Probab. Theory Relat. Fields 124, 574–594 (2002)
Digital Object Identifier (DOI) 10.1007/10.1007/s00440-002-0225-3

Christian Mazza · Didier Piau

Products of correlated symmetric matrices and q-Catalan numbers

Received: 15 February 2002 / Revised version: 25 May 2002 /
Published online: 24 October 2002 –

Abstract. The well known convergence of the spectrum of large random symmetric matrices, due to Wigner, holds for products of correlated symmetric matrices with general entries. The limiting moments coincide with weighted enumeration of permutations, or of rooted trees. When the correlations are Markovian, the limiting first moments are closely related to Carlitz-Riordan q-Catalan numbers. As a consequence, these moments asymptotically exhibit a phase transition, with respect to the correlation coefficient. The critical correlations can be computed as the least positive zero of q-hypergeometric functions. Similar methods allow to recover some results due to Logan, Mazo, Odlyzko and Shepp.

1. Introduction

Wigner (1955, 1957, 1958) proved that the spectral measure of a wide class of symmetric random matrices of dimension N converges, in the $N \to \infty$ limit, to the semicircle law. Since then, this observation has been extended to numerous classes of matrices, and often completed by central limit theorems. More recently, large deviations principles have been derived, see Ben Arous and Guionnet (1997) and Guionnet and Zeitouni (2000). The literature on the subject is too rapidly expanding to allow for completeness, and we advise the interested reader to consult the two above-mentioned papers for references.

Broadly speaking, two strategies are used to study random large matrices. Either one starts from an explicit formula that gives the distribution of the spectrum, if one is available, as for example in Gaussian settings. Otherwise, one has to go back to Wigner's original technique and enumerate the terms of some relevant expansions. This paper falls in the latter category and proves that an analogue of Wigner's result holds for some products of, possibly non Gaussian, correlated symmetric matrices. In the definition below, the laws of the entries $\gamma_{i,j}$ are not required to be Gaussian or symmetric.

Definition 1. *The N-dimensional random matrices $\Gamma_N := (\gamma_{i,j})_{1 \le i,j \le N}$ are called reduced Wigner matrices if the following holds:*

Ch. Mazza, D. Piau: Université Claude Bernard Lyon–I, LaPCS – Domaine de Gerland, 50, avenue Tony-Garnier, 69366 Lyon Cedex 07 France.
e-mail: `Christian.Mazza@univ-lyon1.fr`; `Didier.Piau@univ-lyon1.fr`,
web: `http://lapcs.univ-lyon1.fr`

Mathematics Subject Classification (2000): Primary: 15A52; Secondary: 05A30, 15A18, 60F05, 82B44

- *Each Γ_N is symmetric, that is, $\gamma_{i,j} = \gamma_{j,i}$.*
- *For $i \le j$, the random variables $\gamma_{i,j}$ are independent and centered.*
- *For $i \ne j$, $\mathbb{E}(\gamma_{i,j}^2) = 1$.*
- *For any $n \ge 2$, $\mathbb{E}(|\gamma_{i,j}|^n) \le c_n$, where c_n is independent of $i \le j$.*

The sequence $\Gamma_N(k) := (\gamma_{i,j}(k))_{1\le i,j\le N}$ of N-dimensional random matrices, indexed by $k \ge 1$, is called a reduced Wigner process of correlation ρ, with $|\rho| \le 1$, if the following holds:

- *Each $\Gamma_N(k)$ is a reduced Wigner matrix, and the sequence $(c_n)_n$ of the last point above is uniform in k.*
- *For $i \le j$, each process $(\gamma_{i,j}(k))_k$ is independent of the others.*
- *For $i \ne j$, the process $(\gamma_{i,j}(k))_k$ is ρ–correlated, that is, for any $k \ge m$,*

$$\mathbb{E}(\gamma_{i,j}(k)\gamma_{i,j}(m)) := \rho^{k-m}. \tag{1}$$

A Wigner matrix (or process) of variance σ^2 is defined as being σ times a reduced Wigner matrix (or process).

Remark 2. If $\rho = 0$, $(\Gamma_N(k))_k$ is an i.i.d. sequence of Wigner matrices. If $\rho = 1$, the matrices $\Gamma_N(k)$ are all equal to the same Wigner matrix $\Gamma_N(1)$. For any $|\rho| \le 1$, the correlation structure (1) entails a Markovian dependence, since one can realize a ρ–correlated Wigner process Γ_N from an i.i.d. sequence $(\Delta_N(k))_k$ of Wigner matrices, as $\Gamma_N(1) := \Delta_N(1)$ and, for $k \ge 1$,

$$\Gamma_N(k+1) := \rho\,\Gamma_N(k) + (1-\rho^2)^{1/2}\Delta_N(k+1).$$

However, see Remark 6 in Section 2.2 for more general correlations. □

The normalization of the entries in Definition 1 implies that Wigner matrices scale like $N^{1/2}$. Introduce the products of reduced Wigner matrices

$$Q_k^N := N^{-k/2} \prod_{m=1}^{k} \Gamma_N(m),$$

and their spectral measure

$$\mu_k^N := N^{-1} \sum_{\lambda} \delta_\lambda,$$

where the sum is over the N eigenvalues λ of Q_k^N. The first order behaviour of $(Q_k^N)^\ell$ is described by its mean normalized trace, that is,

$$N^{-1}\,\mathbb{E}(\mathrm{tr}(Q_k^N)^\ell) =: B_{k,\ell}^N(\rho).$$

One sets $B_k^N(\rho) := B_{k,1}^N(\rho)$ and $\mu_N := \mu_1^N$. The case $\rho = 1$ is Wigner's case since then, $Q_k^N = N^{-k/2}\,\Gamma_N(1)^k$ and $\Gamma_N(1)$ is a Wigner matrix. Thus, $B_{k,\ell}^N(1)$ is the mean of the $(k\ell)$th moment of μ_N. Wigner solved this case, showing that the numbers $B_k^N(1)$ converge to the moments of the now famous semicircle distribution

$$\nu_1(\mathrm{d}w) := (2\pi)^{-1}(4-w^2)^{1/2}\mathbf{1}_{|w|\le 2}\,\mathrm{d}w, \tag{2}$$

whose odd moments are zero and even moments are the Catalan numbers

$$C_k := \frac{(2k)!}{k!\,(k+1)!}. \tag{3}$$

Wigner assumed that the random variables $\gamma_{i,j}$ have symmetric laws, but this assumption is not necessary.

When the reduced Wigner process $(\Gamma_N(k))_k$ is Gaussian and under the additional assumption that $\mathbb{E}(\gamma_{i,i}(k)^2) = 2$, Khorunzhy (2001) proved in a recent preprint that $B_k^N(\rho)$ converges for any $\rho \in [0; 1]$, when $N \to \infty$, and he proposed a combinatorial interpretation of the nonzero limits $B_{2k}(\rho)$.

On the other hand, Logan, Mazo, Odlyzko and Shepp (1983), denoted by LMOS below, studied a similar problem, in dimension $N = 1$ and in the Gaussian case, that we describe now. Let $(\gamma_k)_k$ denote a one-dimensional Gauss-Markov process, that is, a stationary centered Gaussian process, of covariance

$$\mathbb{E}(\gamma_k \gamma_m) := \sigma^2\, \rho^{k-m},$$

for any integers $k \geq m$, where $\sigma^2 > 0$ and $|\rho| \leq 1$. The polyspectrum of γ is defined as

$$\mathbb{E}(\gamma_1 \cdots \gamma_k) =: \sigma^k\, A_k(\rho).$$

Each $A_{2k+1}(\rho)$ is zero, since $\gamma_1 \cdots \gamma_{2k+1}$ is centered. LMOS motivate their study of the asymptotics of $A_{2k}(\rho)$ when $k \to \infty$, by the fact that $\gamma_1 \cdots \gamma_k$ is a toy model of the product of noncommuting correlated matrices, which arise in the analysis of learning curves for adaptive systems. LMOS write the generating series of $(A_{2k}(\rho))_k$ as a continued fraction. This representation yields the existence of a critical value ρ_A such that $A_{2k}(\rho) \to 0$ exponentially fast if $\rho \in [0; \rho_A)$ and $A_{2k}(\rho) \to \infty$ exponentially fast if $\rho \in (\rho_A; 1]$. Finally, LMOS estimate numerically

$$\rho_A \simeq .563007169.$$

2. Results

We study the $N \to \infty$ limit of the spectral measures μ_k^N. We prove that, for fixed $k \geq 1$, $\ell \geq 1$ and $|\rho| \leq 1$, the ℓth moment $B_{k,\ell}^N(\rho)$ of μ_k^N converges when $N \to \infty$ to a limit $B_{k,\ell}(\rho)$. This, and other considerations, imply that μ_k^N converges weakly, in probability, to a deterministic measure $\nu_{k,\rho}$, with bounded support. A new phenomenon due to the correlation structure arises, namely a phase transition for the limiting first moments $B_k(\rho) := B_{k,1}(\rho)$. For large values of ρ, $B_k(\rho) \to \infty$ when $k \to \infty$, for smaller values of ρ, $B_k(\rho) \to 0$.

Our methods are primarily combinatorial and they mimick Wigner's original one. As mentioned above, no Gaussianity or symmetry is required. About the hypotheses in Definition 1, see Remark 6 in Section 2.2. We write every limiting moment as a weighted enumeration of involutions or, alternatively, of rooted planar trees. In general, the generating functions considered are continued fractions, and in fact, the first moments are linked to well known combinatorial objects, called q-Catalan numbers. The next sections state our mathematical results.

2.1. Convergence of the first moments

We first give a direct proof of the convergence of $B_k^N(\rho)$ when $N \to \infty$, see Theorem 1 below, recovering Khorunzhy's result about the Gaussian case in a simpler way. (Additionally, we correct some misprints that mar Khorunzhy (2001).) We find that the limits $B_{2k+1}(\rho) = 0$ and that $B_{2k}(\rho)$ is related to some q-analogue $C_k(q)$ of the Catalan numbers C_k, introduced by Carlitz and Riordan (1964), through the simple formula

$$B_{2k}(\rho) = \rho^k \, C_k(q) \quad \text{for} \quad q := \rho^2. \tag{4}$$

To prove this, we use a general correspondence, systematized by Flajolet (1980), that links Stieltjes continued fractions and some enumerations of marked lattice paths. Thus, we exhibit some family of paths that $B_{2k}(\rho)$ enumerates. This is based on well known bijections between sets of permutations and sets of lattice paths, and yields immediately the generating functions of $B_{2k}(\rho)$ as an explicit continued fraction.

The same idea works in the one-dimensional case that LMOS established. We give a direct proof that the generating function of $(A_{2k}(\rho))_k$ is indeed the continued fraction that LMOS exhibited, see Equation (12) in Section 6. The paths that $A_{2k}(\rho)$ and $B_{2k}(\rho)$ enumerate are called Dyck paths by combinatoricians and Bernoulli excursions by probabilists, see Takacs (1991) for instance. Theorems 1, 2 and 3 below use definitions of Section 3.1, that we recall briefly in the statement of the theorems.

Theorem 1. *Let $\mathcal{J}(n)$ denote the set of the involutions of $\{1, 2, \dots, n\}$ with no fixed points and no crossing, $D(\sigma)$ the diameter of σ, $\mathcal{D}(n)$ the set of the Dyck paths of length n, and* $\mathrm{Area}(c)$ *the area under the Dyck path c. Then, when $N \to \infty$, $B_{2k+1}^N(\rho)$ converges to zero, and $B_{2k}^N(\rho)$ converges to*

$$B_{2k}(\rho) := \sum_{\sigma \in \mathcal{J}(2k)} \rho^{D(\sigma)} = \sum_{c \in \mathcal{D}(2k)} \rho^{\mathrm{Area}(c)}.$$

Recall that we do assume in dimension one the Gaussian distribution.

Theorem 2. *Let $\mathcal{I}(2k)$ denote the set of the involutions of $\{1, 2, \dots, 2k\}$ with no fixed points and, for $c \in \mathcal{D}(2k)$, denote* $\mathrm{desc}(c) := \{n;\ c_n = c_{n-1} - 1\}$. *Then,*

$$A_{2k}(\rho) = \sum_{\rho \in \mathcal{I}(2k)} \rho^{D(\sigma)} = \sum_{c \in \mathcal{D}(2k)} \rho^{\mathrm{Area}(c)} \prod_{n \in \mathrm{desc}(c)} c_{n-1}. \tag{5}$$

2.2. Convergence of the spectral measures

The method outlined above yields the convergence of $B_{k,\ell}^N$ as well, see Theorem 3. This establishes the full analogue of Wigner's result.

For any involution $\sigma \in \mathcal{I}(2n)$ with $n \geq 1$, denote by $D(k, \sigma)$ the diameter modulo k of σ, that is,

$$D(k, \sigma) := \frac{1}{2} \sum_{i=1}^{2n} |\{i\}_k - \{\sigma(i)\}_k|,$$

where $\{i\}_k \in \{1, 2, \dots, k\}$ and $\{i\}_k$ is congruent to i modulo k.

Theorem 3. *When $N \to \infty$, μ_k^N converges weakly, in probability, to a deterministic measure $\nu_{k,\rho}$. The ℓth moment $B_{k,\ell}(\rho)$ of $\nu_{k,\rho}$ is the limit of $B_{k,\ell}^N(\rho)$, and equals the following weighted enumeration of involutions:*

$$B_{k,\ell}(\rho) := \sum_{\sigma \in \mathcal{J}(k\ell)} \rho^{D(k,\sigma)}.$$

Remark 3. If $(\Gamma_N(k))_k$ is a Wigner process of correlation ρ, then the process $((-1)^k\, \Gamma_N(k))_k$ is Wigner of correlation $(-\rho)$. Hence,

$$B_{2k}(-\rho) = (-1)^k\, B_{2k}(\rho), \quad A_{2k}(-\rho) = (-1)^k\, A_{2k}(\rho).$$

Likewise, $\nu_{2k+1,-\rho} = \nu_{2k+1,\rho}$, and $\nu_{2k,-\rho}$ is the law of $(-1)^k$ times a random variable of law $\nu_{2k,\rho}$. Thus, one can, and we will, assume that $\rho \in [0; 1]$. □

Remark 4. Special cases of Theorem 3 are as follows. When $k\ell$ is odd, $\mathcal{J}(k\ell)$ is empty and $B_{k,\ell}(\rho) = 0$. Thus, the odd moments of $\nu_{2k+1,\rho}$ are zero, that is, $\nu_{2k+1,\rho}$ is symmetric with respect to the origin. When $\rho \geq 0$, all the moments of $\nu_{2k,\rho}$ are real positive. The support of $\nu_{k,\rho}$ is a subset of the disc $|w| \leq 2^k$. Assume that Wigner's law ν_1 defined by Equation (2) is the law of W. Then, $\nu_{1,\rho} = \nu_1$ and $\nu_{k,1}$ is the law of W^k (both results are consequences of Wigner's case), and $\nu_{2,\rho}$ is the law of $\rho\, W^2$ (see Remark 13 in Section 4.2). □

Remark 5. When $|\rho| < 1$, Q_k^N is not symmetric for $k \geq 2$. However, the crucial relation

$$\int w^\ell \, \mathrm{d}\mu(w) = N^{-1} \operatorname{tr}(Q)^\ell$$

is valid for any matrix Q of normalized spectral measure μ. To see this, use the triangular form $Q = PTP^{-1}$ of Q, where T is (upper) triangular and the diagonal elements of T are the eigenvalues of Q. □

Remark 6. Our proofs reveal that the hypotheses in Definition 1 can be relaxed. The law of $\gamma_{i,j}$ in Γ_N may vary with N, as long as $\mathbb{E}(\gamma_{i,j}^2) \to \sigma^2$ when $N \to \infty$, uniformly in i and j, and

$$\mathbb{E}(|\gamma_{i,j}|^k) \leq c_k\, N^{(k/2)-1}$$

for every $k \geq 2$, where c_k is uniform in N, i and j. Similar modifications are possible as regards Wigner processes.

One can also consider other correlation structures. Namely, replace (1) by the more general

$$\mathbb{E}(\gamma_{i,j}(k)\gamma_{i,j}(m)) := \sigma^2\, c(k,m),$$

where $c(\cdot,\cdot)$ is a correlation function. In particular, $c(k,k) = 1$, $|c(k,m)| \leq 1$ and $c(k,m) = c(m,k)$. Then, Theorem 1 holds with

$$B_{2k} := \sigma^{2k} \sum_{\sigma \in \mathcal{J}(2k)} \prod_i c(i, \sigma(i)),$$

and Theorem 3 holds with

$$B_{k,\ell} := \sigma^{k\ell} \sum_{\sigma \in \mathcal{J}(k\ell)} \prod_i c(\{i\}_k, \{\sigma(i)\}_k),$$

where each product runs over the integers i such that $i < \sigma(i)$. In the rest of the paper, we refrain from looking for the weakest possible hypotheses. □

Remark 7. The objects and the methods of this paper are clearly related to noncommutative probability, see Voiculescu, Dykema and Nica (1992) and Speicher (1990) for instance. We do not pursue this idea here. □

2.3. Asymptotics of the moments

To get some insight about the first order behaviour of $\nu_{k,\rho}$, we turn to the asymptotic behaviour of $(B_{2k}(\rho))_k$. A useful fact here is that Ramanujan wrote the generating function of $C_k(q)$ as the ratio of two q-hypergeometric functions, see Equations (8) and (9) in Section 3.3.

We prove that $(B_{2k})_{k\geq 0}$ exhibits a phase transition in the following sense: there exists a critical value $\rho_B \in (0; 1)$, such that

- $B_{2k}(\rho) \to 0$ exponentially fast for $\rho \in [0; \rho_B)$,
- $B_{2k}(\rho) \to \infty$ exponentially fast for $\rho \in (\rho_B; 1]$.

Theorem 4. *For any $\rho \geq 0$, $(B_{2k}(\rho))^{1/k}$ converges to $\beta(\rho) \geq 0$ as $k \to \infty$. If $\rho > 1$, $\beta(\rho) = +\infty$. On the interval* $[0; 1]$*, β is continuous and strictly increasing. Finally, $\beta(0) = 0$ and $\beta(1) = 4$.*

Corollary 8. *If $\sigma^2 \in [0; 1/4]$, $\sigma^{2k}\, B_{2k}(\rho) \to 0$ for every $\rho \in [0; 1]$.*
If $\sigma^2 > 1/4$, there exists a unique $\rho_B(\sigma^2) \in (0; 1)$, such that $\sigma^{2k}\, B_{2k}(\rho)$ converges to $+\infty$ if $\rho > \rho_B(\sigma^2)$, and to 0 if $\rho < \rho_B(\sigma^2)$.
The critical value $\rho = \rho_B(\sigma^2)$ solves

$$\sigma^2\, \beta(\rho) = 1.$$

For instance, $\rho_B := \rho_B(1) \in (.660; .683)$.

Further properties of β are in Section 5. In Section 3.3, we write $\beta = \beta(\rho)$ as the largest root of $F(\rho^2, \rho/\beta) = 0$ for a given q-hypergeometric function F. This reads

$$\sum_{k\geq 0} (-1)^k\, \rho^{2k^2-k}\, \beta^{-k}/(\rho^2)_k = 0.$$

From the implicit functions theorem for analytic functions, β is real analytic on $[0; 1)$. The critical value $\rho = \rho_B(\sigma^2)$ is the root of smallest modulus of the explicit q-series $F(\rho^2, \rho\sigma^2) = 0$. Very few terms of this q-series yield accurate numerical estimations, such as

$$\rho_B \simeq .662901485.$$

Remark 9. In terms of the diameter $e_{k,\rho}$ of the support of the measure $\nu_{k,\rho}$, one should have $e_{2k,\rho} \to 0$ for $\rho < \rho_B$ and $e_{2k,\rho} \to \infty$ for $\rho > \rho_B$. However, this supposes to interchange the limits $N \to \infty$ and $k \to \infty$, something we have not checked. □

Remark 10. The limit $\gamma(q)$ of $C_k(q)^{1/k}$ exhibits no phase transition as in Theorem 4, since, for instance, our proofs imply that

$$C_k(q) \geq (1+q)^{k-1}.$$

On the other hand, introduce the generating function $C(q, z)$ of the sequence $(C_k(q))$ (see Section 3.3). From Odlyzko and Wilf (1988), the coefficient of q^k in the expansion of $C(q, q)$ as a series of powers of q behaves like $(q_{\mathrm{OW}})^{-k}$, with

$$q_{\mathrm{OW}} \simeq .5761487699142756.$$

Thus, the radius of convergence of $C(q, q)$ is q_{OW}. In other words, putting together these results, using the monotonicity of $C_k(q)$ with respect to q and its submultiplicativity with respect to k, see Lemma 16 below, one gets that $(\rho_B)^2 \leq q_{\mathrm{OW}} < \rho_B$ and

- if $q < (\rho_B)^2$, then $C_k(q) \leq (\rho_B)^{-k} \leq q^{-k/2}$ and $\gamma(q) \leq \rho_B^{-1} \leq q^{-1/2}$,
- if $q < q_{\mathrm{OW}}$, then $C_k(q) \leq (q_{\mathrm{OW}})^{-k} \leq q^{-k}$ and $\gamma(q) \leq q_{\mathrm{OW}}^{-1} \leq q^{-1}$,
- if $q \leq 1$, then $C_k(q) \leq 4^k$ and $\gamma(q) \leq 4$.

Although q-Catalan numbers are classical combinatorial objects, such quantifications of their asymptotic behaviour do not seem customary. □

The situation in dimension one is quite similar. We know from LMOS that $(A_{2k})_k$ exhibits a phase transition. Theorem 5 includes this result.

Theorem 5. *For any $\rho \geq 0$, $(A_{2k}(\rho))^{1/k}$ converges to $\alpha(\rho) \geq 0$ as $k \to \infty$. If $\rho \geq 1$, $\alpha(\rho) = +\infty$. On the interval* $[0; 1)$*, α is continuous and strictly increasing. Finally, $\alpha(0) = 0$ and $\alpha(1-) = +\infty$.*
For any $\sigma^2 \neq 0$, there exists a unique $\rho_A(\sigma^2) \in$ $(0; 1)$*, such that $\sigma^{2k} A_{2k}(\rho)$ converges to $+\infty$ if $\rho > \rho_A(\sigma^2)$, and to 0 if $\rho < \rho_A(\sigma^2)$. The critical value $\rho = \rho_A(\sigma^2)$ solves*

$$\sigma^2 \alpha(\rho) = 1.$$

For instance, $\rho_A := \rho_A(1) \in$ $(.543; .619)$.

We conjecture that α is real analytic on $[0; 1)$. The fact that $\alpha(\rho)$ is infinite for $\rho \geq 1$ comes from the exact expression

$$A_{2k}(1) = \frac{(2k)!}{2^k\, k!}.$$

Theorem 5 provides the behaviour of $\mathbb{E}(\gamma_1 \cdots \gamma_{2k})$ for any σ^2. Recall that, for $\sigma^2 = 1$, LMOS estimate numerically

$$\rho_A \simeq .563007169.$$

We have not been able to treat the continued fraction of LMOS, see Equation (12) in Section 6, like the generating series of $B_{2k}(\rho)$, see Equation (8) in Section 3.3, so as to get its exact radius of convergence. In this respect, and somewhat ironically, the $N \to \infty$ limit is easier to handle than the one-dimensional case that LMOS had set out to solve as "a valuable guide to more realistic situations."

However, and in a somewhat different direction, we provide in Sections 5 and 6 explicit geometric bounds of $A_{2k}(\rho)$ and $C_k(q)$, valid for every finite $k \geq 1$. This yields direct proofs of some numerical values stated in our theorems, and more generally, by inversion, bounds of $\rho_A(\sigma^2)$ and $\rho_B(\sigma^2)$ for every σ^2. Finally, we provide in Section 7 additional bounds of $\gamma(q)$ and $\alpha(\rho)$, deduced from the continued fraction representations.

The rest of the paper is organized as follows. Section 3 provides a quick reminder of the combinatorial objects that we use in our proofs, mainly involutions and Dyck paths, the bijections between them, and some well known continued fractions that are related to q-Catalan numbers. Section 4 proves Theorems 1 and 3. Section 5 starts the asymptotic study of the numbers $B_{2k}(\rho)$, thus proving Theorem 4. The same study is achieved in Section 6 for the numbers $A_k(\rho)$ and Theorem 5. Section 7 complete this with further asymptotic results. For the convenience of the reader, we finally provide in Section 8 the probably well known proof of an integration by parts formula for Gaussian vectors, used in Section 6.

3. Combinatorial tools

3.1. Involutions and Dyck paths

We use the following notations. For $k \geq 1$, $[k] := \{1, 2, \dots, k\}$, $\mathcal{I}(k)$ is the set of the involutions of $[k]$ with no fixed point, $\mathcal{J}(k)$ is the subset of $\mathcal{I}(k)$ of the involutions σ with no crossing. This means that the configurations

$$i < j < \sigma(i) < \sigma(j)$$

do not appear in $\sigma \in \mathcal{J}(k)$. For odd values of k, $\mathcal{I}(k)$ and $\mathcal{J}(k)$ are empty.

Definition 11. *Let $i \in \mathrm{cr}(\sigma)$ denote the fact that $i < \sigma(i)$. Let $D(\sigma)$ denote the diameter of σ, that is, the sum of the diameters of its cycles. When $\sigma \in \mathcal{I}(2k)$, σ has only 2-cycles, thus*

$$D(\sigma) := \sum_{i \in \mathrm{cr}(\sigma)} \sigma(i) - i = \frac{1}{2} \sum_{i=1}^{2k} |i - \sigma(i)|.$$

Let $\mathcal{D}(2k)$ be the set of the Dyck paths of length $2k$, that is, of the sequences $c := (c_n)_{0 \leq n \leq 2k}$ of nonnegative integers such that

$$c_0 = c_{2k} = 0, \quad c_n - c_{n-1} = \pm 1, \quad n \in [2k].$$

Thus, exactly k indices $n \in [2k]$ correspond to ascending steps (c_{n-1}, c_n), that is, to steps when $c_n = c_{n-1} + 1$. We denote this by $n \in \mathrm{asc}(c)$. The k others indices correspond to descending steps, that is, to steps when $c_n = c_{n-1} - 1$, and we denote this by $n \in \mathrm{desc}(c)$. For odd values of k, let $\mathcal{D}(k)$ denote the empty set. Let $\mathrm{Area}(c)$ denote the area under the path c, that is,

$$\mathrm{Area}(c) := c_1 + c_2 + \cdots + c_{2k}.$$

Finally, $\mathcal{M}(2k)$ is the set of the marked Dyck paths (c, a), where $c \in \mathcal{D}(2k)$ and the sequence $a := (a_n)_{1 \le n \le 2k}$ is a mark of c, that is

$$\begin{aligned} &\text{if } n \in \mathrm{asc}(c), && a_n = 1, \\ &\text{if } n \in \mathrm{desc}(c), && a_n \in [c_{n-1}]. \end{aligned}$$

3.2. Bijections

We make use of bijections between $\mathcal{D}(2k)$ and $\mathcal{J}(2k)$, and between $\mathcal{M}(2k)$ and $\mathcal{I}(2k)$. Their construction is in Biane (1993) and we rephrase it slightly for our purposes.

If $c \in \mathcal{D}(2k)$, $\psi(c) := \sigma \in \mathcal{J}(2k)$ is an involution which maps each element of $\mathrm{desc}(c)$ to a smaller element of $\mathrm{asc}(c)$. Thus, $\mathrm{cr}(\sigma) = \mathrm{asc}(c)$. More specifically, if $n \in \mathrm{desc}(c)$, $\sigma(n)$ is the greatest $m \le n$ such that $(c_{m-1}, c_m) = (c_n, c_{n-1})$. This defines a bijection $\psi : \mathcal{D}(2k) \to \mathcal{J}(2k)$, such that

$$\mathrm{Area}(c) = D(\sigma).$$

In fact, ψ is the restriction of a bijection $\varphi : \mathcal{M}(2k) \to \mathcal{I}(2k)$, that we describe now. Here again, $\varphi(c, a) := \sigma$ is such that,

$$\text{if } n \in \mathrm{desc}(c), \text{ then } \sigma(n) \in \mathrm{asc}(c) \text{ and } \sigma(n) < n. \tag{6}$$

Since $c \ge 0$, $\mathrm{asc}(c) \cap [n]$ has greater cardinality than $\mathrm{desc}(c) \cap [n]$, for any $n \in [2k]$. If $n \ge 2$ is the first element of $\mathrm{desc}(c)$, $[n-1] \subset \mathrm{asc}(c)$, and one sets

$$\sigma(n) := a_n \in [c_{n-1}] = [n-1].$$

More generally, let $n \in \mathrm{desc}(c)$. For any choice of the images of the preceding descents which respects the rule (6), a careful enumeration of $[n-1]$ shows that there remains exactly c_{n-1} ascents before n that are not the image of a descent before n. Choose the a_nth greatest ascent before n as $\sigma(n)$.

Since one can reconstruct (c, a) from σ, the map φ is a bijection. If a is minimal, that is, if $a_n = 1$ for every n, then $\varphi(c, a) = \psi(c)$. Finally, $D(\varphi(c, a))$ is independent of a. Thus,

$$D(\varphi(c, a)) = D(\psi(c)) = \mathrm{Area}(c).$$

3.3. Continued fractions and q-Catalan numbers

Recall that the generating series of the ordinary Catalan numbers C_k, defined by (3), is

$$C(z) := \sum_{k\geq 0} C_k\, z^k = (1-(1-4z)^{1/2})/(2z).$$

This follows easily from $C_0 = 1$ and from the recursion relation

$$C_{k+1} = \sum_{n=0}^{k} C_n\, C_{k-n}.$$

For any $|q| \leq 1$, Carlitz and Riordan (1964) uniquely define q-Catalan numbers $C_k(q)$ by $C_0(q) := 1$ and

$$C_{k+1}(q) := \sum_{n=0}^{k} q^n\, C_n(q)\, C_{k-n}(q). \tag{7}$$

For instance,

$$C_1(q) = 1, \quad C_2(q) = 1+q, \quad C_3(q) = 1+2q+q^2+q^3,$$

and the usual Catalan numbers are $C_k = C_k(1)$. The q-Catalan generating function satisfies

$$C(q,z) := \sum_{k\geq 0} C_k(q)\, z^k = 1 + z\, C(q,z)\, C(q,qz).$$

Iterating this equation yields $C(q,z)$ as a continued fraction. To state this, let $[z] := z$ and

$$[z_1, z_2, \ldots, z_k] := \frac{z_1}{1-[z_2,\ldots,z_k]}.$$

Note the minus sign in the denominator. Let $[z_1, z_2, \ldots]$ denote the limit of $[z_1, z_2, \ldots, z_k]$ when $k \to \infty$, if this limit exists. Finally, let $(q)_0 = 1$ and $(q)_k = (q)_{k-1}\,(1-q^k)$. Then, $C(q,z)$ is the generalized Rogers-Ramanujan continued fraction, that is,

$$C(q,z) = [1, z, qz, q^2 z, \ldots, q^k z, \ldots]. \tag{8}$$

From chapter 7 of Andrews (1976), this is the ratio of two q-hypergeometric functions, that is, $C(q,z) = F(q,qz)/F(q,z)$, where

$$F(q,z) := \sum_{k\geq 0} (-1)^k\, q^{k^2-k}\, z^k/(q)_k. \tag{9}$$

4. Spectral measures by enumeration

We prove Theorem 1 in a detailed way and Theorem 3 more quickly. Both proofs follow Wigner's combinatorial argument.

4.1. Proof of Theorem 1

The mean non normalized trace adds the numbers

$$t(i) := \mathbb{E}(\gamma_{i_0 i_1}(1)\gamma_{i_1 i_2}(2)\cdots\gamma_{i_{2k-1} i_{2k}}(2k)),$$

for the sequences $i := (i_n)_{0\le n\le 2k}$ in $[N]$ such that $i_0 = i_{2k}$. Since the involved random variables are centered, we can assume that each edge (j, j') of $[N]\times[N]$, together with the reversed edge (j', j), does not appear at all, or appears at least twice in $t(i)$. Let $s(i)$ denote the support of i, that is,

$$s(i) := \{i_n\ ;\ 0 \le n \le 2k\}.$$

The contribution $t(i)$ is invariant by conjugacy, that is, $t(i) = t(i')$ if there exists a bijection from $s(i)$ to $s(i')$ that sends i_n to i'_n for any n. There exist $N(N-1)\cdots(N-s+1) \le N^s$ sequences in the conjugacy class of i, where s is the cardinality of $s(i)$, and the total number of conjugacy classes is finite and fixed with k. Since $B_k^N(\rho)$ involves a normalization by $N^{1+(k/2)}$, the classes such that $s \le (1+k)/2$ disappear in the $N \to \infty$ limit. This implies that the limit is zero when k is odd. From now on, we consider $B_{2k}^N(\rho)$.

The remaining sequences i involve exactly $1+k$ different elements i_n in $[N]$. Each pair $\{j, j'\}$ in the sequence i is such that $j \ne j'$ and it appears once as $(j, j') = (i_{n-1}, i_n)$ and once as $(j', j) = (i_{m-1}, i_m)$ with $m > n$. Thus,

$$\mathbb{E}(\gamma_{j,j'}(n)\gamma_{j',j}(m)) = \rho^{m-n}.$$

The unique involution $\sigma \in \mathcal{I}(2k)$ such that $\sigma(n) := m$ for the couples (n, m) that are described above characterizes the conjugacy class of i, and, from the definitions,

$$t(i) = \rho^{D(\sigma)}.$$

Consider the undirected graph $g(i)$ whose vertices are the elements of $s(i)$ and whose edges are the pairs $\{j, j'\}$ that appear in i. Viewed as a directed path on $s(i)$, the sequence i performs a walk on $g(i)$ that crosses twice every edge. Since the cardinality of $s(i)$ is $k+1$, $g(i)$ has no cycle. Thus, i performs the standard (left most, say) walk on the tree $g(i)$, rooted at i_0, see Takacs (1991) for instance.

Viewing the (conjugacy class of the) sequence i as the Bernoulli excursion $c \in \mathcal{D}(2k)$ which is encoded by $\sigma = \psi(c)$, this shows that $B_{2k}(\rho)$ is the sum over $\sigma \in \mathcal{J}(2k)$ of $\rho^{D(\sigma)}$, or the sum over $c \in \mathcal{D}(2k)$ of $\rho^{\mathrm{Area}(c)}$. This ends the proof of Theorem 1.

4.2. Proof of Theorem 3

A nice feature of the problem is that μ_k^N and $\mathbb{E}(\mu_k^N)$ are asymptotically close, in the $N \to \infty$ limit. We begin with $\mathbb{E}(\mu_k^N)$, and write the trace of $(Q_k^N)^\ell$, as in the proof of Theorem 1. Due to the normalization by $N^{1+(k\ell)/2}$, the remaining terms correspond to walks on rooted tree on $(k\ell)/2$ edges and $1+(k\ell)/2$ nodes, every edge $\{j, j'\}$ appearing twice, at some time instants $n < m$. Hence, $k\ell$ must be even and we assume this below. The only difference with Theorem 1 is the contribution of (n, m) to the correlation. To study this, we need one more definition.

Definition 12. *For $i \geq 1$, let $\{i\}_k$ denote the unique integer in $[k]$ such that k divides $i - \{i\}_k$. Introduce the diameter $D(k,\sigma)$ modulo k of σ as*

$$D(k,\sigma) := \frac{1}{2}\sum_i |\{i\}_k - \{\sigma(i)\}_k|.$$

For instance, if $\sigma \in \mathcal{J}(k)$, $\{i\}_k = i$ for any $i \in [k]$, hence $D(k,\sigma) = D(\sigma)$.

Each product considered above can be decomposed into ℓ blocks of size k. If both time instants n and m belong to the same block, the contribution of (n,m) is ρ^{m-n}. If n and m belong to different blocks, their contribution is

$$\rho^{|\{m\}_k - \{n\}_k|}.$$

Thus, the moment of order ℓ of μ_k^N converges to

$$B_{k,\ell}(\rho) := \sum_{\sigma \in \mathcal{J}(k\ell)} \rho^{D(k,\sigma)}.$$

The moments $B_{k,\ell}(\rho)$ specify a unique distribution, for example because Carleman's condition holds trivially, since

$$B_{k,\ell}(\rho) \leq B_{k,\ell}(1) = C_{k\ell/2} \leq 2^{k\ell}.$$

To end the proof of Theorem 3, it remains to check that the variance of any integral with respect to μ_k^N goes to zero as $N \to \infty$. From Bienaymé-Chebychev bound, this would imply that μ_k^N and $\mathbb{E}(\mu_k^N)$ have the same weak limit, in probability. Fix $\ell \geq 1$, and let $Q_N := (Q_k^N)^\ell$. One has to show that

$$\mathbb{E}((\mathrm{tr}\, Q_N)^2) - \mathbb{E}(\mathrm{tr}\, Q_N)^2$$

goes to zero when $N \to \infty$. Without the normalization by $N^{2+k\ell}$, this is the sum over all the paths i and j of length $k\ell$, such that $i_0 = i_{k\ell}$ and $j_0 = j_{k\ell}$, of the correlations

$$\mathbb{E}(t(i)\, t(j)) - \mathbb{E}(t(i))\, \mathbb{E}(t(j)).$$

There are two classes of couples (i,j). First, paths i and j can have no common edge. Then, $t(i)$ and $t(j)$ are independent and the correlation is zero. Second, there can exist at least one edge, common to i and j. This leaves N times less choices for i and j, meaning at most $N^{1+k\ell}$ possible couples. Because one normalizes by $N^{2+k\ell}$, this contribution vanishes, when $N \to \infty$, and the proof is complete.

Remark 13. Consider the special cases $k = 1$ and $k = 2$. First, $\{\cdot\}_1 = 1$, hence $D(1,\cdot) = 0$. Thus, $B_{1,\ell}$ is the cardinality of $\mathcal{J}(\ell)$, that is, zero if ℓ is odd, and $C_{\ell/2}$ if ℓ is even. This describes ν_1. Turning to the case $k = 2$, note that $\{2i+1\}_2 = 1$ and $\{2i\}_2 = 2$ for any i. For any $\sigma \in \mathcal{J}(n)$, the pairs $\{i, \sigma(i)\}$ are made of integers of opposite parities. Thus,

$$D(2,\sigma) = \ell,$$

for any $\sigma \in \mathcal{J}(2\ell)$, and $B_{2,\ell}(\rho)$ is ρ^ℓ times the cardinal of $\mathcal{J}(2\ell)$, that is,

$$B_{2,\ell}(\rho) = \rho^\ell\, C_\ell.$$

Recall that W denotes a random variable of law ν_1, defined by (2), whose moments are given by $B_{2\ell} = C_\ell$ and $B_{2\ell+1} = 0$. Thus, $\nu_{2,\rho}$ is the distribution of $\rho\, W^2$. □

Remark 14. The supremum $e_{k,\rho}$ of the support of $\nu_{k,\rho}$ is

$$\lim_{\ell\to\infty} B_{k,2\ell}^{1/2\ell}(\rho),$$

thus $e_{k,\rho} \leq 2^k$. This bound is obvious from the fact that the support of the semicircle law ν_1 is $[-2, 2]$. Can one compute $e_{k,\rho}$ in the general case? One knows that $e_{k,1} = 2^k$, $e_{1,\rho} = 2$ and $e_{2,\rho} = 4\rho$. Can one show that the support of $\nu_{k,\rho}$ is (a subset of) the interval $[0; e_{k,\rho}]$ or $[-e_{k,\rho}; e_{k,\rho}]$ of the real line, depending on the parity? □

5. Phase transition for *q*-Catalan numbers

The starting point to the asymptotic study of $B_{2k}(\rho)$ is Lemma 15, that links $B_{2k}(\rho)$ to q-Catalan numbers.

Lemma 15. *Setting* $B_0(\rho) := 1$, *for any* $k \geq 0$,

$$B_{2k+2}(\rho) = \sum_{n=0}^{k} \rho^{2n+1}\, B_{2n}(\rho)\, B_{2k-2n}(\rho). \tag{10}$$

Proof of Lemma 15. Decompose the sum that defines $B_{2k+2}(\rho)$ in Theorem 1 along the values of $\sigma(1)$. Due to the non crossing property, $\sigma(1) = 2n + 2$ for a given $0 \leq n \leq k$. Then, σ is the juxtaposition of three pieces, namely, the cycle $\{1, 2n + 2\}$, which has diameter $2n + 1$, the restriction σ_- of σ to $\{2, \cdots, 2n + 1\}$, which corresponds to a given involution of $\mathcal{J}(n)$, and the restriction σ_+ of σ to $\{2n + 3, \cdots, 2k + 2\}$, which corresponds to a given involution of $\mathcal{J}(k - n)$. When $n = 0$ or $n = k$, σ_- or σ_+ does not appear, and we set $D(\sigma_-) = 0$ or $D(\sigma_+) = 0$. Then, for any n,

$$D(\sigma) = 2n + 1 + D(\sigma_-) + D(\sigma_+),$$

a fact which implies (10). □

Thus, B_{2k} is a polynomial function of $\rho \geq 0$, of valuation k, degree k^2, with nonnegative integer coefficients. A direct consequence of Equations (10) and (7) is Equation (4) in Section 2.1. We now prove Theorem 4 and Corollary 8. See Remark 19 below for additional properties of β.

Lemma 16. *The sequences* $(A_{2k})_k$, $(B_{2k})_k$ *and* $(C_k)_k$ *are submultiplicative, in the following sense. For* $D_k := A_{2k}$ *or* $D_k := B_{2k}$ *or* $D_k := C_k$, *and for* $u \geq 0$,

$$D_{k+m}(u) \geq D_k(u)\, D_m(u).$$

Thus, $A_{2k}(\rho)^{1/k} \to \alpha(\rho)$, $B_{2k}(\rho)^{1/k} \to \beta(\rho)$ *and* $C_k(q)^{1/k} \to \gamma(q)$, *where* $\alpha(\rho)$, $\beta(\rho)$ *and* $\gamma(q)$ *are defined on* $[0; +\infty)$ *and valued in* $[0; +\infty]$. *Finally,*

$$\beta(\rho) = \rho\, \gamma(\rho^2).$$

Lemma 17. *For $q \in [0; 1)$ and $k \geq 1$,*

$$(1+q)^{k-1} \leq C_k(q) \leq \gamma_2(q)^{k-1}, \tag{11}$$

where $\gamma_2(q)$ is the unique positive solution of $\gamma^2 - \gamma = q/(1-q)$. Thus, $\gamma_2(q) \leq (1-q)^{-1}$.

Corollary 18. *For any $q < 1$ and any $\rho < 1$,*

$$(1+q) \leq \gamma(q) \leq \gamma_2(q) \leq (1-q)^{-1},$$

$$\rho\,(1+\rho^2) =: \beta_1(\rho) \leq \beta(\rho) \leq \beta_2(\rho) := \rho\gamma_2(\rho^2) \leq \rho/(1-\rho^2).$$

Remark 19. The function $\beta(\rho)/\rho$ is nondecreasing. Thus, for any $\rho \in [0; 1]$, $\rho \leq \beta(\rho) \leq 4\rho$. Using results of this section and of Section 7, this is refined, for $\rho \in [0; 1)$, by

$$\beta_1(\rho) \leq \beta(\rho) \leq 2\,\beta_1(\rho), \quad \beta(\rho) \leq \beta_2(\rho),$$

with

$$\beta_1(\rho) := \rho(1+\rho^2), \quad \beta_2(\rho) := \frac{1}{2}\,\rho\left(1 + \left(\frac{1+3\rho^2}{1-\rho^2}\right)^{1/2}\right).$$

Thus, $\beta(\rho) = \rho + \rho^3 + O(\rho^7)$ when $\rho \to 0+$. (See Remark 27 in Section 7 for more about this.) The best upper bound of β is β_2 for values of ρ up to a certain value, and $2\beta_1$ for higher values of ρ. The value $\beta(1) = 4$ comes from the exact expression of $B_{2k}(1) = C_k$, the usual kth Catalan number. Finally, one can prove that the critical point $\rho = \rho_B$ is the unique solution of the equation

$$\sum_{k \geq 0} \rho^{2k+1} B_{2k}(\rho) = 1.$$

□

Proof of Theorem 4. From (7), $C_k(q)$ is nondecreasing. Hence, for any $\rho \geq \rho'$, $\beta(\rho) \geq \beta(\rho')\,\rho/\rho'$, and β is strictly increasing.

Assume that $\rho > 1$. Iterating the inequality $C_{k+1}(q) \geq q^k\,C_k(q)$, which is a byproduct of (7), one gets $C_{k+1}(q) \geq q^{k(k-1)/2}$. Thus, $B_{2k}(\rho) \geq \rho^{k^2}$ and $\beta(\rho)$ is infinite. We assume now that $\rho < 1$ and we consider $C_k(q)$.

We first show that $\gamma(q) \to 4$ when $q \to 1$. Let $z > 1/4$ be close to $1/4$. Since the generating function $C(1, \cdot)$ diverges at $1/4$, consider the first truncation of $C(1, \cdot)$, say

$$[1, z, z, \ldots, z],$$

that takes a negative value at z. By continuity, the same holds for

$$[1, z, qz, \ldots, q^k z],$$

if $q < 1$ is large enough. For such values of q, the radius of convergence of $C(q, \cdot)$ is less than z. This means that $\gamma(q) \geq 1/z$.

The continuity of β on (0; 1) is a consequence of Proposition 20 below, since the continued fraction $B(\rho, z) := C(\rho^2, \rho z)$ obviously satisfies hypothesis (H) in this proposition, and $\beta(\rho)$ is the inverse of the radius of convergence of $B(\rho, \cdot)$. Note that (H) covers the one-dimensional case as well. We conclude the proof of Theorem 4, translating bounds of β into bounds of ρ_B.

From Corollary 18, $\beta_1 \leq \beta \leq \beta_2$, where β_1 and β_2 are increasing. Assume that $\beta_1(\rho_1) = \beta_2(\rho_2) = 1$. Then, if $\rho > \rho_1$, $\beta(\rho) \geq \beta_1(\rho) > 1$, hence $\rho_B \leq \rho_1$. Likewise, $\rho_B \geq \rho_2$. Finally,

$$\rho_1 \simeq .6823278040, \quad \rho_2 \simeq .6609925319.$$

□

Proposition 20. *Set $CF(q, z) := [1, f_1(q)\, z, \ldots, f_k(q)\, z, \ldots]$. Assume that each real valued function f_k is continuous on (0; 1), nonnegative, nondecreasing, and that, for any $q \in (0; 1)$, $f_k(q) \to 0$ when $k \to \infty$, uniformly in the following sense.*

(H) *For any $q \in (0; 1)$ and any $\varepsilon > 0$, there exists a neighborhood $U \subset (0; 1)$ of q and a finite integer N, such that, for all $r \in U$ and $k \geq N$, $f_k(r) \leq f_N(r) \leq \varepsilon$.*

Then, the radius of convergence $R(q)$ of the holomorphic function $CF(q, \cdot)$ is a nonincreasing continuous function of q.

Proof of Proposition 20. We prove that R is continuous on (0; 1), the sense of variation of R being obvious. Let $q \in (0; 1)$ and $z < R(q)$. Choose $\varepsilon > 0$ such that $(1 + \varepsilon)^2 z < R(q)$. Choose U and N from (H). Since $f_k(q) \to 0$ when $k \to \infty$, replace N by a larger value, if necessary, to ensure that $f_N(q) \leq \varepsilon$. Since f_1, $\ldots$, f_N are continuous at q, replace U by a smaller neighborhood, if necessary, to ensure that $f_k(r) \leq (1 + \varepsilon)\, f_k(q)$ for all $r \in U$ and $k \leq N$. Finally, introduce

$$[[u]] := [u, u, \ldots] = (1 - (1 - 4u)^{1/2})/2,$$

and note that $[[u]] = u + o(u)$ when $u \to 0$. Hence, one can assume that $[[u]] \leq (1 + \varepsilon)\, u$, for every nonnegative $u \leq 2\varepsilon\, R(q)$. For any $r \in U$,

$$[f_{N+1}(r)\, z, f_{N+2}(r)\, z, \ldots] \leq [[f_N(r)\, z]] \leq [[(1 + \varepsilon)\, f_N(q)\, z]],$$

which is at most $(1 + \varepsilon)^2 f_N(q)\, z$, since $f_N(q) \leq \varepsilon$ and $z < R(q)$. Hence,

$$CF(r, z) \leq [1, (1 + \varepsilon)\, f_1(q)\, z, \ldots, (1 + \varepsilon)\, f_{N-1}(q)\, z, (1 + \varepsilon)^2 f_N(q)\, z],$$

which is at most $CF(q, (1 + \varepsilon)^2 z)$. This converges, since $(1 + \varepsilon)^2 z < R(q)$. We proved that $R(r) \geq z$, that is, that R is lower semicontinuous.

On the other hand, let $R_k(q)$ denote the radius of convergence of the kth truncated continued fraction $CF_k(q, \cdot)$ of $CF(q, \cdot)$. If $z \leq (1 - \varepsilon)\, R_k(q)$, one can show that $CF_k(q, z) \leq 1/\varepsilon$. This implies that the nonincreasing sequence R_k of continuous functions converges to R when $k \to \infty$. Hence, R is upper semicontinuous. □

We now prove Lemmas 16 and 17. The submultiplicativity in Lemma 16 is a consequence of the following definition.

Definition 21. *If σ is a permutation of $[k]$ and τ a permutation of $[m]$, the concatenation $\sigma \star \tau$ of σ and τ is the permutation of $[k+m]$ defined by*

$$\begin{aligned} \sigma \star \tau(n) &:= \sigma(n) && \text{if } n \le k, \\ &:= k + \tau(n-k) && \text{if } n > k. \end{aligned}$$

Remark 22. If $(\sigma, \tau) \in \mathcal{I}(k) \times \mathcal{I}(m)$, then $\sigma \star \tau \in \mathcal{I}(k+m)$. If σ and τ have no crossing cycle, neither has $\sigma \star \tau$. Finally, the map $(\sigma, \tau) \mapsto \sigma \star \tau$ is injective. □

Proof of Lemma 16. From Remark 22, $(A_{2k}(\rho))_k$ and $(B_{2k}(\rho))_k$ are submultiplicative. For instance, $(B_{2k}(\rho))^{1/k}$ converges to

$$\beta(\rho) := \sup_{k \ge 1} (B_{2k}(\rho))^{1/k}, \quad \beta(\rho) \in [0; +\infty].$$

The result for $(C_k(q))_k$ follows from (4). □

Proof of Lemma 17. Since $C_1 = 1$, (11) holds for $k = 1$. Assume that the lower bound holds for C_n, $n \le k$, with $k \ge 1$. Then, (7) yields

$$\begin{aligned} C_{k+1}(q) &= (1+q^k)\, C_k(q) + \sum_{n=1}^{k-1} q^i\, C_n(q)\, C_{k-n}(q) \\ &\ge (1+q^k)\,(1+q)^{k-1} + (1+q)^{k-2}\,(q + \cdots + q^{k-1}). \end{aligned}$$

Note that $k \ge 1$ ensures that the terms $n = 0$ and $n = k$ are indeed different. Also, the last sum is null if $k = 1$. Thus, $C_{k+1}(q) \ge (1+q)^k$, as soon as

$$(1+q^k)\,(1+q) + (q + \cdots + q^{k-1}) \ge (1+q)^2.$$

This holds if $k \ge 2$ (one cancels the $q^k\,(1+q)$ term and keeps only two terms in the sum), and this holds as an equality if $k = 1$. Note that this part of the proof is valid for any $q \ge 0$. Assume now that $q < 1$. Likewise, the recursion for the upper bound holds as soon as, for any $k \ge 1$,

$$\begin{aligned} \gamma^2 &\ge (1+q^k)\,\gamma + (q + \cdots + q^{k-1}) \\ &= \gamma + \frac{q}{1-q} + q^k \left(\gamma - \frac{1}{1-q}\right). \end{aligned}$$

Since q^k decreases from q to 0 when k increases from 1 to ∞, it is enough to check the inequality at $k = 1$ and $k = \infty$. This yields the conditions $\gamma \ge 1 + q$, and $\gamma \ge \gamma_2(q)$, with

$$\gamma_2(q)^2 = \gamma_2(q) + \frac{q}{1-q}.$$

The second condition implies the first one. Note that $\rho = \rho_2$ in the proof of Theorem 4 is defined by $\rho\, \gamma_2(\rho^2) = 1$, that is,

$$\frac{1}{\rho^2} = \frac{1}{\rho} + \frac{\rho^2}{1-\rho^2},$$

that is: $\rho^4 - \rho^3 + \rho^2 + \rho - 1 = 0$. Since this polynomial has a unique real nonnegative root, the proof is complete. □

6. Gauss-Markov products in dimension one

Proof of Theorem 2. A repeated use of the integration by part formula for Gaussian vectors, see Lemma 28 in Section 8, yields the first equality of Equation (5). To see this, set $\gamma := (\gamma_1, \dots, \gamma_{2k})^t$, $F(\gamma) := \gamma_2 \cdots \gamma_{2k}$, and use Lemma 28. Each $\partial_n F(\gamma)$ is the product $\gamma_2 \cdots \gamma_k$ where one omits γ_n. Use Lemma 28 again to compute $\mathbb{E}(\partial_n F(\gamma))$, with the term γ_m of smallest index in lieu of γ_1 and the remaining product in lieu of $F(\gamma)$. After k applications of the lemma, one gets A_{2k} as a sum of products of $\mathbb{E}(\gamma_n \gamma_m)$. Replacing each of these expectations by the corresponding power of ρ, and noting that each involution comes into play exactly once, one gets the first equality of Equation (5). The bijection φ reduces the sum over involutions to a sum over marked paths $(c, a) \in \mathcal{M}(2k)$. Then, $D(\sigma) = \text{Area}(c)$ is independent of a, and the number of marks of each path c, that is, the cardinality of the set $\{a\,;\ (c, a) \in \mathcal{M}(2k)\}$, is

$$\prod_{n \in \text{desc}(c)} c_{n-1}.$$

This yields the second part of (5). □

Thus, $A_{2k}(\rho)$ is an enumeration of marked paths, or an enumeration of involutions. The latter is a special case of the enumerations considered by Flajolet (1980). Hence, the expression of the generating function of $(A_{2k}(\rho))_k$ as a continued fraction in LMOS is straighforward, yielding

$$\begin{aligned} A(\rho, z) &:= 1 + \sum_{k \geq 1} A_{2k}(\rho)\, z^k \\ &= [1, \rho z, 2\rho^3 z, \dots, k\, \rho^{2k-1}\, z, \dots]. \end{aligned} \tag{12}$$

We stick in this section to the formulation of Equation (5). Like $B_{2k}(\rho)$, $A_{2k}(\rho)$ can be defined for any ρ. This polynomial function has valuation k, degree k^2, and nonnegative integer coefficients. We turn to the proof of Theorem 5.

Lemma 23. *For $\rho \in [0; 1)$ and $k \geq 1$,*

$$A_{2k-2}(\rho)\, \frac{\rho - \rho^{4k+3}}{1 - \rho^2} \leq A_{2k}(\rho) \leq A_{2k-2}(\rho)\, \frac{\rho + \rho^3}{1 - \rho^2}.$$

Corollary 24. *For $\rho \in [0; 1)$,*

$$\frac{\rho}{1 - \rho^2} =: \alpha_3(\rho) \leq \alpha(\rho) \leq \alpha_4(\rho) := \frac{\rho + \rho^3}{1 - \rho^2}.$$

Thus, $\alpha(0+) = 0$ and $\alpha(1-) = +\infty$.

Proof of Theorem 5. From Lemma 16, $(A_{2k})_k$ is submultiplicative. Thus, $(A_{2k}(\rho))^{1/k}$ converges to

$$\alpha(\rho) := \sup_{k \geq 1} (A_{2k}(\rho))^{1/k}, \quad \alpha(\rho) \in [0; +\infty].$$

Like $B_{2k}(\rho)/\rho^k$, $A_{2k}(\rho)/\rho^k$ is nondecreasing, hence α is strictly increasing on its domain. The continuity of α on (0; 1) is a consequence of Proposition 20, since $\alpha(\rho)$ is the inverse of the radius of convergence of $A(\rho, \cdot)$.

From Lemma 23, $\rho_A \geq \rho_4$ where $\alpha_4(\rho_4) = 1$, that is, $\rho_4 \simeq .5436890125$. Likewise, $\rho_A \leq \rho_3$ where $\alpha_3(\rho_3) = 1$, that is, $\rho_3 \simeq .6180339890$. □

Proof of Lemma 23. Let σ be an involution of $[2k + 2]$ with no fixed point, and define $n+1 := \sigma(1)$. Then, $1 \leq n \leq 2k+1$ and the cycle $\{1, n+1\}$ has diameter n. The other cycles contribute as follows. Let $\tau(m) = \sigma(m+1)-1$ if $m \in [n-1]$, and $\tau(m) = \sigma(m+2)-2$ if $n \leq m \leq 2k$. Thus, $\tau \in \mathcal{I}(k)$. Let $\kappa_n(\sigma)$ denote the number of cycles of σ that cross $n + 1$, that is, the number of i such that $i < n + 1 < \sigma(i)$. Then,

$$D(\tau) = D(\sigma) - n - \kappa_n(\sigma).$$

Thus, $D(\sigma) - n \geq D(\tau)$, and, when n is even, $D(\sigma) - n \geq D(\tau) + 1$, since at least one cycle of σ crosses $n + 1$. Furthermore, for a fixed n, the map $\sigma \mapsto \tau$ is one-to-one, from the set of the involutions $\sigma \in \mathcal{I}(k + 1)$, such that $\sigma(1) = n + 1$, to $\mathcal{I}(k)$. Assuming that $\rho < 1$, one gets

$$A_{2k+2}(\rho) \leq \sum_{n\geq 0} \rho^{2n+1} \sum_{\tau} \rho^{D(\tau)} + \sum_{n\geq 0} \rho^{2n+2} \sum_{\tau} \rho^{D(\tau)+1} = A_{2k}(\rho)\, \frac{\rho + \rho^3}{1 - \rho^2}.$$

On the other hand, $\kappa_n(\sigma) \leq \min\{n - 1, 2k + 1 - n\}$, hence $D(\sigma)$ is at most $D(\tau) + 2n - 1$. This yields

$$A_{2k+2}(\rho) \geq \sum_{n=1}^{2k+1} \sum_{\tau} \rho^{D(\tau)} \rho^{2n-1} = A_{2k}(\rho) \sum_{n=0}^{2k} \rho^{2n+1}.$$

□

7. Asymptotic estimates

The continued fraction expressions of $A(\rho, z)$ and $C(q, z)$ in (12) and (8) yield bounds of $\alpha(\rho)$ and $\gamma(q)$, thus of $\beta(\rho)$. We skip the proofs of Lemmas 25 and 26 below, which can be deduced from Theorems 11.2 and 14.1 of Wall (1948).

Lemma 25. *Set $q_k := q^{k-1}$. Then, $\gamma(q) \geq \gamma_5(q)$, where $\gamma = \gamma_5(q)$ is the largest root of*

$$\gamma^2 - (q_1 + q_2 + q_3 + q_4)\,\gamma + q_1\, q_3 + q_1\, q_4 + q_2\, q_4 = 0.$$

A consequence is $\gamma(q) > q_1 + q_2 = 1 + q$. On the other hand,

$$\gamma(q) \leq 2 \sup\{q_k + q_{k+1}\,;\ k \geq 1\} = 2\,(1 + q).$$

Lemma 26. *Set $\rho_k := k\,\rho^{2k-1}$. Then, $\alpha(\rho) \geq \alpha_6(\rho)$, where $\alpha = \alpha_6(\rho)$ is the largest root of*

$$\alpha^2 - (\rho_1 + \rho_2 + \rho_3 + \rho_4)\,\alpha + \rho_1\, \rho_3 + \rho_1\, \rho_4 + \rho_2\, \rho_4 = 0.$$

A consequence is $\alpha(\rho) > \rho_1 + \rho_2 = \rho + 2\rho^3$. *On the other hand,*

$$\alpha(\rho) \le 2 \sup\{\rho_k + \rho_{k+1}\,;\ k \ge 1\}.$$

This implies for example that $\alpha(\rho) \le 2\,(\rho + 2\rho^3)$ *if* $\rho \le 3^{-1/4}$, *and that*

$$\lim_{\rho \to 1} (1-\rho)\,\alpha(\rho) = 2/\mathrm{e} \simeq .7357588824.$$

Thus, $\rho_B \le \rho_5$ where $\rho_5\,\gamma_5(\rho_5^2) = 1$, that is, ρ_5 solves

$$\rho^{10} + \rho^8 + \rho^6 + 1 = \rho^7 + \rho^5 + \rho^3 + \rho.$$

Hence, $\rho_5 \simeq .6629288547$. Likewise, $\rho_A \le \rho_6$ where $\alpha_6(\rho_6) = 1$, that is, ρ_6 solves

$$8\rho^{10} + 4\rho^8 + 3\rho^6 + 1 = 4\rho^7 + 3\rho^5 + 2\rho^3 + \rho.$$

Hence, $\rho_6 \simeq .5630593728$. Note that ρ_5 and ρ_6 are approximations of ρ_B and ρ_A, up to $3 \cdot 10^{-5}$ and $6 \cdot 10^{-5}$, respectively.

Remark 27. The representation (9) of $C(q,z)$ yields an expansion of $\gamma(q)$ along powers of q. For instance, considering the first four terms, one gets

$$\begin{aligned}(1-q)\,(1-q^2)\,(1-q^3)\,\gamma^3 - (1-q^2)\,(1-q^3)\,\gamma^2 \\ + (1-q^3)\,q^2\,\gamma - q^6 + q^{12} = O(q^{13}),\end{aligned}$$

that is,

$$\begin{aligned}\gamma(q) = 1 + q + q^3 - q^4 + 2\,q^5 - 3\,q^6 + 6\,q^7 - 12\,q^8 + 25\,q^9 \\ -52\,q^{10} + 111\,q^{11} - 241\,q^{12} + O(q^{13}).\end{aligned}$$

Computations with the help of Maple 6 ©, up to order q^{156}, seem to indicate that the expansion of $\gamma(q)$ is

$$\gamma(q) = 1 + \sum_{n \ge 1} (-1)^{n+1}\, c_n\, q^n,$$

with nonnegative integers c_n, such that the sequence c_{n+1}/c_n is nondecreasing for $n \ge 6$, and converges to a finite limit $c > 2.4$. Thus, one would have $c_n = c^{n+o(n)}$, as $n \to \infty$.

Finally, elementary computations show that $\gamma(q)$ is related to the least positive zero $q_0(z)$ of the generalised Rogers-Ramanujan continued fraction, that is, with our notations, of the function $q \mapsto F(q, qz)$. More precisely,

$$1/\gamma(q) = q\, q_0^{-1}(q),$$

where q_0^{-1} is the inverse function of q_0. See Theorem 6.2 of Berndt, Huang, Sohn and Son (2000), and the sequence `A050203` in Sloane's On-Line Encyclopedia of Integer Sequences (our Reference [11]), for expansions of $q_0(z)$ along the powers of $1/z$. □

8. Gaussian integration by parts

For the sake of completeness, we prove an integration by parts formula for Gaussian vectors, that we used above, and which is probably well known since it yields one easy proof of Wick's formula. The notations of this section are independent of the rest of the paper.

Lemma 28. *Let $\gamma := (\gamma_1, \ldots, \gamma_k)^{\mathrm{t}}$ denote a centered Gaussian vector, F a smooth function such that $F(\gamma)$ is integrable, and* $\mathrm{grad} F$ *the gradient of F. Then,*

$$\mathbb{E}(\gamma_1 F(\gamma)) = \mathbb{E}(\gamma_1 \gamma) \cdot \mathbb{E}(\mathrm{grad} F(\gamma)) = \sum_{n=1}^{k} \mathbb{E}(\gamma_1 \gamma_n)\, \mathbb{E}(\partial_n F(\gamma)).$$

Proof of Lemma 28. Conditionally on γ_1, γ is distributed like $\gamma_1 \alpha + \delta$, where δ is a centered Gaussian vector, independent of γ_1, and where $\alpha := \mathbb{E}(\gamma_1 \gamma)/\sigma^2$. Thus, α is a deterministic vector.

Fix δ, set $\sigma^2 := \mathbb{E}(\gamma_1^2)$ and $g_\delta(\gamma_1) := F(\gamma_1 \alpha + \delta)$. The usual integration by parts of the function $x\, g_\delta(x)\, \mathrm{e}^{-x^2/2\sigma^2}$ in dimension 1 yields

$$\mathbb{E}(\gamma_1 g_\delta(\gamma_1)) = \sigma^2\, \mathbb{E}(g'_\delta(\gamma_1)).$$

There is δ-almost surely no boundary term because $F(\gamma)$ is integrable. Since $\sigma^2 \alpha = \mathbb{E}(\gamma_1 \gamma)$, and $g'_\delta = \alpha \cdot \mathrm{grad} F$, one gets

$$\mathbb{E}(\gamma_1 F(\gamma) \mid \delta) = \mathbb{E}(\gamma_1 \gamma) \cdot \mathbb{E}(\mathrm{grad} F(\gamma_1 \alpha + \delta) \mid \delta).$$

Taking expectations of both sides yields the lemma. □

References

1. Andrews G.E.: *The theory of partitions*, volume 2 of *Encyclopedia of Mathematics and its Applications*. Addison-Wesley, Reading, Mass., 1976
2. Ben Arous, G., Guionnet, A.: Large deviations for Wigner's law and Voiculescu's non-commutative entropy. *Probab. Theory Related Fields*, **108**, 517–542 (1997)
3. Berndt, B.C., Huang, S.S., Sohn, J., Son, S.H.: Some theorems on the Rogers-Ramanujan continued fraction in Ramanujan's lost notebook. *Trans. Amer. Math. Soc.*, **352**, 2157–2177 (2000)
4. Biane, P.: Permutations suivant le type d'excédance et le nombre d'inversions et interprétation combinatoire d'une fraction continue de Heine. *European J. Combin.*, **14**, 277–284 (1993)
5. Carlitz, L., Riordan, J.: Two element lattice permutation numbers and their q-generalization. *Duke Math. J.*, **31**, 371–388 (1964)
6. Flajolet, P.: Combinatorial aspects of continued fractions. *Discrete Math.*, **32**, 125–161 (1980)
7. Guionnet, A., Zeitouni, O.: Concentration of the spectral measure for large matrices. *Electron. Comm. Probab.*, **5**, 119–136 (electronic), 2000
8. Khorunzhy, A.: Products of random matrices and q-Catalan numbers. Preprint `math.CO/0104074`

9. Logan, B.F., Jr., Mazo, J.E., Odlyzko, A.M., Shepp, L.A.: On the average product of Gauss-Markov variables. *Bell System Tech. J.*, **62**, 2993–3006 (1983)
10. Odlyzko, A.M., Wilf, H.F.: The Editor's corner: n coins in a fountain. *Amer. Math. Monthly*, **95**, 840–843 (1988)
11. Sloane, N.J.A.: The On-Line Encyclopedia of Integer Sequences. `www.research.att.com/~njas/sequences/`
12. Speicher, R.: A new example of "independence" and "white noise". *Probab. Theory Related Fields*, **84**, 141–159 (1990)
13. Voiculescu, D.V., Dykema, K.J., Nica, A.: *Free random variables*. American Mathematical Society, Providence, RI, 1992. A noncommutative probability approach to free products with applications to random matrices, operator algebras and harmonic analysis on free groups
14. Wall, H.S.: *Analytic theory of continued fractions*. Van Nostrand, New York, 1948
15. Wigner, E.P.: Characteristic vectors of bordered matrices with infinite dimensions. *Ann. of Math.*, **62**, 548–564 (1955)
16. Wigner, E.P.: Characteristic vectors of bordered matrices with infinite dimensions. II. *Ann. of Math.*, **65**, 203–207 (1957)
17. Wigner, E.P.: On the distribution of the roots of certain symmetric matrices. *Ann. of Math.*, **67**, 325–327 (1958)