{j-1}, G_j | T_{k+1}, \dots, T_{k+\ell}] m_1[T_j, \dots, T_k], \quad j \in [k - 1] \\
 K_3^{(j)} &:= m_1 m_1[T_1, \dots, T_{j-1}, G_j] m_\kappa[T_j, \dots, T_k | T_{k+1}, \dots, T_{k+\ell}], \quad j \in [2, k] \\
 K_4^{(r,s,t)} &:= m_1 \langle M_{[r]} \odot M_{[s,t]} \rangle \langle (M_{[r,k]} A_k) \odot M_{(t,\dots,k+\ell,k+1,\dots,s)} \rangle, \\
 &\quad r \in [k], \quad k + 1 \leq s \leq t \leq k + \ell.
 \end{aligned}$$

As the argument is similar, we fix j resp. r, s, t and omit the superscripts for the following discussion.

Structure of K_1 : By the induction hypothesis, $m_\kappa[T_2, \dots, T_{k-1}, G_k A_k A_1 | T_{k+1}, \dots, T_{k+\ell}]$ factorizes into expressions involving only deterministic matrices or spectral parameters, respectively, and further has the structure specified on the right-hand side of (2.12) in terms of $z_2, \dots, z_{k+\ell}$ and the matrices $A_2, \dots, A_{k-1}, A_k A_1, A_{k+1}, \dots, A_{k+\ell}$. As a consequence, the matrices A_k and A_1 always occur together in the matrix products. On the level of the indices of the deterministic matrices, we may reinterpret this as an element $\pi \in \overline{NCP}(k, \ell)$ with a cycle $(\dots k 1 \dots)$ or an element of

$\pi \in NCP(k) \times NCP(\ell)$ with a block $\{\dots, k, 1, \dots\}$ depending on the rest of the underlying structure. The treatment of the two cases is identical and we consider them in parallel. As the indices 1 and k always occur together in $K(\pi)$, the index 1 must occur separated in $K^{-1}(\pi)$, either as a fixed point (1) or a singleton set $\{1\}$, to match the structure in (2.12). Note that the spectral parameter z_1 only appears in the prefactor m_1 . Hence, setting the functions $\psi_{K^{-1}(\pi), (1)}$ resp. $\psi_{K^{-1}(\pi), \{1\}}$ equal to $m(z_1)$ yields the missing contribution. It follows that K_1 matches the structure on the right-hand side of (2.12). Note that all ψ_i associated with permutations without a fixed point (1) and partitions without a singleton block $\{1\}$, respectively, are equal to zero for K_1 .

Structure of K_2 : We apply the induction hypothesis for $m_k[T_1, \dots, T_{j-1}, G_j | T_{k+1}, \dots, T_{k+\ell}]$ as well as (1.22) for $m_1[T_j, \dots, T_k]$ to rewrite K_2 as a sum of terms that naturally factorize into expressions involving only $z_1, \dots, z_{k+\ell}$ or $A_1, \dots, A_{k+\ell}$, respectively. It remains to check for the structure on the right-hand side of (2.12), i.e., that each summand can be associated with an annular non-crossing permutation π or a marked element $\pi \in NCP(k) \times NCP(\ell)$ such that the terms involving spectral parameters factorize according to cycles resp. blocks of π and the contribution of the deterministic matrices factorizes according to the cycles resp. blocks in the Kreweras complement $K(\pi)$. As treatment of the cases $\pi \in \overline{NCP}(k, \ell)$ and $\pi \in NCP(k) \times NCP(\ell)$ is identical, we consider them in parallel.

Observe that the induction hypothesis and (1.22) already prescribe the desired complement structure for the indices of the spectral parameters and deterministic matrices occurring in $m_1 m_k [T_1, \dots, T_{j-1}, G_j | T_{k+1}, \dots, T_{k+\ell}]$ and $m_1 [T_j, \dots, T_k]$, respectively. As we may visualize the elements of $NCP(j)$ with their Kreweras complement on an interval by cutting the boundary of the labeled disk (cf. Fig. 1), we can draw both factors of K_2 onto the same annulus. The result is visualized on the left of Fig. 18. For simplicity, we omit most of the intermediate labels and only add some of the matrices associated with the vertices on the midpoints of the arcs between the labels in red.

Note that the interval is placed such that the orientation inherited from the disk aligns with the orientation of the underlying annulus. Moreover, Fig. 18 matches the picture of a (k, ℓ) -annulus up to the label j occurring twice and the label on the midpoint of the arch connecting the two copies of j that is associated with the identity matrix. Since this identity matrix does not influence the value of K_2 , we may remove the label corresponding to Id from the picture in Fig. 18. This leaves $k + \ell$ labels at the midpoints of arches along the annulus, one associated with each matrix $A_1, \dots, A_{k+\ell}$. On the level of the indices of the deterministic matrices, each term in K_2 can thus be identified with an element $\pi \in \overline{NCP}(k, \ell)$ resp. an element of $\pi \in NCP(k) \times NCP(\ell)$. As the two labels j now occur next to each other, we merge them to obtain the (k, ℓ) -annulus as the structure underlying the indices of the spectral parameters. The result is visualized on the right of Fig. 18. Note that the labels now match Definition 1.19 exactly.

It is readily checked that the cycle resp. block structure obtained from merging the two labels j matches $K^{-1}(\pi)$ and that any cycles resp. blocks that were connected in this step can be interpreted as a cycle of an element in $\overline{NCP}(k, \ell)$ resp. a block of an element of $NCP(k) \times NCP(\ell)$. In particular, the contribution of the spectral parameters factorizes as claimed on the right-hand side of (2.12). Comparing the

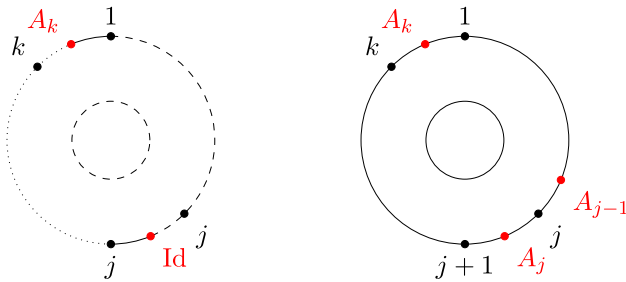


Fig. 18 Visualization of the indices in $m_1 m_k [T_1, \dots, T_{j-1}, G_j | T_{k+1}, \dots, T_{k+\ell}]$ (dashed) and in $m_1 [T_j, \dots, T_k]$ (dotted) before the rewriting (left) and the indices after merging the two labels j (right). (Color figure online)

permutations resp. partitions contributing to each $K_2^{(j)}$ and noting that the right-hand side of (2.12) is linear in each ψ_i , the sum $\sum_j K_2^{(j)}$ also has the desired structure.

Structure of K_3 : The treatment of K_3 is analogous to that of K_2 .

Structure of K_4 : From the formula for $M_{(\dots)}$ in (1.18), it follows that K_4 can be written as a sum of terms that naturally factorize into expressions involving only $z_1, \dots, z_{k+\ell}$ or $A_1, \dots, A_{k+\ell}$, respectively. In particular, the part that involves deterministic matrices always consists of two factors of the form $\langle (\prod_{j \in I_1} A_j) \odot (\prod_{j \in I_2} A_j) \rangle$ with index sets $I_1 \subseteq [k]$ and $I_2 \subseteq [k+1, \dots, k+\ell]$ due to the Hadamard product in K_4 . It remains to check for the structure on the right-hand side of (2.12).

Using the same trick as for K_2 and K_3 , we can visualize the terms involved in K_4 on the same annulus. The result is sketched in Fig. 19. Again, we omit most of the labels and add the matrices associated with the vertices on the midpoints of the arcs in red. To avoid overcrowding the labels in the interior of the inner circle, we further visualize the two circles separately.

Observe that whenever a label occurs twice, both copies are placed next to one another and the label at the midpoint of the arch connecting them is always associated with the identity matrix. As the identity matrices do not contribute to K_4 , we may argue as before and remove the corresponding labels from the picture in Fig. 19. Note that this leaves $k + \ell$ labels at the midpoints of arches along the annulus, one

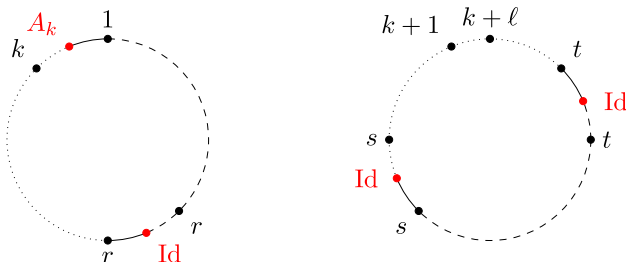


Fig. 19 Visualization of the indices in $m_1 \langle M_{[r]} \odot M_{[s,t]} \rangle$ (dashed) and $\langle (M_{[r,k]} A_k) \odot M_{(t, \dots, k+\ell, k+1, \dots, s)} \rangle$ (dotted) on the outer (left) and inner circle (right) of the (k, ℓ) -annulus (pictured separately). (Color figure online)

associated with each matrix $A_1, \dots, A_{k+\ell}$, such that we can reinterpret the structure on the level of the indices of the deterministic matrices as an element $\pi \in \overrightarrow{NCP}(k, \ell)$ resp. $\pi \in NCP(k) \times NCP(\ell)$. Since the two copies of r, s , and t are now placed right next to their counterpart, we merge them to obtain a (k, ℓ) -annulus as the structure underlying the indices of the spectral parameters. It is readily checked that the cycle resp. block structure obtained from this step matches $K^{-1}(\pi)$ and that any cycles resp. blocks that were connected by merging two identical labels can be interpreted as a cycle of an element in $\overrightarrow{NCP}(k, \ell)$ resp. a block of an element of $NCP(k) \times NCP(\ell)$. In particular, whenever the merging of the labels creates two new cycles, the cycles can be drawn onto the (k, ℓ) -annulus without crossing. It follows that the contribution of the spectral parameters factorizes as claimed on the right-hand side of (2.12). Comparing the permutations resp. partitions contributing to each $K_4^{(r,s,t)}$ and noting that the right-hand side of (2.12) is linear in each ψ_i , summing over r, s, t preserves the structure of the term.

The remaining terms with prefactor $q_{1,k}\langle A_k \rangle$ can be rewritten similarly. As the factor $\langle A_k \rangle$ always occurs separately, the structure underlying the indices of the deterministic matrices is an element $\pi \in \overrightarrow{NCP}(k, \ell)$ resp. $\pi \in NCP(k) \times NCP(\ell)$ with a fixed point (k) resp. a singleton set $\{k\}$. Hence, K^{-1} must contain a cycle $(\dots k 1 \dots)$ resp. a block $\{\dots, k, 1, \dots\}$ in which 1 and k occur together. Recalling that $q_{1,k} = \frac{m_1 m_k}{1 - m_1 m_k}$, it is ensured that the part of the term depending on z_1, \dots, z_k contains a factor $\psi_i(\dots, z_k, z_1, \dots)$. With this modification, we can use the same argument as for K_1, \dots, K_4 to conclude that all terms contributing to the recursion match the structure on the right-hand side of (2.12). Comparing the permutations resp. partitions contributing to each term and recalling that the right-hand side of (2.12) is linear in each ψ_i , it follows that the same holds for their sum, i.e., for $m_\kappa[\alpha|\beta]$. This concludes the induction step. □

4.2 Proof of Theorems 2.7 and 2.8

The proofs of Theorems 2.7 and 2.8 are similar to those of Theorems 2.4 and 2.6, respectively. We, therefore, mainly focus on the necessary modifications below.

Proof of Theorem 2.7 The overall argument is analogous to the proof of Theorem 2.4 in Sects. 3.1 and 3.2 up to the definition of the graphs appearing in the combinatorial formula for $\tilde{m}_\sigma[\cdot|\cdot]$. Recalling that the structure of the recursion for $\tilde{m}_\sigma[\cdot|\cdot]$ is similar to the one for $\tilde{m}_{GUE}[\cdot|\cdot]$, the multi-set of graphs can be constructed as described in Definition 3.2. However, since the source term involves $m^{\#,\sigma}[\cdot]$, the resulting multi-set of graphs will carry the same kind of vertex coloring. By replacing \mathcal{G}_4 in Definition 3.2 by

$$\mathcal{G}_4^\sigma := \bigcup_{j=1}^{\ell} \tau \left(\left\{ \Gamma \in NCG^\#(\{1, \dots, k, k+j, k+j-1, \dots, k+1, k+\ell, k+j\}) \mid \Gamma \text{ has edge } (1, k+j) \right\} \right).$$

where $\# = (0, \dots, 0, 1, \dots, 1)$ with k zeros and $k + \ell + 1$ ones, we obtain a family $\mathcal{G}^\sigma(k, \ell)$ of graphs with the desired properties. Recall that map τ was introduced in Definition 3.1. An example in the current setting is given in Fig. 20. Note the different arrangement of the indices on the left compared to Fig. 9, which results from the structure of the source term in the recursion for $\tilde{m}_\sigma[\cdot|\cdot]$. Further, τ does not influence the coloring of the vertices.

Similar to the proof of Lemma 3.3, we obtain

$$\tilde{m}_\sigma[1, \dots, k|k + 1, \dots, k + \ell] = \left(\prod_{s=1}^{k+\ell} m_s \right) \sum_{\Gamma \in \mathcal{G}^\sigma(k, \ell)} \prod_{(i, j) \in E(\Gamma)} q_{i, j}^\#$$

where $q_{i, j}^\#$ is as in Lemma A.2, i.e., $q_{i, j}^\# = q_{i, j} = \frac{m_i m_j}{1 - m_i m_j}$ whenever the edge (i, j) connects two vertices of the same color and $q_{i, j}^\# = \frac{\sigma m_i m_j}{1 - \sigma m_i m_j}$ otherwise. From here, the remaining steps are carried out as in the proof of Theorem 2.4.

Proof of Theorem 2.8 We use again proof by induction. As the recursions for $m_\omega[\cdot|\cdot]$ and $m_\kappa[\cdot|\cdot]$ are the same up to the source term, it only remains to show that

$$K_5^{(j)} := \langle (M_{[k]} A_k) \odot M_{(k+j, \dots, k+\ell, k+1, \dots, k+j)} \rangle, \quad j \in [\ell]$$

is of the form (2.14) and that summing up the contributions on the right-hand side of the recursion (2.4) does not break the structure. We start by noting that each $K_5^{(j)}$ contains the matrices $A_1, \dots, A_{k+\ell}$ exactly once and that the indices involved in the first and second factor of the Hadamard product match the indices on the outer and inner circle of the (k, ℓ) -annulus, respectively. The desired structure now follows from (1.22) and the fact that any annular non-crossing permutation of the (k, ℓ) -annulus can be decomposed uniquely into a composition of two permutations that only act on the inner and outer circle, respectively. Note that the index $k + j$ occurring twice in $M_{(k+j, \dots, k+\ell, k+1, \dots, k+j)}$ does not influence the structure of the Kreweras complement. When visualizing the term on a labeled disk, the two vertices labeled $k + j$ lie next to each other and the vertex on the midpoint of the arch connecting the two copies is associated with the identity matrix, i.e., its effect is not visible when the matrix product in $M_{(k+j, \dots, k+\ell, k+1, \dots, k+j)}$ is evaluated.

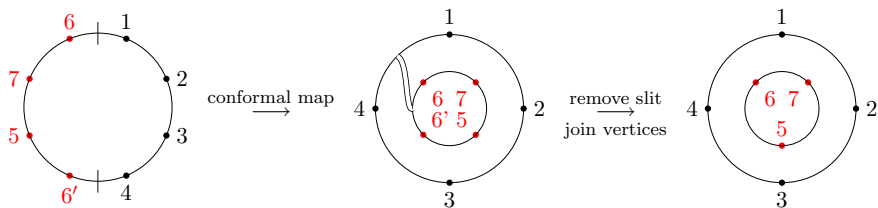


Fig. 20 The geometry of the transformation τ for $k = 4, \ell = 3$, and $j = 2$. (Color figure online)

A Proofs for the CLT in the Resolvent Case

A.1 Proof of Theorem 1.10 (Global Law with Transposes)

The proof of Theorem 1.10 is, modulo careful bookkeeping of the transposes, similar to the proof of the averaged local law in [6, Thm. 3.4]. In particular, an explicit formula for $m^{\#, \sigma}[\cdot]$ and the associated free cumulants is obtained along the way. For the convenience of the reader, we give a brief overview of the necessary changes. As there is nothing to prove if $\sigma = 1$ and the case $\sigma = -1$ needs an additional argument, consider $\sigma \in (-1, 1)$ first. We start by introducing a renormalization that captures the general second moment structure of W more precisely (cf. [4] and [7]). Let

$$\underline{Wf(W)} := Wf(W) - \tilde{\mathbb{E}}\tilde{W}(\partial_{\tilde{W}}f)(W) \tag{A.1}$$

with \tilde{W} an independent copy of W . This yields, e.g.,

$$\underline{WG_1} = WG_1 + \langle G_1 \rangle G_1 + \frac{\sigma}{N} G_1^t G_1 + \frac{\tilde{\omega}_2}{N} \text{diag}(G_1) G_1, \tag{A.2}$$

$$\begin{aligned} \underline{WT_1 \dots T_k} &= \underline{WG_1} A_1 T_{[2,k]} + \sum_{j=2}^k (\langle T_{[1,j]} G_j \rangle T_{[j,k]} + \frac{\sigma}{N} (T_{[1,j]} G_j)^t T_{[j,k]}) \\ &\quad + \frac{\tilde{\omega}_2}{N} \text{diag}(T_{[1,j]} G_j) T_{[j,k]}. \end{aligned} \tag{A.3}$$

Note that a simpler renormalization can be obtained by choosing \tilde{W} to be an independent GUE matrix instead of an independent copy of W . For the formulas (A.2) and (A.3) above, the difference between the two renormalizations is negligible. However, it becomes significant if the matrix product involves at least one transpose of a resolvent (cf. [4, Rem. 4.3]). As an example, consider the normalized trace of

$$\underline{WG_1^t} = WG_1^t + \frac{1}{N} G_1 G_1^t + \sigma \langle G_1 \rangle G_1^t + \frac{\tilde{\omega}_2}{N} \text{diag}(G_1) G_1^t$$

where $\sigma \langle G_1 \rangle^2 \sim 1$.

To obtain (1.23), we rewrite $G_1^\sharp A_1 \dots G_k^\sharp A_k$ in terms of $\underline{WG_1^\sharp A_1 \dots G_k^\sharp A_k}$ and estimate the resulting terms. We start by considering the case $A_1 = \dots = A_k = \text{Id}$. For $k = 2$, we obtain, e.g.,

$$\underline{WG_1 G_2^t} = \underline{WG_1} G_2^t + \frac{1}{N} (G_1 G_2^t)^t G_2^t + \sigma \langle G_1 G_2^t \rangle G_2^t + \frac{\tilde{\omega}_2}{N} \text{diag}(G_1 G_2^t) G_2^t$$

which yields

$$\begin{aligned} \left(1 - \sigma m_1 m_2 + \mathcal{O}_{\prec} \left(\frac{1}{N}\right)\right) \langle G_1 G_2^t \rangle &= m_1 m_2 - m_1 \underline{WG_1 G_2^t} + \frac{m_1}{N} (\sigma \langle G_1^t G_1 G_2^t \rangle \\ &\quad + \tilde{\omega}_2 \langle \text{diag}(G_1) G_1 G_2^t \rangle) \end{aligned}$$

$$+ \text{diag}(G_1 G_2^t G_2^t) + \mathcal{O}_<\left(\frac{1}{N}\right)$$

by (A.2), the resolvent identity $WG - zG = \text{Id}$, (1.11), and the local law for $\langle G_j \rangle$. Noting that $|\langle W G_1^t G_2 \rangle| = \mathcal{O}_<(N^{-1})$ following the bounds in [4, Thm. 4.1] and that the term with prefactor N^{-1} is also of lower order (cf. [7, Eq. (69)–(70)]), we obtain (1.23) with

$$m_1[G_1, G_2^t] = m^{(0,1),\sigma}[1, 2] = \frac{m_1 m_2}{1 - \sigma m_1 m_2}.$$

Note that the only effect of the inclusion of transposes lies in the emergence of the stability factor $1 - \sigma m_1 m_2$. However, as $|1 - \sigma m_1 m_2| > 1 - |\sigma|$ for $|\sigma| < 1$, the term is bounded from below and does not change the outcome of the estimates. The analog for general $k \geq 2$ follows by induction over the number of resolvents. Similar to [6, Thm. 3.4], the above argument allows us to extract a recursion that is satisfied by the deterministic approximation of $\langle G_1^\# \dots G_k^\# \rangle$ up to an $\mathcal{O}_<(N^{-1})$ error. By defining $m^{\#, \sigma}[\cdot]$ to satisfy this recursion exactly, i.e., defining it by (1.24), yields the desired statement. The recursion for $m_1[G_1^\# A_1, \dots, G_k^\# A_k]$ is obtained analogously.

The explicit formula (1.23) now follows by solving this recursion. As the treatment of the deterministic matrices is completely analogous to [6, Thm. 3.4], we restrict the discussion to the case $A_1 = \dots = A_k = \text{Id}$, i.e., to (1.24). Here, we obtain that $m^{\#, \sigma}[\cdot]$ and the associated free cumulants have a representation in terms of non-crossing graphs that mirror (1.16) and (1.17).

Definition A.1 (Bicolored NCG) Let $k \in \mathbb{N}$ and denote by $\#$ a binary vector of length k . For every $\Gamma \in NCG(k)$, color the vertex $j \in \{1, \dots, k\}$ red if the j th entry of $\#$ is 1, otherwise color it black. We call the resulting set of graphs *bicolored (disk) non-crossing graphs* on $\{1, \dots, k\}$ and denote it by $NCG^\#(k)$.

Lemma A.2 Let $k \in \mathbb{N}$ and fix a binary vector $\#$ of length k . Then

$$m^{\#, \sigma}[1, \dots, k] = \left(\prod_{s=1}^k m_s\right) \sum_{\Gamma \in NCG^\#(k)} q_\Gamma^\# \tag{A.4}$$

$$m_\circ^{\#, \sigma}[1, \dots, k] = \left(\prod_{s=1}^k m_s\right) \sum_{\Gamma \in NCG_c^\#(k)} q_\Gamma^\# \tag{A.5}$$

where $NCG_c^\#(k)$ denotes the connected graphs in $NCG^\#(k)$ and $q_\Gamma := \prod_{(i,j) \in E(\Gamma)} q_{i,j}^\#$ with $E(\Gamma)$ denoting the edge set of Γ . The edge weights $q_{i,j}^\#$ are such that $q_{i,j}^\# := q_{i,j} = \frac{m_i m_j}{1 - m_i m_j}$ whenever the edge (i, j) connects two vertices of the same color and $q_{i,j}^\# := \frac{\sigma m_i m_j}{1 - \sigma m_i m_j}$ whenever (i, j) connects a red vertex to a black one.

The proof is analogous to [6, Lem. 5.2] using the recursive structure of $NCG(k)$. Note that the modified edge weights as well as the factors $q_{1,k}^\sharp$ and $c_{1,j}$ in the recursion account for the stability factor $1 - \sigma m_i m_j$ arising whenever the product $G_1^\sharp \dots G_k^\sharp$ involves both resolvents and their transposes. We illustrate Lemma A.2 with an example.

Example A.3 Let $k = 3$ and pick spectral parameters $z_1, z_2, z_3 \in \mathbb{C}$ with $|\Im z_j| \gtrsim 1$ as well as $A_1 = A_2 = A_3 = \text{Id}$. By Theorem 1.10, the deterministic approximation of $\langle G_1 G_2^t G_3^t \rangle$ is given by $m^{\#, \sigma}[1, 2, 3]$ with $\# = (0, 1, 1)$. We visualize the elements of the set $NCG^\#(3)$ in Fig. 21. For a better overview, the edges are drawn as solid or dashed according to their contribution to q_Γ^\sharp . We use this sketch to compute

$$m^{\#, \sigma}[1, 2, 3] = m_1 m_2 m_3 \sum_{\Gamma \in NCG^\#} = \frac{m_1 m_2 m_3}{(1 - \sigma m_1 m_2)(1 - \sigma m_1 m_3)(1 - m_2 m_3)}.$$

Note that we thus reobtain [7, Eq. (49)] on macroscopic scales if $z_2 = z_3$.

It remains to consider the case $\sigma = -1$. Here, the Wigner matrix is of the form $W = D + iS$ with a diagonal matrix D and a skew-symmetric matrix S . Note that whenever the diagonal part is equal to zero, the resolvent $R(z_j) = (iS - z_j)^{-1}$ satisfies $R(z_j)^t = -R(-z_j)$. This allows treating the proof of (1.23) for $\langle R(z_1)^\sharp \dots R(z_k)^\sharp \rangle$ with $R(z_j)^\sharp \in \{R(z_j), R(z_j)^t\}$ analogous to the case $\langle G_1 \dots G_k \rangle$. Recall that this requires in particular a local law for $\langle R(z_j) \rangle$, which holds even if the Wigner matrix has zero diagonal. The general case follows from the bound

$$\langle G_1^\sharp \dots G_k^\sharp \rangle = \langle R(z_1)^\sharp \dots R(z_k)^\sharp \rangle + \mathcal{O}_< \left(\frac{1}{N} \right) \tag{A.6}$$

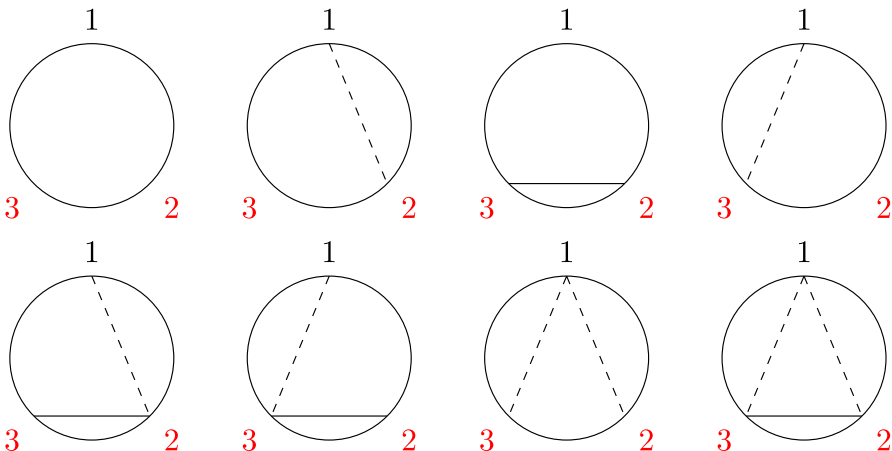


Fig. 21 The elements of $NCG^\#(3)$ for $\# = (0, 1, 1)$. The solid edges contribute a factor of $\frac{m_i m_j}{1 - m_i m_j}$ to q_Γ^\sharp while dashed edges contribute a factor $\frac{\sigma m_i m_j}{1 - \sigma m_i m_j}$. (Color figure online)

which reduces the proof of (1.23) to the previously considered case $D = 0$. The proof of (A.6) is given for $\langle G_1 G_2^t \rangle$ in [7, App. B]. By expanding $G_3^\sharp, \dots, G_k^\sharp$ analogously, the argument readily extends to $k \geq 3$ resolvents. \square

A.2 Proof of Theorem 2.3 (CLT for Resolvents)

We follow the general outline of the proof of [27, Thm. 3.6] to identify the necessary modifications for the general model described by Assumption 1.1. As the formulas in the general case are derived analogously, we assume w.l.o.g. $\langle A_k \rangle = 0$ throughout to keep the following equations short. The first step is the analog of [27, Lem. 2.3] that characterizes the difference between $\langle T_{[1,k]} \rangle - m_1[T_1, \dots, T_k]$ and $\langle T_{[1,k]} \rangle - \mathbb{E}\langle T_{[1,k]} \rangle$, thus connecting the statistics X_α back to the local law (1.19).

Lemma A.4 *Let $k \in \mathbb{N}$ and fix spectral parameters z_1, \dots, z_n with $|\Im z_j| \gtrsim 1$ and $\max_j |z_j| \leq N^{100}$ as well as bounded deterministic matrices A_1, \dots, A_k such that $\langle A_k \rangle = 0$. Then,*

$$\mathbb{E}\langle T_1 \dots T_k \rangle = m_1[T_1, \dots, T_k] + \frac{1}{N} \mathcal{E}[T_1, \dots, T_k] + \mathcal{O}\left(\frac{N^\varepsilon}{N^{3/2}}\right) \tag{A.7}$$

with $m_1[\cdot]$ as in (1.21) and a set function $\mathcal{E}[\cdot]$ that satisfies $\mathcal{E}[\emptyset] = 0$ as well as the recursion

$$\begin{aligned} \mathcal{E}[T_1, \dots, T_k] &= m_1(\mathcal{E}[T_2, \dots, T_{j-1}, G_k A_k A_1] + \sum_{j=1}^{k-1} \mathcal{E}[T_1, \dots, T_{j-1}, G_j] m_1[T_j, \dots, T_k]) \\ &\quad + \sum_{j=2}^k m_1[T_1, \dots, T_{j-1}, G_j] \mathcal{E}[T_j, \dots, T_k] + \frac{\tilde{\omega}_2}{N} \sum_{j=1}^k \langle M_{[j]} \odot (M_{[j,k]} A_k) \rangle \\ &\quad + \frac{\sigma}{N} \sum_{j=1}^k m_1[G_j^t A_{j-1}^t, \dots, G_2^t A_1^t, G_1^t, T_j, \dots, T_k] \\ &\quad + \kappa_4 \sum_{1 \leq r \leq s \leq t \leq k} \langle M_{[r]} \odot M_{[s,t]} \rangle \langle M_{[r,s]} \odot (M_{[t,k]} A_k) \rangle, \end{aligned}$$

where \odot denotes the Hadamard product and $m_2[\cdot]$ is interpreted as in (1.25).

Proof We note the following modifications to the proof of [27, Lem. 2.3]. Using the renormalization in (A.1) with \tilde{W} being an independent copy of W (compared to the independent GUE matrix \tilde{W} used in [27]), we obtain the relation

$$\begin{aligned} \langle T_{[1,k]} \rangle - m_1[T_1, \dots, T_k] &= m_1(-\langle W T_{[1,k]} \rangle \\ &\quad + \langle (T_{[2,k]} G_k A_k A_1) \rangle - m_1[T_2, \dots, T_{k-1}, G_k A_k A_1]) \\ &\quad + \sum_{j=1}^{k-1} (\langle (T_{[1,j]} G_j) \rangle - m_1[T_1, \dots, T_{j-1}, G_j]) m_1[T_j, \dots, T_k] \end{aligned}$$

$$\begin{aligned}
 & + \sum_{j=2}^k m_1[T_1, \dots, T_{j-1}, G_j] \langle (T_{[j,k]}) - m_1[T_j, \dots, T_k] \rangle \\
 & + \frac{\sigma}{N} \sum_{j=1}^k \langle (T_{[1,j]} G_j)^t T_{[j,k]} \rangle \\
 & + \frac{\tilde{\omega}_2}{N} \sum_{j=1}^k \langle \text{diag}(T_{[1,j]} G_j) T_{[j,k]} \rangle + \mathcal{O}_{\prec} \left(\frac{1}{N^2} \right). \tag{A.8}
 \end{aligned}$$

Recall that the deterministic approximation $m_1[\cdot]$ is independent of the value of σ and $\tilde{\omega}_2$ (cf. [6, Thm. 3.4]). The terms in the last line of (A.8) involve an additional factor of N^{-1} . Applying Theorem 1.10 as well as the isotropic local law (1.20) thus yields

$$\begin{aligned}
 \langle (T_{[1,j]} G_j)^t T_{[j,k]} \rangle & = m_1[G_j^t A_{j-1}^t, \dots, G_2^t A_1^t, G_1^t, T_j, \dots, T_k] + \mathcal{O}_{\prec} \left(\frac{1}{N} \right) \\
 \langle \text{diag}(T_{[1,j]} G_j) T_{[j,k]} \rangle & = \frac{1}{N} \sum_{r=1}^N (T_{[1,j]} G_j)_{rr} (T_{[j,k]})_{rr} \\
 & = \frac{1}{N} \sum_{r=1}^N (M_{[j]} M_{[j,k]} A_k)_{rr} + \mathcal{O}_{\prec} \left(\frac{1}{\sqrt{N}} \right)
 \end{aligned}$$

and we can replace the respective terms in (A.8) by their deterministic approximation. Note that this also changes the error term to $\mathcal{O}_{\prec}(N^{-3/2})$. Computing the expectation of (A.8) is now analogous to [27, Lem. 2.3]. In particular, the cumulant expansion

$$N \mathbb{E} \langle \underline{W T_{[1,k]}} \rangle = \sum_{n \geq 2} \sum_{x, y \in [N]} \sum_{v \in \{xy, yx\}^n} \frac{\kappa(xy, v)}{n!} \mathbb{E} \partial_v (T_{[1,k]})_{yx}$$

yields the same result, since the term does not involve any second moments of the entries of W .

Next, we identify the limiting covariance of two modes X_α and X_β . This yields an analog of [27, Lem. 2.5] for the model in Assumption 1.1 on macroscopic scales and thus constitutes the base case for the induction argument in the proof of Theorem 2.3.

Lemma A.5 *Fix $k, \ell \in \mathbb{N}$ and let α, β be two multi-indices of length k and ℓ , respectively. Assume that the Wigner matrix W satisfies Assumption 1.1 and pick spectral parameters $z_1, \dots, z_{k+\ell}$ with $|\Im z_j| \gtrsim 1$ and $\max_j |z_j| \leq N^{100}$ as well as deterministic matrices $A_1, \dots, A_{k+\ell}$ with $\|A_j\| \lesssim 1$. Then, for any $\varepsilon > 0$,*

$$N^2 \mathbb{E} X_\alpha^{(k,a)} X_\beta^{(\ell,b)} = m_2[\alpha|\beta] + \mathcal{O} \left(\frac{N^\varepsilon}{\sqrt{N}} \right),$$

with $m_2[\cdot]$ as introduced in Definition 2.1.

Proof of Lemma A.5 The proof is analogous to that of [27, Lem. 2.5] with one key difference. When evaluating the cumulant expansion

$$N^2 \mathbb{E}(\langle \underline{W} T_{[1,k]} \rangle X_\beta^{(\ell)}) = N \sum_{n \geq 1} \sum_{x, y \in [N]} \sum_{v \in \{xy, yx\}^n} \frac{\kappa(xy, v)}{n!} \mathbb{E} \partial_v((T_{[1,k]})_{yx} X_\beta^{(\ell)}),$$

the second moment structure of the model in Assumption 1.1 has to be taken into account for evaluating the $n = 1$ term. We obtain

$$\begin{aligned} & N \sum_{x, y \in [N]} \sum_{v \in \{xy, yx\}} \kappa(xy, v) \mathbb{E} \partial_v((T_{[1,k]})_{yx} X_\beta^{(\ell)}) \\ &= - \sum_{j=1}^{\ell} \mathbb{E}(\langle T_{[1,k]} T_{[k+j, k+\ell]} T_{[1, k+j]} G_{k+j} \rangle + \sigma \langle T_{[1,k]} (T_{[k+j, k+\ell]} T_{[1, k+j]} G_{k+j})^t \rangle \\ & \quad + \widetilde{\omega}_2(\text{diag}(T_{[1,k]}) \text{diag}(T_{[k+j, k+\ell]} T_{[1, k+j]} G_{k+j}))), \end{aligned}$$

which, compared to the computation in the proof of [27, Lem. 2.5], additionally involves σ and $\widetilde{\omega}_2$. The deterministic approximation of the terms follow from (1.19), Theorem 1.10 and (1.20), respectively, which yields the source terms \mathfrak{s}_{GUE} , \mathfrak{s}_σ , and \mathfrak{s}_ω in (2.4). The rest of the expansion evaluates exactly as its counterpart in [27] since the term does not involve any second moments of the entries of W .

The rest of the proof of Theorem 2.3 is analogous to that of [27, Thm. 3.6], i.e., we apply induction on the number of factors in (2.10) using $\mathbb{E} X_\alpha = 0$ and Lemma A.5 as base cases. □

B Additional Proofs and Computations

B.1 Proof of Lemma 1.17

Using (1.16) for the ordered multi-set $S = \{z_1, \dots, z_j, \dots, z_k, z_j\}$, rewrite the left-hand side of (1.32) as

$$m[1, \dots, j, \dots, k, j] = m_1 \dots m_{j-1} m_j^2 m_{j+1} \dots m_k \sum_{\Gamma \in NCG[\{1, \dots, k, j\}]} \prod_{(a,b) \in E(\Gamma)} q_{a,b}. \tag{B.1}$$

Recalling that $m[S]$ is invariant under any permutation of the elements of S , we pick an ordering in which the two j 's occur in two consecutive positions and visualize the corresponding non-crossing graphs by equidistantly arranging the vertices on a circle. In this picture, the edge e connecting both j 's cannot be involved in any crossing, even in an arbitrary planar graph on the given vertices (see left of Fig. 22). Hence, for any non-crossing graph with edges $\{e_1, \dots, e_n\} \not\ni e$, the graph with edge set $\{e_1, \dots, e_n, e\}$ is also non-crossing. In particular, every non-crossing graph that involves e has a counterpart that does not. Next, note that there is at most one vertex l

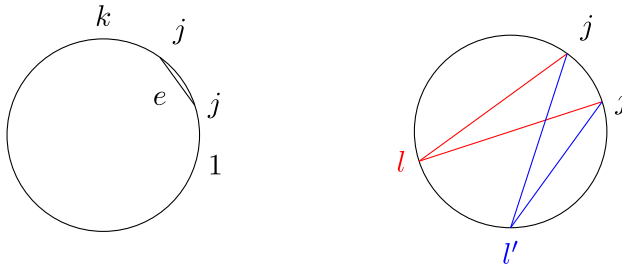


Fig. 22 An edge between two consecutive vertices never results in a crossing (left), but connecting two distinct vertices to both copies of j always does (right). (Color figure online)

among $1, \dots, j-1, j+1, \dots, k$ that is connected to both copies of j , as having two distinct vertices l, l' with this property results in a crossing (see right of Fig. 22).

Consider the disjoint decomposition

$$NCG(\{1, \dots, j, \dots, k, j\}) = \mathcal{S}_1 \cup \left(\bigcup_{l \in [k] \setminus \{j\}} \mathcal{S}_2^{(l)} \right)$$

where $\mathcal{S}_2^{(l)}$ contains all $\Gamma \in NCG(\{1, \dots, j, \dots, k, j\})$ for which the vertex $l \in [k] \setminus \{j\}$ is connected to both copies of j and \mathcal{S}_1 contains any remaining Γ . This implies

$$\sum_{\Gamma \in NCG(\{1, \dots, k, j\})} \prod_{(a,b) \in E(\Gamma)} q_{a,b} = \sum_{\Gamma \in \mathcal{S}_1} \prod_{(a,b) \in E(\Gamma)} q_{a,b} + \sum_{l \in [k] \setminus \{j\}} \sum_{\Gamma \in \mathcal{S}_2^{(l)}} \prod_{(a,b) \in E(\Gamma)} q_{a,b}. \tag{B.2}$$

Next, merge both vertices with the label j . The resulting graphs are either an element of $NCG(k)$ or arise from an element of $NCG(k)$ by adding a loop (j, j) . Recalling that

$$q_{j,j} = \frac{m_j^2}{1 - m_j^2} = m'_j \tag{B.3}$$

where the second equality follows from (1.12), we can factor out $(1 + m'_j)$ on the right-hand side of (B.2) and reduce to summation over $NCG(k)$ without further restriction. Note that this also results in the edge (l, j) doubling whenever $\Gamma \in \mathcal{S}_2^{(l)}$, i.e., we obtain an extra factor of $q_{j,l}$ in this case. Hence,

$$\sum_{\Gamma \in NCG(\{1, \dots, k, j\})} \prod_{(a,b) \in E(\Gamma)} q_{a,b} = (1 + m'_j) \left(\sum_{\Gamma \in NCG(\{1, \dots, k\})} \prod_{(a,b) \in E(\Gamma)} q_{a,b} \right) \left(1 + \sum_{l \in [k] \setminus \{j\}} q_{j,l} \right).$$

Noting that (B.3) is equivalent to $m_j(1 + m'_j) = \frac{m'_j}{m_j}$ and applying (1.16) to recover $m[1, \dots, k]$ yields (1.32). □

B.2 Proof of Corollary 2.11

We start by evaluating $sc[i_1, \dots, i_n]$ and $sc[i_1, \dots, i_n|i_{n+1}, \dots, i_{n+m}]$ before considering the associated free cumulant functions. First, note that

$$sc[i_1, \dots, i_n] = \int_{-2}^2 x^n \rho_{sc}(x) dx = \begin{cases} 0, & \text{if } n \text{ odd,} \\ C_{n/2}, & \text{if } n \text{ even,} \end{cases}$$

where C_0, C_1, C_2, \dots denote the Catalan numbers. In particular, $sc[i_1, \dots, i_n]$ coincides with the number of non-crossing pairings of the set $[n]$. It readily follows that $sc_o[i_1, i_2] = 1$ and that $sc_o[i_1, \dots, i_n] = 0$ whenever n is odd. Moreover, $sc_o[i_1, \dots, i_n] = 0$ for any even $n \geq 4$. The latter follows inductively from (1.14) by writing

$$sc_o[i_1, \dots, i_n] = sc[i_1, \dots, i_n] - \sum_{\pi \in NCP(n) \setminus \{[n]\}} \prod_{B \in \pi} sc_o[B], \tag{B.4}$$

and observing that only pairings contribute to the sum on the right-hand side of (B.4), i.e., the two terms cancel.

Next, we compute $sc[i_1, \dots, i_n|i_{n+1}, \dots, i_{n+m}]$. Observe that by (1.26) and Theorem 2.9,

$$sc[i_1, \dots, i_n|i_{n+1}, \dots, i_{n+m}] = \lim_{N \rightarrow \infty} \mathbb{E}[(\text{Tr } W^n - \mathbb{E}W^n)(\text{Tr } W^m - \mathbb{E}W^m)].$$

The limit on the right-hand side is well-known in the free probability literature (see, e.g., [21]) and hence readily identified as the number of non-crossing pairings of the (n, m) -annulus. Solving (1.26) for $sc_{oo}[i_1, \dots, i_n|i_{n+1}, \dots, i_{n+m}]$, it follows inductively that $sc_{oo}[i_1, \dots, i_n|i_{n+1}, \dots, i_{n+m}] = 0$ for any n, m . Hence, $\Phi_{\pi_1 \times \pi_2, U_1 \times U_2}(f_1, \dots, f_{k+\ell}) = 0$ and

$$\Phi_{\pi}(f_1, \dots, f_{k+\ell}) = \begin{cases} 1, & \text{if } \pi \in \overrightarrow{NCP}_2(k, \ell), \\ 0, & \text{otherwise,} \end{cases}$$

which is the claim. Note that the error in (2.16) evaluates to $\mathcal{O}(N^{\varepsilon-1/2})$ as $\|f_i\|_{HP}$ for $i = 1, \dots, k$ resp. $\|f_j\|_{HQ}$ for $j = k + 1, \dots, k + \ell$ are N -independent constants in the macroscopic regime. □

Acknowledgements I am very grateful to László Erdős for suggesting the topic and many valuable discussions during my work on the project. Partially supported by ERC Advanced Grant ‘‘RMTBeyond’’ No. 101020331.

Funding Open access funding provided by Institute of Science and Technology (IST Austria).

Data Availability Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

Declarations

Conflict of interest The author has no conflict of interest to declare that are relevant to the content of this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Bai, Z.D., Yao, J.: On the convergence of the spectral empirical process of Wigner matrices. *Bernoulli* **11**, 1059–1092 (2005)
2. Bao, Z., He, Y.: Quantitative CLT for linear eigenvalue statistics of Wigner matrices. *Ann. Appl. Probab.* **33**(6), 5171–5207 (2023)
3. Borot, G., Guionnet, A.: Asymptotic expansion of β matrix models in the one-cut regime. *Commun. Math. Phys.* **317**, 447–483 (2013)
4. Cipolloni, G., Erdős, L., Schröder, D.: Eigenstate thermalization hypothesis for Wigner matrices. *Commun. Math. Phys.* **388**(2), 1005–1048 (2021)
5. Cipolloni, G., Erdős, L., Schröder, D.: Optimal multi-resolvent local laws for Wigner matrices. *Electron. J. Probab.* **27**, 1–38 (2022)
6. Cipolloni, G., Erdős, L., Schröder, D.: Thermalization for Wigner matrices. *J. Funct. Anal.* **282**(8), 109394 (2022)
7. Cipolloni, G., Erdős, L., Schröder, D.: Functional central limit theorems for Wigner matrices. *Ann. Appl. Probab.* **33**(1), 447–489 (2023)
8. Collins, B., Mingo, J., Śniady, P., Speicher, R.: Second order freeness and fluctuations of random matrices: III. Higher order freeness and free cumulants. *Doc. Math.* **12**, 1–70 (2007)
9. Diaz, M., Jaramillo, A., Pardo, J.C.: Fluctuations for matrix-valued Gaussian processes. *Ann. Henri Poincaré* **58**(4), 2216–2249 (2022)
10. Diaz, M., Mingo, J.A.: On the analytic structure of second-order non-commutative probability spaces and functions of bounded Fréchet variation. *Random Matrices Theory Appl.* **12**(1), 2250044 (2022)
11. Guionnet, A.: Large deviation upper bounds and central limit theorems for non-commutative functionals of Gaussian large random matrices. *Ann. Inst. H. Poincaré Probab. Stat.* **38**, 341–384 (2002)
12. He, Y., Knowles, A.: Mesoscopic eigenvalue density correlations of Wigner matrices. *Probab. Theory Relat. Fields* **177**, 147–216 (2020)
13. Johansson, K.: On fluctuations of eigenvalues of random Hermitian matrices. *Duke Math. J.* **91**(1), 151–204 (1998)
14. Khorunzhi, A.M., Khoruzhenko, B.A., Pastur, L.A.: On the $1/N$ corrections to the Green functions of random matrices with independent entries. *J. Phys. A Math. Gen.* **28**, L31 (1995)
15. Khorunzhi, A.M., Khoruzhenko, B.A., Pastur, L.A.: Asymptotic properties of large random matrices with independent entries. *J. Math. Phys.* **37**, 5033–5060 (1996)
16. Landon, B., Sosoé, P.: Almost-optimal bulk regularity conditions in the CLT for Wigner matrices. Preprint (2022). [arXiv:2204.03419](https://arxiv.org/abs/2204.03419)
17. Lytova, A.: On non-Gaussian limiting laws for certain statistics of Wigner matrices. *Zh. Mat. Fiz. Anal. Geom.* **9**, 536–581 (2013)
18. Lytova, A., Pastur, L.: Central limit theorem for linear eigenvalue statistics of the Wigner and the sample covariance random matrices. *Metrika* **69**, 153–172 (2009)
19. Lytova, A., Pastur, L.: Fluctuations of matrix elements of regular functions of Gaussian random matrices. *J. Stat. Phys.* **134**, 147–159 (2009)

20. Male, C.: Freeness over the diagonal and global fluctuations of complex Wigner matrices. *J. Oper. Theory* **85**(1), 153–181 (2021)
21. Male, C., Mingo, J.A., Peché, S., Speicher, R.: Joint global fluctuations of complex Wigner and deterministic matrices. *Random Matrices Theory Appl.* **11**(2), 2250015 (2022)
22. Mingo, J.A., Speicher, R.: *Free Probability and Random Matrices*, vol. 35. Fields Institute Research Monographs. Springer, New York (2017)
23. Mingo, J.A., Nica, A.: Annular noncrossing permutations and partitions, and second-order asymptotics for random matrices. *Int. Math. Res. Not.* **2004**(28), 1413–1460 (2004)
24. Mingo, J.A., Speicher, R.: Second order freeness and fluctuations of random matrices I: Gaussian and Wishart random matrices and cyclic Fock spaces. *J. Funct. Anal.* **235**(1), 226–270 (2006)
25. Redelmeier, C.E.I.: Real and quaternionic second-order free cumulants and connections to matrix cumulants. Preprint (2018). [arXiv:1808.10589v2](https://arxiv.org/abs/1808.10589v2)
26. Redelmeier, C.E.I.: 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)
27. Reker, J.: Multi-point functional central limit theorem for Wigner matrices. Preprint (2023). [arXiv:2307.11028](https://arxiv.org/abs/2307.11028)
28. Riabov, V.: Mesoscopic eigenvalue statistics for Wigner-type matrices. Preprint (2023). [arXiv:2301.01712](https://arxiv.org/abs/2301.01712)
29. Shcherbina, M.: Central limit theorem for linear eigenvalue statistics of the Wigner and sample covariance random matrices. *Zh. Mat. Fiz. Anal. Geom.* **7**, 176–192 (2011)
30. Shcherbina, M.: Fluctuations of linear eigenvalue statistics of β matrix models in the multi-cut regime. *J. Stat. Phys.* **151**, 1004–1034 (2013)
31. Sosoe, P., Wong, P.: Regularity conditions in the CLT for linear eigenvalue statistics of Wigner matrices. *Adv. Math.* **249**, 37–87 (2013)
32. Wigner, E.: Characteristic vectors of bordered matrices with infinite dimensions. *Ann. Math.* **62**, 548–564 (1955)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.