The Spectra of Principal Submatrices in Rotationally Invariant Hermitian Random Matrices and the Markov–Krein Correspondence

Katsunori Fujie and Takahiro Hasebe

Department of Mathematics, Hokkaido University, North 10 West 8, Kita-Ku, Sapporo 060-0810, Japan
E-mail address: kfujie@eis.hokudai.ac.jp

Department of Mathematics, Hokkaido University, North 10 West 8, Kita-Ku, Sapporo 060-0810, Japan
E-mail address: thasebe@math.sci.hokudai.ac.jp
URL: https://www.math.sci.hokudai.ac.jp/∼thasebe/

Abstract. We prove a concentration phenomenon on the empirical eigenvalue distribution (EED) of the principal submatrix in a random hermitian matrix whose distribution is invariant under unitary conjugacy; for example, this class includes GUE (Gaussian Unitary Ensemble) and Wishart matrices. More precisely, if the EED of the whole matrix converges to some deterministic probability measure $\mathfrak{m}$, then the difference of rescaled EEDs of the whole matrix and of its principal submatrix concentrates at the Rayleigh measure (in general, a Schwartz distribution) associated with $\mathfrak{m}$ by the Markov–Krein correspondence. For the proof, we use the moment method with Weingarten calculus and free probability. At some stage of calculations, the proof requires a relation between the moments of the Rayleigh measure and free cumulants of $\mathfrak{m}$. This formula is more or less known, but we provide a different proof by observing a combinatorial structure of non-crossing partitions.

1. Introduction

The Markov–Krein correspondence

$$\int_{\mathbb{R}} \frac{1}{1 - zx} \, d\mathfrak{m}(x) = \exp\left[\int_{\mathbb{R}} \log \frac{1}{1 - zx} \, d\tau(x) \right], \qquad z \in \mathbb{C} \setminus \mathbb{R} \tag{1.1}$$

provides a bijection between the probability measures $\mathfrak{m}$ on $\mathbb{R}$, called transition measures, and certain Schwartz distributions τ. In many examples τ is a signed measure, and in such a case τ is called the Rayleigh measure of $\mathfrak{m}$. In general, τ is the derivative (in the sense of Schwartz

distribution) of a so-called Rayleigh function; see Kerov (1998) for further details. The Markov–Krein correspondence appears in different contexts to describe interlacing sequences: limit shapes of large random Young diagrams (Logan and Shepp (1977); Veršik and Kerov (1977); Biane (1998, 2001)); roots of two orthogonal polynomials of large consecutive degrees (Kerov (1993)); eigenvalues of large random matrices and of their principal submatrices, first in the case of randomly rotated real Wigner matrices (Kerov (1993)), and then Wigner and Wishart matrices (without random rotation) (Bufetov (2013)). There are also situations where the distribution τ above appears as a probability measure: Poisson–Dirichlet processes (see Kerov (1998, Section 4.1) and references therein); self-decomposable distributions for monotone convolution (Franz et al. (2020)); Harish-Chandra–Izykson–Zuber integral of rank one at a high temperature regime (Mergny and Potters (2021)). The reason why the same correspondence appears in different contexts seems still unclear to the authors.

In this paper, we prove a concentration phenomenon analogous to those in Bufetov (2013); Kerov (1993) in the setting of rotationally invariant hermitian random matrices. Although our main result seems to be a "folklore theorem" in random matrix theory, there is no proof, and even, only Goel and Yao (2020) among the literature states it explicitly as a conjecture to the authors' best knowledge.

Let X_N be a hermitian random matrix of size N whose distribution is invariant under conjugacy by unitary matrices and let $\Lambda_N = (\lambda_1^{(N)} \leq \cdots \leq \lambda_N^{(N)})$ be its eigenvalues. It is known that a diagonalization $X_N = U_N D_N U_N^*$ exists, where $D_N = \mathrm{diag}(\lambda_1^{(N)}, \lambda_2^{(N)}, \ldots, \lambda_N^{(N)})$ and U_N is a Haar unitary random matrix of size N and independent of D_N (see Collins and Male (2014, Proposition 6.1)).

For the principal submatrix $\tilde{X}_N$ made by removing the last row and column of X_N, Cauchy's interlacing law says that the eigenvalues $\tilde{\Lambda}_N = (\tilde{\lambda}_1^{(N)} \leq \cdots \leq \tilde{\lambda}_{N-1}^{(N)})$ of $\tilde{X}_N$ interlace with Λ_N (see Tao (2012, Exercise 1.3.14)):

$$\lambda_1^{(N)} \leq \tilde{\lambda}_1^{(N)} \leq \lambda_2^{(N)} \leq \tilde{\lambda}_2^{(N)} \leq \cdots \leq \lambda_{N-1}^{(N)} \leq \tilde{\lambda}_{N-1}^{(N)} \leq \lambda_N^{(N)}.$$

In many examples, the empirical eigenvalue distribution $\mathfrak{m}_N = (1/N) \sum_{i=1}^{N} \delta_{\lambda_i^{(N)}}$ of the random matrix X_N converges, as $N \to \infty$, to a non-random probability measure, and we do assume so. Then it is not hard to see (at least with a mild assumption) that the empirical eigenvalue distribution $\tilde{\mathfrak{m}}_N$ of $\tilde{X}_N$ also converges to the same limit. Our main result roughly says that the Rayleigh measure

$$\widehat{\tau}_N := N\mathfrak{m}_N - (N-1)\tilde{\mathfrak{m}}_N = \sum_{i=1}^{N} \delta_{\lambda_i^{(N)}} - \sum_{j=1}^{N-1} \delta_{\tilde{\lambda}_j^{(N)}}$$

is close to the Rayleigh measure τ_N linked to the transition measure $\mathfrak{m}_N$ by the Markov–Krein correspondence. Note that τ_N is of the form

$$\tau_N = \sum_{i=1}^{N} \delta_{\lambda_i^{(N)}} - \sum_{j=1}^{N-1} \delta_{\eta_j^{(N)}}$$

where $(\eta_1^{(N)} \leq \cdots \leq \eta_{N-1}^{(N)})$ is a sequence also interlacing with Λ_N (see Kerov (1998, Eq. (2))).

Since our arguments are based on the moment method, we denote by $M_k(\zeta)$ for simplicity the k-th moment of a measure or Schwartz distribution ζ when it is well defined. It should be noted here that if a probability measure $\mathfrak{m}$ has finite moments of all orders, then τ defined via (1.1) also has finite moments of all orders (see Aronszajn and Donoghue (1964, Theorem A (d)) and Kerov (1998, Section 3.4)). Furthermore, for convenience of statements, let $\widehat{\mathfrak{m}}_N$ be the transition measure associated to the Rayleigh measure $\widehat{\tau}_N$; then the main result can alternatively be phrased that $\widehat{\mathfrak{m}}_N$ is close to $\mathfrak{m}_N$.

The precise statement of the main result is as follows, which answers to a conjecture announced by Goel and Yao (2020).

Theorem 1.1. *Let $\mathfrak{m}_N, \tau_N, \widehat{\mathfrak{m}}_N, \widehat{\tau}_N$ be as above, $\mathfrak{m}$ be a probability measure on $\mathbb{R}$ and τ be related to $\mathfrak{m}$ via* (1.1). *Assume that*

$$\sup_{N \geq 1} \mathbb{E}[M_k(\mathfrak{m}_N)] < \infty \quad and \quad M_k(\mathfrak{m}) < \infty, \qquad k \in 2\mathbb{N} \tag{1.2}$$

and $\mathfrak{m}_N$ converges in moments to $\mathfrak{m}$ in probability:

$$\lim_{N \to \infty} \mathbb{P}[|M_k(\mathfrak{m}_N) - M_k(\mathfrak{m})| \geq \epsilon] = 0, \qquad k \in \mathbb{N}, \ \epsilon > 0. \tag{1.3}$$

Then we have

$$\lim_{N \to \infty} \| M_k(\widehat{\tau}_N) - M_k(\tau) \|_{L^2} = 0, \qquad k \in \mathbb{N},$$

and

$$\lim_{N \to \infty} \mathbb{P}[|M_k(\widehat{\mathfrak{m}}_N) - M_k(\mathfrak{m})| \geq \epsilon] = 0, \qquad k \in \mathbb{N}, \ \epsilon > 0.$$

In particular, if the moment problem for $\{M_k(\mathfrak{m})\}_{k \geq 1}$ is determinate then $\widehat{\mathfrak{m}}_N$ weakly converges to $\mathfrak{m}$ in probability:

$$\lim_{N \to \infty} \mathbb{P}\left[\left| \int_{\mathbb{R}} f(x) \, \widehat{\mathfrak{m}}_N(dx) - \int_{\mathbb{R}} f(x) \, \mathfrak{m}(dx) \right| \geq \epsilon \right] = 0, \qquad f \in C_b(\mathbb{R}), \ \epsilon > 0.$$

Remark 1.2.   (i) Since the relation between the moments $\{M_n(\mathfrak{m})\}_{n \in \mathbb{N}}$ (resp. $\{M_n(\mathfrak{m}_N)\}_{n \in \mathbb{N}}$) and $\{M_k(\tau)\}_{k \in \mathbb{N}}$ (resp. $\{M_k(\tau_N)\}_{k \in \mathbb{N}}$) is the same as that between complete symmetric functions and Newton power sums (see (3.1) below), the convergence (1.3) holds if and only if τ_N converges in moments to τ in probability.
 (ii) (1.2) and (1.3) imply the convergence of moments in L^p norm for every $p \in [1, \infty)$; see Proposition A.3.
(iii) The assumptions (1.2) and (1.3) are satisfied by appropriately normalized Gaussian Unitary Ensemble (GUE) Hiai and Petz (2000, Theorem 4.1.5), where $\mathfrak{m}$ is Wigner's semicircle law $(1/(2\pi))\sqrt{4 - x^2} \, dx$. For GUE (actually, more general Wigner matrices), a finer result on the fluctuation of $\widehat{\tau}_N$ from τ is also known in Erdős and Schröder (2018) stated in the language of rectangular Young diagrams; see also Sodin (2017).

The proof is based on Weingarten calculus and free probability which allow us to compute the moments of the principal submatrix:

$$\mathbb{E} \circ \mathrm{Tr}[(\tilde{X}_N)^k] = \mathbb{E} \circ \mathrm{Tr}[D_N U_N P_N U_N^* D_N U_N P_N U_N^* \cdots D_N U_N P_N U_N^*], \tag{1.4}$$

where $P_N = \mathrm{diag}(1, 1, \ldots, 1, 0)$.

In fact, the joint distribution of $(\tilde{\lambda}_1^{(N)} \leq \cdots \leq \tilde{\lambda}_{N-1}^{(N)})$ is explicit under the condition that $(\lambda_1^{(N)} \leq \cdots \leq \lambda_N^{(N)})$ is a constant sequence; it is proportional to the Vandermond determinant Baryshnikov (2001, Proposition 4.2) (see also the expository paper Faraut (2015)). Using this explicit formula might be an alternative approach for computing (1.4) and hence for a proof of Theorem 1.1; however, the authors are not sure whether this direction is promising.

At some stage of calculations of (1.4) with Weingarten calculus, it turns out that the following formula (1.5) is crucial.

Theorem 1.3. *Suppose that $\mathfrak{m}$ is a probability measure on $\mathbb{R}$ with finite moments of all orders and τ be defined via* (1.1). *Then the formula*

$$M_k(\tau) = \sum_{\rho \in NC(k)} (k + 1 - |\rho|) \, R_\rho(\mathfrak{m}) \tag{1.5}$$

holds for every $k \in \mathbb{N}$, where $NC(k)$ is the set of non-crossing partitions of $\{1, \ldots, k\}$ and $R_\rho(\mathfrak{m})$ is the free cumulant of $\mathfrak{m}$.

This formula gives an explicit combinatorial relation between two bases in the Kerov-Olshanski algebra: the moments of τ and free cumulants of $\mathfrak{m}$. It can be easily proved by combining known formulas for complete symmetric functions as follows. The moments of τ and the free cumulants of $\mathfrak{m}$ can be identified with the elements $\{p_n(A)\}_{n\geq 1}$ and $\{(-1)^n e_n^*(A)\}_{n\geq 1}$ in Lassalle (2009), respectively; the latter fact is noted on page 2242 of Lassalle (2009). Combining (4.5) and the formula right before (4.10) in Lassalle (2009) allows one to express $\{p_n^*(A)\}_{n\geq 1}$ in terms of $\{e_n(A)\}_{n\geq 1}$ as a sum over integer partitions. Applying the involution gives a formula that expresses $\{p_n(A)\}_{n\geq 1}$ in terms of $\{(-1)^n e_n^*(A)\}_{n\geq 1}$. This formula can be transformed into the sum over non-crossing partitions via Nica and Speicher (2006, Corollary 9.12), which amounts to Theorem 1.3.

In this paper, we provide a different proof of Theorem 1.3 based on non-crossing partitions. A key observation is that the coefficient $k + 1 - |\rho|$ coincides with the cardinality of the Kreweras complement $K(\rho)$. Since the coefficients are so simple, one may expect that there is a combinatorial structure behind. Indeed, we will introduce the notion of "Kreweras decomposition" of a non-crossing partition and count the number of such decompositions as a crucial ingredient of the proof.

After this introduction, this paper is structured as follows. Section 2 consists of preliminaries on Weingarten calculus, free probability and symmetric groups. Section 3 provides the proof of Theorem 1.1, as well as an alternative proof of Theorem 1.3 as mentioned. Some results on convergence of random measures are proved in Appendix A.

2. Preliminaries

In this section, we introduce standard notions in free probability and related fields for later use in the proof of the main results.

2.1. *Weingarten calculus.* Computation of mixed moments of Haar unitary random matrices U_N and deterministic matrices is called Weingarten calculus. For $\sigma \in \mathfrak{S}_k$, let $\mathrm{Tr}_\sigma[A_1, A_2, \ldots, A_k]$ be the product of traces according to the cycle decomposition of σ; for example if $\sigma = (1,3,2,5)(4)(6,9)(7,8)$ then $\mathrm{Tr}_\sigma[A_1, A_2, \ldots, A_9] = \mathrm{Tr}(A_1 A_3 A_2 A_5)\,\mathrm{Tr}(A_4)\,\mathrm{Tr}(A_6 A_9)\,\mathrm{Tr}(A_7 A_8)$. Similarly, for a sequence $\{\alpha_n\}_{n\geq 1} \subset \mathbb{C}$ we define α_σ to be the product of α_n's according to the sizes of cycles; in the above example, $\alpha_\sigma = \alpha_4 \alpha_1 \alpha_2{}^2$.

Let A_i, B_i $(i = 1, \ldots, k)$ be $N \times N$ matrices. Then

$$\mathbb{E} \circ \mathrm{Tr}_\sigma[A_1 U_N B_1 U_N{}^*, \ldots, A_k U_N B_k U_N{}^*]$$
$$= \sum_{\substack{\sigma_1, \sigma_2, \sigma_3 \in \mathfrak{S}_k \\ \sigma_1 \sigma_2 \sigma_3 = \sigma}} \mathrm{Tr}_{\sigma_1}[A_1, \ldots, A_k]\,\mathrm{Tr}_{\sigma_2}[B_1, \ldots, B_k]\,\mathrm{Wg}(\sigma_3, N)$$

for all $\sigma \in \mathfrak{S}_k$. In particular, in the case of $\sigma = \gamma_k = (1, 2, \ldots, k)$ the above formula specializes to

$$\mathbb{E} \circ \mathrm{Tr}[(A_1 U_N B_1 U_N{}^*) \cdots (A_k U_N B_k U_N{}^*)]$$
$$= \sum_{\sigma, \pi \in \mathfrak{S}_k} \mathrm{Tr}_\sigma[A_1, \ldots, A_k]\,\mathrm{Tr}_\pi[B_1, \ldots, B_k]\,\mathrm{Wg}(\pi^{-1}\sigma^{-1}\gamma_k, N), \tag{2.1}$$

see Collins and Śniady (2006, Proposition 2.3). The coefficients $\mathrm{Wg}(\sigma, N)$ are called the Weingarten function. Its asymptotic behavior for large N is known in the form

$$N^{k+|\sigma|}\,\mathrm{Wg}(\sigma, N) = \mu_k(\sigma) + O\left(\frac{1}{N^2}\right), \qquad \sigma \in \mathfrak{S}_k. \tag{2.2}$$

The number $|\sigma|$, called the length function, is the minimal number l for which σ can be written as a product of l transpositions, and the number $\mu_k(\sigma)$ above is expressed in terms of the Catalan

numbers $C_n = (2n)!/(n!(n+1)!)$ as

$$\mu_k(\sigma) = \prod_{1 \leq j \leq l} (-1)^{|\pi_j|} C_{|\pi_j|}$$

where $\sigma = \pi_1 \cdots \pi_l$ is the cycle decomposition of σ; see Collins and Matsumoto (2017, Theorem 2.7).

2.2. *Free cumulants.* This section summarizes notations and facts on free cumulants. The reader is referred to Nica and Speicher (2006) for further details. Let L be a finite linearly ordered set. A partition of L is the collection of nonempty disjoint subsets of L whose union is L. For a partition $\rho = \{B_1, B_2, ..., B_r\}$ of L, each element B_i is called a block and the cardinality r is denoted by $|\rho|$. It is referred to as a crossing partition if there are two blocks B_i, B_j ($i \neq j$) and elements $a, b \in B_i, c, d \in B_j$ such that $a < c < b < d$. Otherwise, it is called a non-crossing partition. The set of the non-crossing partitions of L will be denoted by $\mathrm{NC}(L)$. In particular, when $L = [k] := \{1, 2, \ldots, k\}$, $\mathrm{NC}([k])$ is simply denoted by $\mathrm{NC}(k)$.

For $\nu, \rho \in \mathrm{NC}(k)$, the notation $\nu \leq \rho$ means that for every $B \in \nu$ there exists $C \in \rho$ such that $B \subseteq C$. This defines a poset structure on $\mathrm{NC}(k)$. The maximum element regarding this partial order is $1_k := \{[k]\}$, the partition consisting of one block $[k]$, and the minimum is $0_k := \{\{1\}, \{2\}, \ldots, \{k\}\}$.

The Kreweras complement of a non-crossing partition $\rho \in \mathrm{NC}(k)$ is defined as follows. Inserting additional points $[\bar{k}] := \{\bar{1}, \bar{2}, \ldots, \bar{k}\}$ to $[k]$, suppose that $L_k = \{1, \bar{1}, 2, \bar{2}, \ldots, k, \bar{k}\}$ is a linearly ordered set with the order as displayed. Take the maximal non-crossing partition ν of $[\bar{k}]$ such that $\rho \cup \nu \in \mathrm{NC}(L_k)$. Then deleting bars over the integers, ν is called the Kreweras complement of ρ and denoted by $K(\rho)$. For convenience, we sometimes keep the bars and regard $K(\rho)$ as a non-crossing partition on $[\bar{k}]$.

Example 2.1. If $\rho = \{\{1,7\}, \{2,5,6\}, \{3\}, \{4\}, \{8,9\}\}$ then the following picture
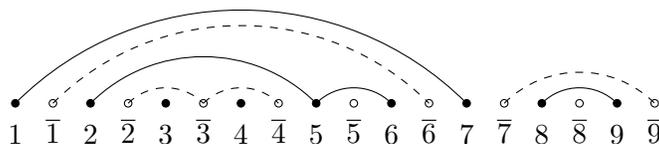
shows that $K(\rho) = \{\{1,6\}, \{2,3,4\}, \{5\}, \{7,9\}, \{8\}\}$.

For a sequence $\{\alpha_n\}_{n \in \mathbb{N}} \subset \mathbb{C}$ and a partition ρ of $[k]$, define

$$\alpha_\rho := \prod_{B \in \rho} \alpha_{|B|}.$$

For a probability measure $\mathfrak{m}$ on $\mathbb{R}$ with finite moments of all orders, the free cumulants $\{\mathrm{R}_k(\mathfrak{m})\}_{k \in \mathbb{N}}$ of $\mathfrak{m}$ are determined recursively by the moment–cumulant formula

$$M_\nu(\mathfrak{m}) = \sum_{\substack{\rho \in \mathrm{NC}(k) \\ \rho \leq \nu}} \mathrm{R}_\rho(\mathfrak{m}), \qquad \nu \in \mathrm{NC}(k), \quad k \in \mathbb{N}. \tag{2.3}$$

Actually, it suffices to take $\nu = 1_k, k = 1, 2, 3, \ldots$ to determine the free cumulants, and then the above formula can be proved for all $\nu \in \mathrm{NC}(k), k = 1, 2, 3, \ldots$. More explicitly, free cumulants can be expressed as

$$\mathrm{R}_\rho(\mathfrak{m}) = \sum_{\substack{\nu \in \mathrm{NC}(k) \\ \nu \leq \rho}} M_\nu(\mathfrak{m}) \tilde{\mu}_k(\nu, \rho), \qquad \rho \in \mathrm{NC}(k), \quad k \in \mathbb{N}, \tag{2.4}$$

where $\tilde{\mu}_k$ is the Möbius function on the poset $\mathrm{NC}(k)$.

2.3. *Non-crossing partitions and symmetric groups.* The set of non-crossing partitions can be embedded into the symmetric group. Here we collect needed facts. For further details, the reader is referred to Nica and Speicher (2006).

The length function (see Section 2.1) on symmetric groups satisfies the following properties: for all $\sigma, \pi \in \mathfrak{S}_k$,

$$|\pi\sigma\pi^{-1}| = |\sigma|,$$
$$|\sigma\pi| \leq |\sigma| + |\pi|,$$
$$|\sigma\pi| \equiv |\sigma| + |\pi| \pmod 2. \tag{2.5}$$

The number $\#(\sigma)$ of cycles in the cycle decomposition of σ is known to satisfy

$$\#(\sigma) + |\sigma| = k.$$

Let d be a metric on $\mathfrak{S}_k$ defined by $d(\sigma, \pi) = |\sigma^{-1}\pi|$. The geodesic set from the unit e to $\gamma_k := (1, \ldots, k)$ is defined by

$$\mathfrak{S}_{\mathrm{NC}}(\gamma_k) = \{\, \sigma \in \mathfrak{S}_k \mid d(e, \sigma) + d(\sigma, \gamma_k) = d(e, \gamma_k) \,(= k - 1) \,\}.$$

For $\sigma, \pi \in \mathfrak{S}_{\mathrm{NC}}(\gamma_k)$, denote by $\sigma \leq \pi$ if σ and π are on a common geodesic and $d(e, \sigma) \leq d(e, \pi)$, namely, if $d(e, \sigma) + d(\sigma, \pi) = d(e, \pi)$, or equivalently, $|\sigma| + |\sigma^{-1}\pi| = |\pi|$.

For a partition $\rho \in \mathrm{NC}(k)$, each block $B = \{i_1, i_2, \ldots, i_p\} \in \rho$ whose elements are arranged in the increasing order associates the cyclic permutation $\pi = (i_1, i_2, \ldots, i_p)$, so that ρ associates the permutation $\mathcal{P}_\rho := \pi_1\pi_2\cdots\pi_l$, where $l = \#(\mathcal{P}_\rho) = |\rho|$. This embedding becomes a poset isomorphism

$$\mathcal{P}\colon \mathrm{NC}(k) \longrightarrow \mathfrak{S}_{\mathrm{NC}}(\gamma_k),$$

see Nica and Speicher (2006, Proposition 23.23). We need the following facts later: for $\sigma = \mathcal{P}_\nu$, $\pi = \mathcal{P}_\rho \in \mathfrak{S}_{\mathrm{NC}}(\gamma_k)$,

(1) the relation $\nu \leq \rho$ holds in $\mathrm{NC}(k)$ if and only if $|\sigma| + |\sigma^{-1}\pi| + |\pi^{-1}\gamma_k| = k - 1$,
(2) $\mu_k(\sigma^{-1}\pi) = \tilde{\mu}_k(\nu, \rho)$,
(3) $\pi^{-1}\gamma_k = \mathcal{P}_{K(\rho)}$; in particular $\#(\pi^{-1}\gamma_k) = |K(\rho)|$.

Similar results hold for $\gamma_k^{(1)}\gamma_k^{(2)} \in \mathfrak{S}_{2k}$ instead of γ_{2k}, where

$$\gamma_k^{(1)} = (1, \ldots, k)(k+1)\cdots(2k) \qquad \text{and} \qquad \gamma_k^{(2)} = (1)\cdots(k)(k+1, \ldots, 2k).$$

Correspondingly, let

$$(1_k^{(1)}, 1_k^{(2)}) := \{\{1, \ldots, k\}, \{k+1, \ldots, 2k\}\},$$

then $\mathcal{P}_{(1_k^{(1)}, 1_k^{(2)})} = \gamma_k^{(1)}\gamma_k^{(2)}$. Via the embedding

$$\mathrm{NC}(k) \times \mathrm{NC}(k) \cong [0_{2k}, (1_k^{(1)}, 1_k^{(2)})] \subset \mathrm{NC}(2k)$$

the restriction of the mapping $\mathcal{P}$ induces an isomorphism between $\mathrm{NC}(k) \times \mathrm{NC}(k)$ and

$$\mathfrak{S}_{\mathrm{NC}}(\gamma_k^{(1)}\gamma_k^{(2)}) = \{\, \sigma \in \mathfrak{S}_{2k} \mid d(e, \sigma) + d(\sigma, \gamma_k^{(1)}\gamma_k^{(2)}) = d(e, \gamma_k^{(1)}\gamma_k^{(2)}) \,(= 2k - 2) \,\}. \tag{2.6}$$

For $(\rho_1, \rho_2) \in \mathrm{NC}(k) \times \mathrm{NC}(k)$ and $\pi = \mathcal{P}_{(\rho_1, \rho_2)} \in \mathfrak{S}_{\mathrm{NC}}(\gamma_k^{(1)}\gamma_k^{(2)})$, the element $\pi^{-1}\gamma_k^{(1)}\gamma_k^{(2)}$ corresponds to $(K(\rho_1), K(\rho_2))$ under the isomorphism $\mathcal{P}$, and in particular $\#(\pi^{-1}\gamma_k^{(1)}\gamma_k^{(2)}) = |K(\rho_1)| + |K(\rho_2)| = |K(\rho)| + 1$, where $\rho = (\rho_1, \rho_2)$ is regarded as a partition in $\mathrm{NC}(2k)$. Note that this relation can be clearly understood in terms of the relative Kreweras complement; however, we will not use this technical notion since it is not directly needed in this paper.

3. Proof of the main results

3.1. *Proof of Theorem 1.3.* Let us start to prove the combinatorial formula (1.5) by induction on the degree k. In this subsection, we keep the assumptions and notation in Theorem 1.3. To begin, the original formula for the Markov–Krein correspondence (1.1) implies the recursive relation

$$M_k(\tau) = k\, M_k(\mathfrak{m}) - \sum_{r=1}^{k-1} M_r(\tau)\, M_{k-r}(\mathfrak{m}), \qquad k \in \mathbb{N}, \tag{3.1}$$

which is exactly the relation satisfied by complete symmetric functions and Newton power sums (Kerov (1998, (3.2.4) and Section 3.4)).

Thanks to the moment-cumulant formula (2.3), the RHS of the desired formula (1.5) may be transformed into

$$\sum_{\rho \in NC(k)} (k+1-|\rho|)\, R_\rho(\mathfrak{m}) = k \sum_{\rho \in NC(k)} R_\rho(\mathfrak{m}) - \sum_{\rho \in NC(k)} (|\rho|-1)\, R_\rho(\mathfrak{m})$$

$$= k\, M_k(\mathfrak{m}) - \sum_{\rho \in NC(k)} (|\rho|-1)\, R_\rho(\mathfrak{m}).$$

Hence, according to the recursive equation (3.1), formula (1.5) is eventually equivalent to

$$\sum_{\rho \in NC(k)} (|\rho|-1)\, R_\rho(\mathfrak{m}) = \sum_{r=1}^{k-1} M_r(\tau)\, M_{k-r}(\mathfrak{m}). \tag{3.2}$$

By the induction hypothesis up to the degree $k-1$ and the moment-cumulant formula, the RHS of (3.2) can be written as

$$\sum_{r=1}^{k-1} \sum_{\substack{\overline{\rho} \in NC(r) \\ \underline{\rho} \in NC(k-r)}} |K(\overline{\rho})|\, R_{\overline{\rho}}(\mathfrak{m})\, R_{\underline{\rho}}(\mathfrak{m}). \tag{3.3}$$

The cardinality $|K(\overline{\rho})|$ can be interpreted as the number of inserting $\underline{\rho}$ into $\overline{\rho}$ in the following way:

(P1) pick $r \in \{1, 2, \ldots, k-1\}$, $\overline{\rho} \in NC(r)$ and $\underline{\rho} \in NC(k-r)$;

(P2) pick a block B of $K(\overline{\rho})$, where $K(\overline{\rho})$ is interpreted as a partition on the points $[\overline{r}]$ interlacing with $[r]$;

(P3) substitute the partition $\underline{\rho}$ into the last point of B.

The steps (P2) and (P3) provide a way to insert $\underline{\rho}$ into $\overline{\rho}$, which yields a non-crossing partition $\rho \in NC(k)$; see also Example 3.1. The sum (3.3) can then be expressed as

$$\sum_{\rho} R_\rho(\mathfrak{m}), \tag{3.4}$$

where ρ runs over all the non-crossing partitions appearing as a result of (P1)–(P3). Note that the same non-crossing partition ρ may appear more than once, and the sum (3.4) needs to count the multiplicity. Actually, in order to have (3.2), we need to demonstrate that each $\rho \in NC(k)$ appears exactly $|\rho|-1$ times. To achieve this, we introduce the notion of Kreweras decomposition of a non-crossing partition, which describes the relation between $\rho, \overline{\rho}$ and $\underline{\rho}$ above.

Example 3.1. For the non-crossing partitions $\overline{\rho} = \{\{1, 7\},\ \{2, 5, 6\},\ \{3\},\ \{4\},\ \{8, 9\}\}$ and $\underline{\rho} = \{\{1, 3\}, \{2\}\}$, the Kreweras complement $K(\overline{\rho})$ is the partition described by the dashed curves below
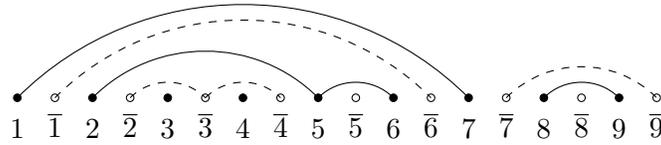
and hence the Kreweras complement has the blocks $\{\overline{1}, \overline{6}\}, \{\overline{2}, \overline{3}, \overline{4}\}, \{\overline{5}\}, \{\overline{7}, \overline{9}\}, \{\overline{8}\}$. According to (P3) we are allowed to place $\underline{\rho}$ at any point of $\{\overline{6}, \overline{4}, \overline{5}, \overline{8}, \overline{9}\}$. For example, if we choose $\overline{4}$ then the resulting non-crossing partition ρ is


Definition 3.2.   (1) For $\nu \in \mathrm{NC}(r)$, a Kreweras point of ν is the last point of a block of the Kreweras complement $K(\nu)$ regarded as a partition on $[\overline{r}]$ that interlaces with $[r]$.
(2) For $\rho \in \mathrm{NC}(k)$, a pair $(\overline{\rho}, \underline{\rho})$ of nonempty disjoint subsets of ρ such that $\overline{\rho} \cup \underline{\rho} = \rho$ and the union of all elements of $\underline{\rho}$ is an interval of $[k]$, that is, there exist some $i < j$ such that

$$\bigcup_{V \in \underline{\rho}} V = [i, j] = \{i, i+1, \ldots, j\}.$$

If the position of $\underline{\rho}$ is a Kreweras point of $\overline{\rho}$, then we call $(\overline{\rho}, \underline{\rho})$ a Kreweras decomposition of ρ, $\overline{\rho}$ an outer partition of ρ and $\underline{\rho}$ an inner partition of ρ.

Example 3.3. The non-crossing partition $\rho = \{\{1, 8\}, \{2, 3\}, \{4, 6, 7\}, \{5\}, \{9, 10\}\}$ can be described as
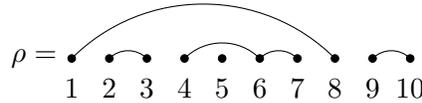
and it has the four inner partitions $\rho_1 = \{\{2, 3\}, \{4, 6, 7\}, \{5\}\}, \rho_2 = \{\{4, 6, 7\}, \{5\}\}, \rho_3 = \{\{5\}\}$, $\rho_4 = \{\{9, 10\}\}$. Any other subsets of ρ are not inner partitions; for example, $\rho' = \{\{2, 3\}\}$ has the support $\{2, 3\}$ of interval form, but the Kreweras complement of $\rho \setminus \rho'$ is described by the dashed curves and white singletons in the picture
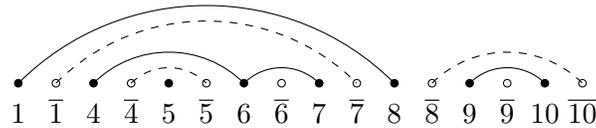
so that the position of the removed block $\{2, 3\}$ was at the point $\overline{1}$, which was not the last point of the block $\{\overline{1}, \overline{7}\}$.

The goal is then to demonstrate that each $\rho \in \mathrm{NC}(k)$ has exactly $|\rho| - 1$ Kreweras decompositions. The proof is based on induction, which depends on the following nesting structure of inner partitions.

Lemma 3.4. *Suppose that $\rho \in \mathrm{NC}(k)$ and its first block which contains 1 divides $[k]$ into (non-empty) l segments $I_1, \ldots, I_l$. Then $\rho_j := \rho|_{I_j}$ is an inner partition of ρ for every j, and moreover, every inner partition of ρ_j is an inner partition of ρ. Conversely, any inner partition of ρ is some ρ_j or its inner partition.*

Proof: It is clear that all ρ_j $(j = 1, \ldots, l)$ are inner partitions of ρ. Then we take any inner partition $\underline{\rho_j}$ of ρ_j for $j = 1, \ldots, l$. Note that the Kreweras complement $K(\rho_j)$ equals $K(\rho)$ restricted to the interval I_j. Hence, since $\underline{\rho_j}$ is at a Kreweras point of the outer partition $\overline{\rho_j} = \rho_j \setminus \underline{\rho_j}$, $\underline{\rho_j}$ is also at a Kreweras point of $\rho \setminus \underline{\rho_j}$.

Conversely, we take any inner partition $\underline{\rho}$ of ρ. By the definition of inner partitions, $\underline{\rho}$ is supported on some interval I_j. If $\underline{\rho}$ contains the first block of ρ_j, then $\underline{\rho}$ equals ρ_j. Otherwise, the support of $\underline{\rho}$ is a sub-interval of I_j which does not intersect the first block of ρ_j, and since $\underline{\rho}$ is at a Kreweras point of the outer partition $\overline{\rho}$, $\underline{\rho}$ is also at a Kreweras point of $\rho_j \setminus \underline{\rho}$. $\qquad\square$

Proposition 3.5. *Let $k \geq 2$. Each $\rho \in \mathrm{NC}(k)$ has exactly $|\rho| - 1$ Kreweras decompositions.*

Proof: The proof runs by induction. It is clear that the statement is true when $k = 2$. Then we assume the statement is true up to $k - 1$ and take $\rho \in \mathrm{NC}(k)$ ($|\rho| > 1$). Suppose that the first block of ρ divides $[k]$ into l segments $I_1, \ldots, I_l$. Then all $\{\rho_j = \rho|_{I_j}\}_{j=1}^{l}$ are inner partitions of ρ. By Lemma 3.4, a subset of ρ is an inner partition of ρ if and only if it is one of $\{\rho_j\}_{j=1}^{l}$ or an inner partition of some ρ_j. Therefore, by the induction hypothesis, the number of inner partitions of ρ is $l + \sum_{j=1}^{l}(|\rho_j| - 1) = |\rho| - 1$. $\qquad\square$

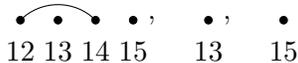*Example* 3.6. We take $\rho \in \mathrm{NC}(27)$ to be
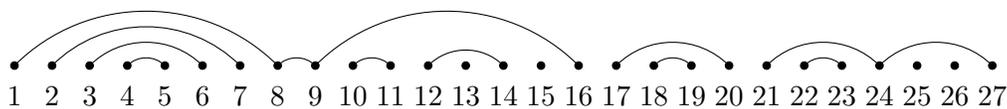
in which $|\rho| = 14$. The three non-crossing partitions
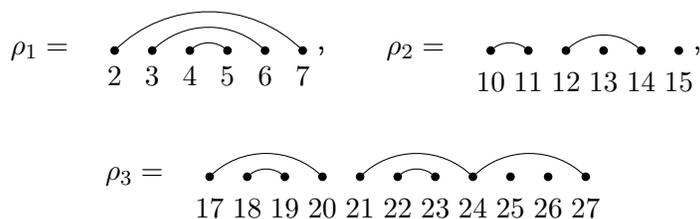
are inner partitions of ρ and the two inner partitions of ρ_1

are inner partitions of ρ. In the same way, the three inner partitions of ρ_2

and the five inner partitions of ρ_3
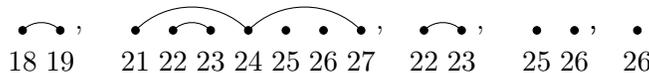
are also inner partitions of ρ. Thus ρ has 13 inner partitions: ρ_1, ρ_2, ρ_3 and the inner partitions of them.

3.2. *Proof of Theorem 1.1.* In this subsection, we follow the notation in Theorem 1.1. The index N is omitted for readability when no confusion occurs. The main part of the proof of Theorem 1.1 is the following.

Theorem 3.7. *Assume that*

$$\sup_{N\geq 1} \mathbb{E}[M_k(\mathfrak{m}_N)] < \infty, \qquad k \in 2\mathbb{N}. \tag{3.5}$$

Then, for every $k \in \mathbb{N}$ and $\ell \in \{1,2\}$, it holds that

$$\mathbb{E}[M_k(\widehat{\tau}_N)^\ell] = \mathbb{E}[M_k(\tau_N)^\ell] + O\left(\frac{1}{N}\right).$$

Remark 3.8. Whether the above result holds for $\ell \geq 3$ is unknown.

Proof: Note first that the assumption (3.5) implies that

$$\sup_{N\geq 1} \mathbb{E}[|M_\sigma(\mathfrak{m}_N)|] < \infty, \qquad \sigma \in \mathfrak{S}_k$$

for every $k \in \mathbb{N}$, thanks to the iterative use of Schwarz inequality and (A.6).

(i) $\ell = 1$. A key of the proof is the calculations of

$$\sum_{j=1}^{N-1} \mathbb{E}[\widetilde{\lambda}_j{}^k] = \mathbb{E} \circ \mathrm{Tr}[(PU^*DUP)^k], \tag{3.6}$$

where $P = \mathrm{diag}(1, \ldots, 1, 0)$. The RHS of (3.6) is calculated into

$$\mathbb{E} \circ \mathrm{Tr}[(PU^*DUP)^k] = \mathbb{E} \circ \mathrm{Tr}[(DUPU^*)^k]$$

$$= \sum_{\sigma,\pi \in \mathfrak{S}_k} \mathbb{E} \circ \mathrm{Tr}_\sigma[D, \ldots, D] \, \mathrm{Tr}_\pi[P, \ldots, P] \, \mathrm{Wg}(\pi^{-1}\sigma^{-1}\gamma_k) \tag{3.7}$$

$$= \sum_{\sigma,\pi \in \mathfrak{S}_k} \mathbb{E} \circ \mathrm{Tr}_\sigma[D, \ldots, D] \, \mathrm{Tr}_{\pi^{-1}\gamma_k}[P, \ldots, P] \, \mathrm{Wg}(\sigma^{-1}\pi)$$

$$= \sum_{\sigma,\pi \in \mathfrak{S}_k} N^{\#(\sigma)} \mathbb{E} \circ \mathrm{tr}_\sigma[D, \ldots, D] (N-1)^{\#(\pi^{-1}\gamma_k)} \, \mathrm{Wg}(\sigma^{-1}\pi), \tag{3.8}$$

where (2.1) was used on the second line and the change of variables $\pi \mapsto \pi^{-1}\gamma_k$ was employed on the third line.

On the other hand, if the projection P is replaced by the identity I in (3.7), then the same calculations lead to

$$\sum_{i=1}^{N} \mathbb{E}[\lambda_i{}^k] = \mathbb{E} \circ \mathrm{Tr}[D^k] = \mathbb{E} \circ \mathrm{Tr}[(DUIU^*)^k]$$

$$= \sum_{\sigma,\pi \in \mathfrak{S}_k} N^{\#(\sigma)} \mathbb{E} \circ \mathrm{tr}_\sigma[D, \ldots, D] N^{\#(\pi^{-1}\gamma_k)} \, \mathrm{Wg}(\sigma^{-1}\pi). \tag{3.9}$$

Taking the difference of (3.8) and (3.9) provides

$$\mathbb{E}[M_k(\widehat{\tau}_N)] = \sum_{i=1}^{N} \mathbb{E}[\lambda_i{}^k] - \sum_{j=1}^{N-1} \mathbb{E}[\widetilde{\lambda}_j{}^k]$$

$$= \sum_{\sigma,\pi \in \mathfrak{S}_k} N^{\#(\sigma)} \mathbb{E} \circ \mathrm{tr}_\sigma[D, \ldots, D] \#(\pi^{-1}\gamma_k) N^{\#(\pi^{-1}\gamma_k)-1} \left(1 + O(N^{-1})\right) \mathrm{Wg}(\sigma^{-1}\pi).$$

Here we use the asymptotic expansion (2.2) of the Weingarten functions to get

$$\mathbb{E}[M_k(\widehat{\tau}_N)] = \sum_{|\sigma|+|\sigma^{-1}\pi|+|\pi^{-1}\gamma_k|=k-1} \#(\pi^{-1}\gamma_k) \mathbb{E}[M_\sigma(\mathfrak{m}_N)] \mu_k(\sigma^{-1}\pi) + O\left(\frac{1}{N}\right). \tag{3.10}$$

Using the isomorphism explained in Section 2.3, we may rewrite (3.10) in terms of non-crossing partitions:

$$\mathbb{E}[\mathrm{M}_k(\widehat{\tau}_N)] = \sum_{\nu \leq \rho \in \mathrm{NC}(k)} |K(\rho)|\, \mathbb{E}[\mathrm{M}_\nu(\mathfrak{m}_N)]\tilde{\mu}_k(\nu, \rho) + O\left(\frac{1}{N}\right)$$

$$= \sum_{\rho \in \mathrm{NC}(k)} |K(\rho)|\, \mathbb{E}[\mathrm{R}_\rho(\mathfrak{m}_N)] + O\left(\frac{1}{N}\right), \tag{3.11}$$

where the cumulant–moment formula (2.4) was used in the last line. Combining (3.11) and Theorem 1.3 implies the desired conclusion.

(ii) $\ell = 2$. Taking the expectation of $\mathrm{M}_k(\widehat{\tau}_N)^2 = (\mathrm{Tr}[D^k] - \mathrm{Tr}[(DUPU^*)^k])^2$ with Weingarten calculus yields

$$\mathbb{E}[\mathrm{M}_k(\widehat{\tau}_N)^2]$$
$$= \mathbb{E} \circ \mathrm{Tr}_{\gamma_k^{(1)}\gamma_k^{(2)}}[(DUIU^*)^k, (DUIU^*)^k] - \mathbb{E} \circ \mathrm{Tr}_{\gamma_k^{(1)}\gamma_k^{(2)}}[(DUIU^*)^k, (DUPU^*)^k]$$
$$\quad - \mathbb{E} \circ \mathrm{Tr}_{\gamma_k^{(1)}\gamma_k^{(2)}}[(DUPU^*)^k, (DUIU^*)^k] + \mathbb{E} \circ \mathrm{Tr}_{\gamma_k^{(1)}\gamma_k^{(2)}}[(DUPU^*)^k, (DUPU^*)^k]$$
$$= \sum_{\sigma, \pi \in \mathfrak{S}_{2k}} N^{\#(\sigma)} \mathbb{E}[\mathrm{M}_\sigma(\mathfrak{m}_N)]\mathcal{T}_k(\pi)\, \mathrm{Wg}(\sigma^{-1}\pi) \tag{3.12}$$

where

$$\mathcal{T}_k(\pi) := \mathrm{Tr}_{\pi^{-1}\gamma_k^{(1)}\gamma_k^{(2)}}[I^k, I^k] - \mathrm{Tr}_{\pi^{-1}\gamma_k^{(1)}\gamma_k^{(2)}}[I^k, P^k] - \mathrm{Tr}_{\pi^{-1}\gamma_k^{(1)}\gamma_k^{(2)}}[P^k, I^k] + \mathrm{Tr}_{\pi^{-1}\gamma_k^{(1)}\gamma_k^{(2)}}[P^k, P^k].$$

Note that, for readability, we use the abbreviation $\mathrm{Tr}_\sigma[A^k, B^k] = \mathrm{Tr}_\sigma[A_1, \ldots, A_k, B_1, \ldots, B_k]$ when $A = A_1 = \cdots = A_k$ and $B = B_1 = \cdots = B_k$.

By using the evident decomposition $I = P + Q$ with $Q = \mathrm{diag}(0, \ldots, 0, 1)$, we have the following expansion

$$\mathrm{Tr}_{\pi^{-1}\gamma_k^{(1)}\gamma_k^{(2)}}[I^k, I^k]$$
$$= \mathrm{Tr}_{\pi^{-1}\gamma_k^{(1)}\gamma_k^{(2)}}[I^k, I^{k-1}, P+Q]$$
$$= \mathrm{Tr}_{\pi^{-1}\gamma_k^{(1)}\gamma_k^{(2)}}[I^k, I^{k-1}, P] + \mathrm{Tr}_{\pi^{-1}\gamma_k^{(1)}\gamma_k^{(2)}}[I^k, I^{k-1}, Q]$$
$$= \mathrm{Tr}_{\pi^{-1}\gamma_k^{(1)}\gamma_k^{(2)}}[I^k, I^{k-2}, P+Q, P] + \mathrm{Tr}_{\pi^{-1}\gamma_k^{(1)}\gamma_k^{(2)}}[I^k, I^{k-1}, Q]$$
$$= \mathrm{Tr}_{\pi^{-1}\gamma_k^{(1)}\gamma_k^{(2)}}[I^k, I^{k-2}, P, P] + \mathrm{Tr}_{\pi^{-1}\gamma_k^{(1)}\gamma_k^{(2)}}[I^k, I^{k-2}, Q, P] + \mathrm{Tr}_{\pi^{-1}\gamma_k^{(1)}\gamma_k^{(2)}}[I^k, I^{k-1}, Q]$$
$$\cdots$$
$$= \mathrm{Tr}_{\pi^{-1}\gamma_k^{(1)}\gamma_k^{(2)}}[I^k, P^k] + \sum_{j=1}^{k} \mathrm{Tr}_{\pi^{-1}\gamma_k^{(1)}\gamma_k^{(2)}}[I^k, I^{k-j}, Q, P^{j-1}]. \tag{3.13}$$

In the same way,

$$\mathrm{Tr}_{\pi^{-1}\gamma_k^{(1)}\gamma_k^{(2)}}[P^k, I^k] = \mathrm{Tr}_{\pi^{-1}\gamma_k^{(1)}\gamma_k^{(2)}}[P^k, P^k] + \sum_{j=1}^{k} \mathrm{Tr}_{\pi^{-1}\gamma_k^{(1)}\gamma_k^{(2)}}[P^k, I^{k-j}, Q, P^{j-1}]. \tag{3.14}$$

Combining (3.13) and (3.14) together we get

$$\mathcal{T}_k(\pi) = \sum_{j=1}^{k} \left(\mathrm{Tr}_{\pi^{-1}\gamma_k^{(1)}\gamma_k^{(2)}}[I^k, I^{k-j}, Q, P^{j-1}] - \mathrm{Tr}_{\pi^{-1}\gamma_k^{(1)}\gamma_k^{(2)}}[P^k, I^{k-j}, Q, P^{j-1}] \right).$$

Again, a similar argument yields

$$\mathcal{T}_k(\pi) = \sum_{i=1}^{k} \sum_{j=1}^{k} \text{Tr}_{\pi^{-1}\gamma_k^{(1)}\gamma_k^{(2)}}[I^{k-i}, Q, P^{i-1}, I^{k-j}, Q, P^{j-1}].$$

When we decompose $\pi^{-1}\gamma_k^{(1)}\gamma_k^{(2)}$ into cycles, the contribution of the cycle which contains Q is at most 1 in $\text{Tr}_{\pi^{-1}\gamma_k^{(1)}\gamma_k^{(2)}}[I^{k-i}, Q, P^{i-1}, I^{k-j}, Q, P^{j-1}]$. Therefore, we get the upper bound

$$\mathcal{T}_k(\pi) = O\left(N^{\#(\pi^{-1}\gamma_k^{(1)}\gamma_k^{(2)})-1}\right) = O\left(N^{2k-|\pi^{-1}\gamma_k^{(1)}\gamma_k^{(2)}|-1}\right). \tag{3.15}$$

Here we also use the asymptotic expansion (2.2) of the Weingarten functions and two elementary facts about the length functions in symmetric groups $\mathfrak{S}_{2k}$: $|\sigma| + |\sigma^{-1}\pi| + |\pi^{-1}\gamma_k^{(1)}\gamma_k^{(2)}| \geq 2k-2$ and $|\sigma| + |\sigma^{-1}\pi| + |\pi^{-1}\gamma_k^{(1)}\gamma_k^{(2)}| \neq 2k-1$ since $|\sigma| + |\sigma^{-1}\pi| + |\pi^{-1}\gamma_k^{(1)}\gamma_k^{(2)}| \equiv |\gamma_k^{(1)}\gamma_k^{(2)}| \equiv 2k-2$ (mod 2) by using the length property (2.5). Applying those facts and (3.15) to (3.12) reveals that

$$\mathbb{E}[\text{M}_k(\widehat{\tau}_N)^2] = \sum_{\substack{\sigma,\pi\in\mathfrak{S}_{2k} \\ |\sigma|+|\sigma^{-1}\pi|+|\pi^{-1}\gamma_k^{(1)}\gamma_k^{(2)}|=2k-2}} N^{-|\sigma|-|\sigma^{-1}\pi|}\mathbb{E}[\text{M}_\sigma(\mathfrak{m}_N)]\mathcal{T}_k(\pi)\mu_{2k}(\sigma^{-1}\pi) + O\left(\frac{1}{N}\right).$$

By using the isomorphism (2.6), the last expression can be rewritten in terms of non-crossing partitions:

$$\mathbb{E}[\text{M}_k(\widehat{\tau}_N)^2] = \sum_{\substack{\nu,\rho\in\text{NC}(2k) \\ \nu\leq\rho\leq(1_k^{(1)},1_k^{(2)})}} \mathbb{E}[\text{M}_\nu(\mathfrak{m}_N)]\tilde{\mu}_{2k}(\nu,\rho)\frac{\mathcal{T}_k(\mathcal{P}_\rho)}{N^{|K(\rho)|-1}} + O\left(\frac{1}{N}\right)$$

$$= \sum_{\substack{\rho\in\text{NC}(2k) \\ \rho\leq(1_k^{(1)},1_k^{(2)})}} \mathbb{E}[\text{R}_\rho(\mathfrak{m}_N)]\frac{\mathcal{T}_k(\mathcal{P}_\rho)}{N^{|K(\rho)|-1}} + O\left(\frac{1}{N}\right)$$

$$= \sum_{\rho_1,\rho_2\in\text{NC}(k)} \mathbb{E}[\text{R}_{\rho_1}(\mathfrak{m}_N)\,\text{R}_{\rho_2}(\mathfrak{m}_N)]\frac{\mathcal{T}_k(\mathcal{P}_{(\rho_1,\rho_2)})}{N^{|K(\rho_1)|+|K(\rho_2)|-2}} + O\left(\frac{1}{N}\right).$$

Note that

$$\mathcal{T}_k(\mathcal{P}_{(\rho_1,\rho_2)}) = \sum_{i=1}^{k}\sum_{j=1}^{k} \text{Tr}_{(K(\rho_1),K(\rho_2))}[I^{k-i}, Q, P^{i-1}, I^{k-j}, Q, P^{j-1}]$$

$$= N^{|K(\rho_1)|+|K(\rho_2)|-2}|K(\rho_1)||K(\rho_2)| + O\left(N^{|K(\rho_1)|+|K(\rho_2)|-3}\right). \tag{3.16}$$

This is because the contribution of a cycle is 0 if it contains both P and Q, and is 1 if it contains Q and no P; from those observations, the main contributions appear when both Q's are at Kreweras points of ρ_1 and ρ_2, respectively, and so (3.16) follows. Hence we arrive at the formula

$$\mathbb{E}[\text{M}_k(\widehat{\tau}_N)^2] = \sum_{\rho_1,\rho_2\in\text{NC}(k)} |K(\rho_1)||K(\rho_2)|\,\mathbb{E}[\text{R}_{\rho_1}(\mathfrak{m}_N)\text{R}_{\rho_2}(\mathfrak{m}_N)] + O\left(\frac{1}{N}\right).$$

Applying Theorem 1.3 to the RHS finishes the proof.   □

Remark 3.9. Note that the calculations for $\ell = 1$ are similar to those in Nica and Speicher (2006, pp.379-393) where asymptotic freeness is proved for matrices rotated by independent Haar unitaries.

Proof of Theorem 1.1: According to Theorem 1.3, $\text{M}_k(\tau_N)^\ell$ is a polynomial on $\{\text{M}_n(\mathfrak{m}_N)\}_{n\in\mathbb{N}}$, so that Proposition A.4 allows us to pass to the limit:

$$\lim_{N\to\infty} \mathbb{E}[\text{M}_k(\tau_N)] = \text{M}_k(\tau) \quad \text{and} \quad \lim_{N\to\infty} \mathbb{E}[\text{M}_k(\tau_N)^2] = \text{M}_k(\tau)^2, \qquad k\in\mathbb{N}.$$

Combining the above and Theorem 3.7 yields that

$$\lim_{N\to\infty} \mathbb{E}[\mathrm{M}_k(\widehat{\tau}_N)] = \mathrm{M}_k(\tau) \quad \text{and} \quad \lim_{N\to\infty} \mathbb{E}[\mathrm{M}_k(\widehat{\tau}_N)^2] = \mathrm{M}_k(\tau)^2, \qquad k \in \mathbb{N},$$

which readily implies $\|\mathrm{M}_k(\widehat{\tau}_N) - \mathrm{M}_k(\tau)\|_{L^2} \to 0$. In particular, $\mathrm{M}_k(\widehat{\tau}_N)$ converges to $\mathrm{M}_k(\tau)$ in probability for every $k \in \mathbb{N}$. Since $\mathrm{M}_k(\widehat{\mathfrak{m}}_N)$ and $\mathrm{M}_k(\mathfrak{m})$ are respectively expressed by a common polynomial evaluated at $\{\mathrm{M}_k(\widehat{\tau}_N)\}_{k\geq 1}$ and $\{\mathrm{M}_k(\tau)\}_{k\geq 1}$, it follows that $\mathrm{M}_k(\widehat{\mathfrak{m}}_N)$ converges to $\mathrm{M}_k(\mathfrak{m})$ in probability. Finally, if the moment problem for $\{\mathrm{M}_k(\mathfrak{m})\}_{k\geq 1}$ is determinate then we conclude that $\widehat{\mathfrak{m}}_N$ weakly converges to $\mathfrak{m}$ in probability by Proposition A.1. $\qquad\square$

Appendix A. Convergence of random probability measures

Some results on the moment method for random measures are collected below. The proofs are basic. Let $\mathfrak{p}, \mathfrak{p}_n, n \in \mathbb{N}$, be random probability measures on $\mathbb{R}$ with an underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ below.

Proposition A.1. *Suppose that $\mathfrak{p}_n, \mathfrak{p}, n \in \mathbb{N}$ have finite moments of all orders almost surely, and the moment problem for $\{\mathrm{M}_k(\mathfrak{p})\}_{k\geq 1}$ is determinate almost surely. If*

$$\lim_{n\to\infty} \mathbb{P}[|\mathrm{M}_k(\mathfrak{p}_n) - \mathrm{M}_k(\mathfrak{p})| \geq \epsilon] = 0, \qquad k \in \mathbb{N}, \ \epsilon > 0, \tag{A.1}$$

then $\mathfrak{p}_n$ weakly converges to $\mathfrak{p}$ in probability:

$$\lim_{n\to\infty} \mathbb{P}\left[\left|\int_{\mathbb{R}} f(x)\, d\mathfrak{p}_n(x) - \int_{\mathbb{R}} f(x)\, d\mathfrak{p}(x)\right| \geq \epsilon\right] = 0, \qquad f \in C_b(\mathbb{R}), \ \epsilon > 0. \tag{A.2}$$

Proof: For later use, we first verify the existence of a subsequence of $\{\mathfrak{p}_n\}_{n\geq 1}$ which weakly converges to $\mathfrak{p}$ almost surely. Let $\Omega_0 \in \mathcal{F}$ be such that $\mathbb{P}[\Omega_0] = 1$ and the moment problem for $\{\mathrm{M}_k(\mathfrak{p}^\omega)\}_{k\geq 1}$ is determinate for all $\omega \in \Omega_0$. For $k = 1$, there exists a subsequence $\{n(1,\ell)\}_{\ell=1}^\infty$ of $\mathbb{N}$ and $\Omega_1 \subset \Omega_0$ such that $\Omega_1 \in \mathcal{F}, \mathbb{P}[\Omega_1] = 1$ and $\mathrm{M}_1(\mathfrak{p}^\omega_{n(1,\ell)})$ converges to $\mathrm{M}_1(\mathfrak{p}^\omega)$ for all $\omega \in \Omega_1$. For $k = 2$, there exists a subsequence $\{n(2,\ell)\}_{\ell=1}^\infty$ of $\{n(1,\ell)\}_{\ell=1}^\infty$ and $\Omega_2 \subset \Omega_1$ such that $\Omega_2 \in \mathcal{F}, \mathbb{P}[\Omega_2] = 1$ and $\mathrm{M}_2(\mathfrak{p}^\omega_{n(2,\ell)})$ converges to $\mathrm{M}_2(\mathfrak{p}^\omega)$ for all $\omega \in \Omega_2$. In this way we obtain subsequences $\{n(k,\ell)\}_{\ell=1}^\infty$ and decreasing subsets Ω_k of probability one for $k \geq 1$. Define $\tilde{\Omega} := \cap_{k\geq 1}\Omega_k$ and $n(\ell) := n(\ell,\ell)$; then $\mathrm{M}_k(\mathfrak{p}^\omega_{n(\ell)})$ converges to $\mathrm{M}_k(\mathfrak{p}^\omega)$ as $\ell \to \infty$ for all $\omega \in \tilde{\Omega}$ and all $k \in \mathbb{N}$. Since the moment problem for the limit sequence is determinate, we conclude by Chung (1968, Theorem 4.5.5) that $\mathfrak{p}^\omega_{n(\ell)}$ weakly converges to $\mathfrak{p}^\omega$ as $\ell \to \infty$ for all $\omega \in \tilde{\Omega}$.

To finish the proof, suppose to the contrary that the desired conclusion (A.2) is false: there exist $f \in C_b(\mathbb{R})$, $\epsilon, \delta > 0$ and a subsequence of $\{\mathfrak{p}_n\}_{n\geq 1}$, denoted by $\{\mathfrak{p}_{n'}\}$, such that for all n'

$$\mathbb{P}\left[\left|\int_{\mathbb{R}} f(x)\, d\mathfrak{p}_{n'}(x) - \int_{\mathbb{R}} f(x)\, d\mathfrak{p}(x)\right| \geq \epsilon\right] \geq \delta. \tag{A.3}$$

However, we can extract a further subsequence of $\{\mathfrak{p}_{n'}\}$ which weakly converges to $\mathfrak{p}$ almost surely as we discussed. For this subsequence, the LHS of (A.3) must tend to zero, a contradiction. $\qquad\square$

Remark A.2. A similar result and proof are found in Grenander (1963, p. 178–180).

Proposition A.3. *Suppose that*

$$\sup_{n\geq 1} \mathbb{E}[\mathrm{M}_k(\mathfrak{p}_n)] < \infty \quad \text{and} \quad \mathbb{E}[\mathrm{M}_k(\mathfrak{p})] < \infty, \qquad k \in 2\mathbb{N}. \tag{A.4}$$

Then the condition (A.1) is equivalent to

$$\|\mathrm{M}_k(\mathfrak{p}_n) - \mathrm{M}_k(\mathfrak{p})\|_{L^p} \to 0, \qquad p \in [1,\infty), \ k \in \mathbb{N}. \tag{A.5}$$

Proof: It suffices to prove that (A.1) implies (A.5); the other direction is well known.

For $p \in [1, \infty)$ choose $\ell \in 2\mathbb{N}$ such that $\ell \geq p$. The Hölder inequality implies that $|\mathrm{M}_k(\mathfrak{p}_n)|^{\ell} \leq \mathrm{M}_{k\ell}(\mathfrak{p}_n)$ and hence

$$\| \mathrm{M}_k(\mathfrak{p}_n)\|_{L^p} \leq \| \mathrm{M}_k(\mathfrak{p}_n)\|_{L^{\ell}} \leq (\mathbb{E}[\mathrm{M}_{k\ell}(\mathfrak{p}_n)])^{\frac{1}{\ell}}. \tag{A.6}$$

Combining the above and (A.4), as well as similar inequalities for $\mathrm{M}_k(\mathfrak{p})$, yields that

$$\sup_{n \in \mathbb{N}} \| \mathrm{M}_k(\mathfrak{p}_n)\|_{L^p} < \infty \quad \text{and} \quad \| \mathrm{M}_k(\mathfrak{p})\|_{L^p} < \infty, \qquad k \in \mathbb{N}, \ p \in [1, \infty). \tag{A.7}$$

By standard arguments we obtain

$$\begin{aligned}
\| \mathrm{M}_k(\mathfrak{p}_n) - \mathrm{M}_k(\mathfrak{p})\|_{L^p}^p &= \mathbb{E}[|\mathrm{M}_k(\mathfrak{p}_n) - \mathrm{M}_k(\mathfrak{p})|^p 1_{\{|\mathrm{M}_k(\mathfrak{p}_n) - \mathrm{M}_k(\mathfrak{p})| \geq \epsilon\}}] \\
&\quad + \mathbb{E}[|\mathrm{M}_k(\mathfrak{p}_n) - \mathrm{M}_k(\mathfrak{p})|^p 1_{\{|\mathrm{M}_k(\mathfrak{p}_n) - \mathrm{M}_k(\mathfrak{p})| < \epsilon\}}] \\
&\leq (\mathbb{E}[|\mathrm{M}_k(\mathfrak{p}_n) - \mathrm{M}_k(\mathfrak{p})|^{2p}])^{1/2} (\mathbb{P}[|\mathrm{M}_k(\mathfrak{p}_n) - \mathrm{M}_k(\mathfrak{p})| \geq \epsilon])^{1/2} + \epsilon^p \\
&\leq (\| \mathrm{M}_k(\mathfrak{p}_n)\|_{L^{2p}} + \| \mathrm{M}_k(\mathfrak{p})\|_{L^{2p}})^p (\mathbb{P}[|\mathrm{M}_k(\mathfrak{p}_n) - \mathrm{M}_k(\mathfrak{p})| \geq \epsilon])^{1/2} + \epsilon^p.
\end{aligned}$$

Applying (A.7) and (A.1) to the above finishes the proof. $\qquad\square$

Proposition A.4. *Suppose that* (A.1) *and* (A.4) *hold. Then*

$$\lim_{n \to \infty} \mathbb{E}[P(\mathrm{M}_1(\mathfrak{p}_n), \mathrm{M}_2(\mathfrak{p}_n), \dots, \mathrm{M}_k(\mathfrak{p}_n))] = \mathbb{E}[P(\mathrm{M}_1(\mathfrak{p}), \mathrm{M}_2(\mathfrak{p}), \dots, \mathrm{M}_k(\mathfrak{p}))]$$

for every $k \in \mathbb{N}$ and every polynomial $P \in \mathbb{C}[x_1, x_2, \dots, x_k]$.

Proof: This is a consequence of Proposition A.3 and the following standard fact: if random variables $Y, Z, Y_n, Z_n, n \in \mathbb{N}$ satisfy $Y_n \to Y$ in L^p and $Z_n \to Z$ in L^p for all $p \in [1, \infty)$, then $Y_n Z_n \to YZ$ in L^p for all $p \in [1, \infty)$. $\qquad\square$

Acknowledgements

References

Aronszajn, N. and Donoghue, W. F. A supplement to the paper on exponential representations of analytic functions in the upper half-plane with positive imaginary part. *J. Analyse Math.*, **12**, 113–127 (1964). MR168769.

Baryshnikov, Y. GUEs and queues. *Probab. Theory Related Fields*, **119** (2), 256–274 (2001). MR1818248.

Biane, P. Representations of symmetric groups and free probability. *Adv. Math.*, **138** (1), 126–181 (1998). MR1644993.

Biane, P. Approximate factorization and concentration for characters of symmetric groups. *Internat. Math. Res. Notices*, **2001** (4), 179–192 (2001). MR1813797.

Bufetov, A. Kerov's interlacing sequences and random matrices. *J. Math. Phys.*, **54** (11), 113302, 10 (2013). MR3137038.

Chung, K. L. *A course in probability theory.* Harcourt, Brace & World, Inc., New York (1968). MR0229268.

Collins, B. and Male, C. The strong asymptotic freeness of Haar and deterministic matrices. *Ann. Sci. Éc. Norm. Supér. (4)*, **47** (1), 147–163 (2014). MR3205602.

Collins, B. and Matsumoto, S. Weingarten calculus via orthogonality relations: new applications. *ALEA Lat. Am. J. Probab. Math. Stat.*, **14** (1), 631–656 (2017). MR3680193.

Collins, B. and Śniady, P. Integration with respect to the Haar measure on unitary, orthogonal and symplectic group. *Comm. Math. Phys.*, **264** (3), 773–795 (2006). MR2217291.

Erdős, L. and Schröder, D. Fluctuations of rectangular Young diagrams of interlacing Wigner eigenvalues. *Int. Math. Res. Not. IMRN*, **2018** (10), 3255–3298 (2018). MR3805203.

Faraut, J. Rayleigh theorem, projection of orbital measures and spline functions. *Adv. Pure Appl. Math.*, **6** (4), 261–283 (2015). MR3403440.

Franz, U., Hasebe, T., and Schleißinger, S. Monotone increment processes, classical Markov processes, and Loewner chains. *Dissertationes Math.*, **552**, 119 (2020). MR4152669.

Goel, G. and Yao, A. A quantized analogue of the Markov-Krein correspondence. *ArXiv Mathematics e-prints* (2020). arXiv: 2011.10724.

Grenander, U. *Probabilities on algebraic structures.* John Wiley & Sons, Inc., New York-London; Almqvist & Wiksell, Stockholm-Göteborg-Uppsala (1963). MR0206994.

Hiai, F. and Petz, D. *The semicircle law, free random variables and entropy*, volume 77 of *Mathematical Surveys and Monographs.* American Mathematical Society, Providence, RI (2000). ISBN 0-8218-2081-8. MR1746976.

Kerov, S. Interlacing measures. In *Kirillov's seminar on representation theory*, volume 181 of *Amer. Math. Soc. Transl. Ser. 2*, pp. 35–83. Amer. Math. Soc., Providence, RI (1998). MR1618739.

Kerov, S. V. Asymptotics of the separation of roots of orthogonal polynomials. *Algebra i Analiz*, **5** (5), 68–86 (1993). MR1263315.

Lassalle, M. Jack polynomials and free cumulants. *Adv. Math.*, **222** (6), 2227–2269 (2009). MR2562783.

Logan, B. F. and Shepp, L. A. A variational problem for random Young tableaux. *Advances in Math.*, **26** (2), 206–222 (1977). MR1417317.

Mergny, P. and Potters, M. Rank one HCIZ at high temperature: interpolating between classical and free convolutions. *ArXiv Mathematics e-prints* (2021). arXiv: 2101.01810.

Nica, A. and Speicher, R. *Lectures on the combinatorics of free probability*, volume 335 of *London Mathematical Society Lecture Note Series.* Cambridge University Press, Cambridge (2006). ISBN 978-0-521-85852-6; 0-521-85852-6. MR2266879.

Sodin, S. Fluctuations of interlacing sequences. *Zh. Mat. Fiz. Anal. Geom.*, **13** (4), 364–401 (2017). MR3733197.

Tao, T. *Topics in random matrix theory*, volume 132 of *Graduate Studies in Mathematics.* American Mathematical Society, Providence, RI (2012). ISBN 978-0-8218-7430-1. MR2906465.

Veršik, A. M. and Kerov, S. V. Asymptotic behavior of the Plancherel measure of the symmetric group and the limit form of Young tableaux. *Dokl. Akad. Nauk SSSR*, **233** (6), 1024–1027 (1977). MR0480398.