MONOTONE HURWITZ NUMBERS AND THE HCIZ INTEGRAL

I. P. GOULDEN, M. GUAY-PAQUET, AND J. NOVAK

Abstract. We give a combinatorial interpretation of the Harish-Chandra-Itzykson-Zuber integral as a generating function for monotone walks on the symmetric groups. This leads to a relation between the logarithm of the HCIZ integral and a desymmetrized version of the double Hurwitz numbers. The link with Hurwitz theory is exploited to prove the complex convergence of the HCIZ free energy under a non-vanishing hypothesis.

Contents

0. Introduction

The Harish-Chandra-Itzykson-Zuber integral,

$$I_N(z) = \int_{U(N)} e^{zN\,\mathrm{Tr}(A_N U B_N U^{-1})} \mathrm{d}U, \tag{0.1}$$

is an important special function which plays a key role in many aspects of random matrix theory, ranging from the fine-scale spectral theory of a single random matrix to the global statistics of several interacting random matrices. For more information

Date: November 25, 2013.
2010 *Mathematics Subject Classification.* Primary 05E10,15B62; Secondary 14N10.
Key words and phrases. Matrix models, enumerative geometry, asymptotic analysis.
IPG and MG-P supported by NSERC.

on the role of the HCIZ integral in random matrix theory, the reader is referred to the surveys [6, 14, 30].

In (0.1), the integration is over the group of $N \times N$ unitary matrices against the normalized Haar measure, z is a complex parameter, and A_N, B_N are any two $N \times N$ complex diagonal matrices. Since $U(N)$ is compact, $I_N(z)$ is an entire function of the complex variable z. If we restrict to A_N, B_N real diagonal, a famous formula of Harish-Chandra [16] and Itzykson and Zuber [18] expresses $I_N(z)$ as a ratio of $N \times N$ determinants. One consequence of the determinantal formula is a link between the HCIZ integral and Schur polynomials. We will not assume A_N, B_N real, and so the determinantal formula does not apply. Instead, we work directly with (0.1) and develop an explicit combinatorial model for its Maclaurin coefficients as a function of z (Theorem 1.3).

Our combinatorial model leads to an interesting connection between the HCIZ integral and a desymmetrized version of Hurwitz theory, a classical topic in enumerative algebraic geometry. Using this connection, we prove that when A_N and B_N are uniformly bounded and converge in moments, the Maclaurin coefficients of the logarithm of $I_N(z)$ admit a topological expansion (Theorem 2.1). We then show that, if $I_N(z)$ is non-vanishing on a neighbourhood of $z = 0$ for N sufficiently large, the HCIZ free energy converges uniformly on compact subsets of a complex domain (Theorem 3.7). We verify this non-vanishing hypothesis in a family of special cases.

1. The leading derivatives theorem

Let $S(d)$ denote the symmetric group acting on $\{1, \ldots, d\}$, and identify $S(d)$ with its (right) Cayley graph as generated by the full conjugacy class of transpositions. Define an edge labelling of the Cayley graph as follows: each edge corresponding to the transposition $\tau = (s\ t)$ is marked by t, the larger of the two numbers interchanged. Thus, emanating from each vertex of the Cayley graph, there is one 2-edge, two 3-edges, three 4-edges, etc.

Definition 1.1. A walk on the Cayley graph of $S(d)$ is said to be *monotone* if the labels of the edges it traverses form a weakly increasing sequence.

The enumeration of monotone walks has been considered by Stanley [27], Biane [1], and Gewurz and Merola [7]. These authors showed that the number of monotone geodesics joining the identity permutation to the full cycle $(1\ 2\ \ldots\ d)$ is the Catalan number

$$\mathrm{Cat}_{d-1} = \frac{1}{d}\binom{2d-2}{d-1}.$$

This was generalized by Murray [22] and Matsumoto and Novak [20], who showed that the number of monotone geodesics joining the identity to any permutation $\pi \in S(d)$ is

$$\prod_{i=1}^{\ell(\beta)} \mathrm{Cat}_{\beta_i - 1},$$

where $\beta = t(\pi)$ is the cycle type of π.

Remarkably, computing the Maclaurin coefficients of the HCIZ integral is equivalent to enumerating monotone walks of arbitrary length, with arbitrary boundary conditions.

Definition 1.2. Given two Young diagrams α and β, each with d cells, let $\vec{W}^r(\alpha,\beta)$ denote the number of r-step monotone walks on $S(d)$ which begin in the the conjugacy class C_α and end in the conjugacy class C_β.

Set

$$p_\alpha(A_N) = \prod_{i=1}^{\ell(\alpha)} \mathrm{Tr}(A_N^{\alpha_i}), \quad p_\beta(B_N) = \prod_{j=1}^{\ell(\beta)} \mathrm{Tr}(B_N^{\beta_j}),$$

the power sum symmetric functions p_α, p_β evaluated on the eigenvalues of A_N, B_N.

Theorem 1.3. *For any $N \geq 1$ and any $1 \leq d \leq N$, we have*

$$I_N^{(d)}(0) = \sum_{r=0}^{\infty} \left(-\frac{1}{N}\right)^r \sum_{\alpha,\beta\vdash d} p_\alpha(A_N)p_\beta(B_N)\vec{W}^r(\alpha,\beta),$$

and this series is absolutely convergent. Equivalently, we have

$$I_N(z) = 1 + \sum_{d=1}^{N} \frac{z^d}{d!} \sum_{r=0}^{\infty} \left(-\frac{1}{N}\right)^r \sum_{\alpha,\beta\vdash d} p_\alpha(A_N)p_\beta(B_N)\vec{W}^r(\alpha,\beta) + O(z^{N+1}),$$

where the O-term is uniform on compact subsets of $\mathbb{C}$.

We call Theorem 1.3 the *leading derivatives theorem*. In the remainder of this section, we give the proof of Theorem 1.3.

1.1. **Differentiation under the integral sign.** Fix two $N \times N$ complex diagonal matrices

$$A_N = \begin{bmatrix} a_1^{(N)} & & \\ & \ddots & \\ & & a_N^{(N)} \end{bmatrix}, B_N = \begin{bmatrix} b_1^{(N)} & & \\ & \ddots & \\ & & b_N^{(N)} \end{bmatrix}.$$

The derivatives of the entire function (0.1) may be computed by differentiation under the integral sign. In particular, the Maclaurin coefficients of $I_N(z)$ are given by

$$\begin{aligned} I_N^{(d)}(0) &= N^d \int\limits_{U(N)} (\mathrm{Tr}\, A_N U B_N U^{-1})^d \mathrm{d}U \\ &= N^d \sum_{i,j} a_{i(1)} \dots a_{i(d)} b_{j(1)} \dots b_{j(d)} \int\limits_{U(N)} |u_{i(1)j(1)} \dots u_{i(d)j(d)}|^2 \mathrm{d}U, \end{aligned}$$

where the summation is over all N^{2d} pairs of functions

$$i, j : \{1, \dots, d\} \to \{1, \dots, N\}$$

and we are omitting the dependence of the a_i's and b_j's on N.

1.2. **The Weingarten function.** Given a quadruple of functions

$$i, j, i', j' : \{1, \dots, d\} \to \{1, \dots, N\},$$

consider the matrix integral

$$\int_{U(N)} u_{i(1)j(1)} \dots u_{i(d)j(d)} \overline{u_{i'(1)j'(1)} \dots u_{i'(d)j'(d)}} \mathrm{d}U. \tag{1.1}$$

This is a generalization of the integral we actually want to compute, which is the case $i = i', j = j'$.

The computation of (1.1) can be addressed using the *Weingarten convolution formula* of Collins and Śniady [4]:

$$\int_{U(N)} u_{i(1)j(1)} \dots u_{i(d)j(d)} \overline{u_{i'(1)j'(1)} \dots u_{i'(d)j'(d)}} \mathrm{d}U = \sum_{\rho,\sigma} \delta_{i,i'\rho} \delta_{j,j'\sigma} \, \mathrm{Wg}_N(\rho^{-1}\sigma),$$

where the summation is over permutations $\rho, \sigma \in S(d)$ and $\mathrm{Wg}_N(\cdot)$ is the *Weingarten function*, which is given by

$$\mathrm{Wg}_N(\pi) = \int_{U(N)} u_{11} \dots u_{dd} \overline{u_{1\pi(1)} \dots u_{d\pi(d)}} \mathrm{d}U$$

for $N \geq d$.

Given a function $i : \{1, \dots, d\} \to \{1, \dots, N\}$, we denote by $\mathrm{Stab}(i)$ the set of permutations $\pi \in S(d)$ such that $i\pi = i$, and given a permutation $\pi \in S(d)$ we denote by $\mathrm{Fix}(\pi)$ the set of functions $i : \{1, \dots, d\} \to \{1, \dots, N\}$ such that $i\pi = i$. Applying the Weingarten convolution formula to $I_N^{(d)}(0)$ with $1 \leq d \leq N$, we obtain

$$\begin{aligned} I_N^{(d)}(0) &= N^d \sum_i \sum_j a_{i(1)} \dots a_{i(d)} b_{j(1)} \dots b_{j(d)} \sum_{\rho \in \mathrm{Stab}(i)} \sum_{\sigma \in \mathrm{Stab}(j)} \mathrm{Wg}_N(\rho^{-1}\sigma) \\ &= N^d \sum_\rho \sum_\sigma \mathrm{Wg}_N(\rho^{-1}\sigma) \sum_{i \in \mathrm{Fix}(\rho)} \sum_{j \in \mathrm{Fix}(\sigma)} a_{i(1)} \dots a_{i(d)} b_{j(1)} \dots b_{j(d)} \\ &= N^d \sum_{\rho,\sigma} \mathrm{Wg}_N(\rho^{-1}\sigma) p_{t(\rho)}(A_N) p_{t(\sigma)}(B_N). \end{aligned} \tag{1.2}$$

1.3. **$1/N$-expansion of the Weingarten function.** In [23] (see also [20]), it was shown that, for any $N \geq d$, the Weingarten function admits the following (absolutely convergent) expansion in powers of $1/N$:

$$\mathrm{Wg}_N(\pi) = \frac{1}{N^d} \sum_{r=0}^{\infty} (-1)^r \frac{\vec{w}^r(\mathrm{id}, \pi)}{N^r},$$

where $\vec{w}^r(\rho, \sigma)$ denotes the number of r-step monotone walks on $S(d)$ from ρ to σ. Thus

$$\mathrm{Wg}_N(\rho^{-1}\sigma) = \frac{1}{N^d} \sum_{r=0}^{\infty} (-1)^r \frac{\vec{w}^r(\mathrm{id}, \rho^{-1}\sigma)}{N^r} = \frac{1}{N^d} \sum_{r=0}^{\infty} (-1)^r \frac{\vec{w}^r(\rho, \sigma)}{N^r}.$$

Plugging this expansion into (1.2) and changing order of summation, we arrive at

$$\begin{aligned} I_N^{(d)}(0) &= \sum_{\rho,\sigma \in S(d)} p_{t(\sigma)}(A_N) p_{t(\rho)}(B_N) \sum_{r=0}^{\infty} (-1)^r \frac{\vec{w}^r(\rho,\sigma)}{N^r} \\ &= \sum_{r=0}^{\infty} \left(-\frac{1}{N} \right)^r \sum_{\rho,\sigma \in S(d)} p_{t(\sigma)}(A_N) p_{t(\rho)}(B_N) \vec{w}^r(\rho,\sigma). \end{aligned}$$

The internal sum may be written

$$\begin{aligned} \sum_{\rho,\sigma \in S(d)} p_{t(\sigma)}(A_N) p_{t(\rho)}(B_N) \vec{w}^r(\rho,\sigma) &= \sum_{\alpha \vdash d} \sum_{\beta \vdash d} p_\alpha(A_N) p_\beta(B_N) \sum_{\rho \in C_\alpha} \sum_{\sigma \in C_\beta} \vec{w}^r(\rho,\sigma) \\ &= \sum_{\alpha \vdash d} \sum_{\beta \vdash d} p_\alpha(A_N) p_\beta(B_N) \vec{W}^r(\alpha,\beta), \end{aligned}$$

and this completes the proof of Theorem 1.3.

2. (Monotone) Hurwitz theory and the HCIZ free energy

As shown by Hurwitz in the 19th century [17], the enumeration of unrestricted walks on the symmetric group with given boundary conditions is equivalent to the enumeration of branched covers of the Riemann sphere with given singular data.

To state this precisely, consider the generating function

$$\mathbf{W} = 1 + \sum_{d=1}^{\infty} \frac{z^d}{d!} \sum_{r=0}^{\infty} \frac{t^r}{r!} \sum_{\alpha,\beta \vdash d} p_\alpha(A) p_\beta(B) W^r(\alpha,\beta),$$

where

$$A = \begin{bmatrix} a_1 & & \\ & a_2 & \\ & & \ddots \end{bmatrix}, \quad B = \begin{bmatrix} b_1 & & \\ & b_2 & \\ & & \ddots \end{bmatrix}$$

are a pair of formal infinite diagonal matrices and $W^r(\alpha,\beta)$ is the total number of r-step walks on $S(d)$ which begin in C_α and end in C_β. Thus $\mathbf{W}$ is an element of the formal power series algebra $\mathbb{Q}[[z,t,a_1,a_2,\dots,b_1,b_2,\dots]]$. Set

$$H^r(\alpha,\beta) = \left[\frac{z^d}{d!} \frac{t^r}{r!} p_\alpha(A) p_\beta(B) \right] \mathbf{H},$$

where $\mathbf{H} = \log \mathbf{W}$ and $[X]Y$ denotes the coefficient of term X in a series Y. By the exponential formula [11], the coefficient $H^r(\alpha,\beta)$ is the number of r-step walks beginning in C_α and ending in C_β whose endpoints and steps together generate a transitive subgroup of $S(d)$.

Hurwitz showed that $H^r(\alpha,\beta)/d!$ may be interpreted as a weighted count of degree d branched covers of the Riemann sphere by a compact, connected Riemann surface such that the covering map has profile α over ∞, β over 0, and simple branching over each of the rth roots of unity. According to the Riemann-Hurwitz formula, such a cover exists if and only if

$$g = \frac{r+2-\ell(\alpha)-\ell(\beta)}{2}$$

is a nonnegative integer, in which case it is the topological genus of the covering surface. We will use the notation $H^r(\alpha,\beta) = H_g(\alpha,\beta)$, with the understanding that r and g determine one another via the Riemann-Hurwitz formula.

The numbers $H_g(\alpha,\beta)$ were first considered from a modern perspective by Okounkov [24], who called them the *double Hurwitz numbers*. Proving a conjecture of Pandharipande [25] in Gromov-Witten theory, Okounkov showed that $\mathbf{H}$ is a solution of the 2D Toda lattice hierarchy of Ueno and Takasaki. It was subsequently shown by Kazarian and Lando [19] that, when combined with the ELSV formula [5], Okounkov's result implies the celebrated Kontsevich-Witten theorem relating intersection theory in moduli spaces of Riemann surfaces to the KdV hierarchy.

2.1. **Monotone Hurwitz numbers.** Mimicking the above classical construction, consider the generating function

$$\vec{\mathbf{W}} = 1 + \sum_{d=1}^{\infty} \frac{z^d}{d!} \sum_{r=0}^{\infty} t^r \sum_{\alpha,\beta \vdash d} p_\alpha(A_N) p_\beta(B_N) \vec{W}^r(\alpha,\beta)$$

enumerating *monotone* walks of all possible lengths and boundary conditions on all of the finite symmetric groups. Note that, due to the monotonicity constraint, the variable t is now an ordinary rather than exponential marker for the walk length statistic. Define the *monotone double Hurwitz numbers* by

$$\vec{H}^r(\alpha,\beta) = \left[\frac{z^d}{d!} t^r p_\alpha(A) p_\beta(B)\right] \vec{\mathbf{H}},$$

where $\vec{\mathbf{H}} = \log \vec{\mathbf{W}}$. Then $\vec{H}^r(\alpha,\beta)$ is the number of r-step monotone walks beginning in C_α and ending in C_β whose endpoints and steps together generate a transitive subgroup of $S(d)$. Alternatively, $\vec{H}^r(\alpha,\beta)/d!$ counts a combinatorially restricted subset of the branched covers counted by $H^r(\alpha,\beta)/d!$.

2.2. **The HCIZ free energy.** The monotone double Hurwitz numbers describe the Maclaurin coefficients of the *logarithm* of the HCIZ integral.

Let Ω_N denote the maximal simply connected open set in $\mathbb{C}$ which contains the origin and avoids the zeros of $I_N(z)$. This domain depends on the matrices A_N and B_N, but is always non-trivial due to the discreteness of the zeros of holomorphic functions and the fact that $I_N(0) = 1$. Let

$$\log I_N(z) = \oint_0^z \frac{I_N'(\zeta)}{I_N(\zeta)} \mathrm{d}\zeta, \quad z \in \Omega_N,$$

denote the principal branch of the logarithm of $I_N(z)$ on Ω_N. From Theorem 1.3 and the exponential formula, it follows that the Maclaurin series of the logarithm is

$$\log I_N(z) = \sum_{d=1}^{N} \frac{z^d}{d!} \sum_{r=0}^{\infty} \left(-\frac{1}{N}\right)^r \sum_{\alpha,\beta \vdash d} p_\alpha(A_N) p_\beta(B_N) \vec{H}^r(\alpha,\beta) + O(z^{N+1}),$$

where the O-term is uniform on compact subsets of Ω_N.

The Maclaurin series of $\log I_N(z)$ can be simplified using the Riemann-Hurwitz formula. We have:

$$\begin{aligned}\log I_N(z) &= \sum_{d=1}^{N} \frac{z^d}{d!} \sum_{\alpha,\beta\vdash d} p_\alpha(A_N)p_\beta(B_N) \sum_{r=0}^{\infty} \left(-\frac{1}{N}\right)^r \vec{H}^r(\alpha,\beta) + O(z^{N+1}) \\ &= \sum_{d=1}^{N} \frac{z^d}{d!} \sum_{\alpha,\beta\vdash d} p_\alpha(A_N)p_\beta(B_N) \sum_{g=0}^{\infty} \left(-\frac{1}{N}\right)^{2g-2+\ell(\alpha)+\ell(\beta)} \vec{H}_g(\alpha,\beta) + O(z^{N+1}) \\ &= N^2 \sum_{d=1}^{N} \frac{z^d}{d!} \sum_{g=0}^{\infty} \frac{1}{N^{2g}} \sum_{\alpha,\beta\vdash d} (-1)^{\ell(\alpha)+\ell(\beta)} \frac{p_\alpha(A_N)}{N^{\ell(\alpha)}} \frac{p_\beta(B_N)}{N^{\ell(\beta)}} \vec{H}_g(\alpha,\beta) + O(z^{N+1})\end{aligned}$$

From this computation, it follows that the Maclaurin series of the *HCIZ free energy*, i.e. the holomorphic function

$$F_N(z) = N^{-2} \log I_N(z), \quad z \in \Omega_N,$$

has the form

$$F_N(z) = \sum_{d=1}^{N} \frac{z^d}{d!} \sum_{g=0}^{\infty} \frac{1}{N^{2g}} \sum_{\alpha,\beta\vdash d} (-1)^{\ell(\alpha)+\ell(\beta)} \frac{p_\alpha(A_N)}{N^{\ell(\alpha)}} \frac{p_\beta(B_N)}{N^{\ell(\beta)}} \vec{H}_g(\alpha,\beta) + O(z^{N+1}),$$

with the O-term uniform on compact subsets of Ω_N. Thus, parameterizing the monotone double Hurwitz numbers by g instead of r, we arrive at a description of $F_N(z)$ which is well-poised for an $N \to \infty$ asymptotic analysis.

2.3. **Asymptotic expansion of Maclaurin coefficients.** We will now consider the asymptotic behaviour of the Maclaurin coefficients of $F_N(z)$ as $N \to \infty$, when A_N, B_N vary regularly with N.

Suppose there exists a nonnegative number M, a nonnegative integer h, and two complex sequences ϕ_k, ψ_k such that:

(1) $\|A_N\|, \|B_N\| \leq M$ for all $N \geq 1$;
(2) For each $k \geq 1$,

$$\frac{1}{N}\operatorname{Tr}(A_N^k) = \phi_k + o\left(\frac{1}{N^{2h}}\right), \quad \frac{1}{N}\operatorname{Tr}(B_N^k) = \psi_k + o\left(\frac{1}{N^{2h}}\right)$$

as $N \to \infty$.

We will summarize these conditions by saying that A_N, B_N are (M,h)-regular with limit moments ϕ_k, ψ_k.

Theorem 2.1. *Suppose that A_N, B_N are (M,h)-regular with limit moments ϕ_k, ψ_k. Then, for each $d \geq 1$, we have*

$$F_N^{(d)}(0) = \sum_{g=0}^{h} \frac{C_{g,d}}{N^{2g}} + o\left(\frac{1}{N^{2h}}\right)$$

as $N \to \infty$, where

$$C_{g,d} = \sum_{\alpha,\beta \vdash d} (-1)^{\ell(\alpha)+\ell(\beta)} \phi_\alpha \psi_\beta \vec{H}_g(\alpha,\beta),$$

and

$$\phi_\alpha = \prod_{i=1}^{\ell(\alpha)} \phi_{\alpha_i}, \quad \psi_\beta = \prod_{j=1}^{\ell(\beta)} \phi_{\beta_j}.$$

Proof. For $1 \leq d \leq N$, we have

$$F_N^{(d)}(0) = \sum_{g=0}^{\infty} \frac{C_{g,d,N}}{N^{2g}},$$

where

$$C_{g,d,N} = \sum_{\alpha,\beta \vdash d} (-1)^{\ell(\alpha)+\ell(\beta)} \frac{p_\alpha(A_N)}{N^{\ell(\alpha)}} \frac{p_\beta(B_N)}{N^{\ell(\beta)}} \vec{H}_g(\alpha,\beta).$$

Since $\|A_N\|, \|B_N\| \leq M$ for all $N \geq 1$, we have

$$|C_{g,d,N}| \leq M^{2d} \sum_{\alpha,\beta \vdash d} \vec{H}_g(\alpha,\beta)$$

for all $N \geq 1$. Since $\vec{H}_g(\alpha,\beta)$ counts certain solutions of the equation

$$\sigma = \rho\tau_1 \dots \tau_r$$

in $S(d)$, with $r = 2g - 2 + \ell(\alpha) + \ell(\beta)$, we have

$$\vec{H}_g(\alpha,\beta) \leq (d!)^{2g+2d}$$

for all $\alpha, \beta \vdash d$. Consequently,

$$|C_{g,d,N}| \leq (d!p(d)M)^{2d}(d!)^{2g},$$

where $p(d)$ is the number of Young diagrams with d cells. Moreover, since A_N, B_N are h-regular with limit moments ϕ_k, ψ_k, we have

$$C_{g,d,N} = C_{g,d} + o\left(\frac{1}{N^{2h}}\right)$$

as $N \to \infty$. We thus have

$$\begin{aligned} F_N^{(d)}(0) &= \sum_{g=0}^{h} \frac{C_{g,d,N}}{N^{2g}} + \sum_{g=h+1}^{\infty} \frac{C_{g,d,N}}{N^{2g}} \\ &= \sum_{g=0}^{h} \frac{C_{g,d}}{N^{2g}} + o\left(\frac{1}{N^{2h}}\right) + O\left(\frac{1}{N^{2h+2}}\right) \end{aligned}$$

as $N \to \infty$, which proves the claim. □

Remark 2.2. The convergence of $F_N^{(d)}(0)$ under the above hypotheses was first stated by Itzykson and Zuber [18], and proved by Collins [2]. Collins obtained the limit of $F_N^{(d)}(0)$ as a double sum over $S(d)$, whereas we present the same limit as a double sum over partitions of d.

Remark 2.3. To the best of our knowledge, Theorem 2.1 is the first result which addresses the sub-leading asymptotics of $F_N^{(d)}(0)$, clearly showing the emergence of a topological expansion in this context.

3. Genus-specific generating functions

The results of the previous section suggest the following.

Conjecture 3.1. *Under the hypotheses of Theorem 2.1, there exists a domain* $\Omega \subseteq \mathbb{C}$ *containing* $z = 0$ *such that*

$$F_N(z) = \sum_{g=0}^{h} \frac{C_g(z)}{N^{2g}} + o\left(\frac{1}{N^{2h}}\right)$$

uniformly on compact subsets of Ω*, where*

$$C_g(z) = \sum_{d=1}^{\infty} \frac{z^d}{d!} C_{g,d}$$

and $C_{g,d}$ *is as in Theorem 2.1.*

Conjecture 3.1 is the complex-variables version of a conjecture formulated by Matytsin in [21]. Matytsin considers the asymptotics of $F_N(z)$ with z real and A_N, B_N real diagonal. He posits the existence of a $1/N^2$ expansion of the HCIZ free energy, and provides a compelling physical argument in favour of the existence of the first term. For a rigorous approach to Matytsin's results based on large deviation theory, see the work of Guionnet and Zeitouni [15] and Guionnet [13]. Here we will take a different approach, and prove directly that the power series $C_g(z)$, which are genus-specific generating functions for the monotone double Hurwitz numbers, are absolutely summable. This will allow us to use techniques from complex function theory to compare $F_N(z)$ to $C_0(z)$, which is a candidate for the large N limit of $F_N(z)$. The end result of this is that we can prove $F_N(z)$ converges to $C_0(z)$ uniformly on compact subsets of a complex domain under a local non-vanishing hypothesis on $I_N(z)$.

3.1. **Monotone simple Hurwitz numbers.** Consider the genus-specific generating functions

$$\vec{\mathbf{S}}_g = \sum_{d=1}^{\infty} \frac{z^d}{d!} \vec{H}_{g,d}$$

for the monotone *simple* Hurwitz numbers

$$\vec{H}_{g,d} = \vec{H}_g(1^d, 1^d).$$

The monotone simple Hurwitz number $\vec{H}_{g,d}$ counts monotone loops of length $r = 2g - 2 + 2d$, based at any given point of $S(d)$, whose steps generate a transitive subgroup of $S(d)$.

According to [9, Theorem 1.4], for any $g \geq 2$ we have

$$\vec{\mathbf{S}}_g = \frac{\zeta(1-2g)}{2-2g} + \frac{1}{(1-6s)^{2g-2}} \sum_{r=0}^{3g-3} \sum_{\mu \vdash r} \frac{c_{g,\mu}(6s)^{\ell(\mu)}}{(1-6s)^{\ell(\mu)}},$$

where ζ is the Riemann zeta function, the $c_{g,\mu}$'s are rational numbers, and s is the unique solution of the functional equation

$$s = z(1-2s)^{-2}$$

in the formal power series algebra $\mathbb{Q}[[z]]$. This equation may be solved by Lagrange inversion, yielding the solution

$$s = \sum_{n=1}^{\infty} \frac{2^{n-1}}{n} \binom{3n-2}{n-1} z^n.$$

Since

$$s' = \sum_{n=0}^{\infty} \binom{3n+1}{n} (2z)^n = {}_2F_1\left(\frac{2}{3}, \frac{4}{3}, \frac{3}{2}; \frac{27}{2} z\right),$$

where ${}_2F_1(a,b,c;z)$ is the Gauss hypergeometric function, $\vec{\mathbf{S}}_g$ extends to a holomorphic function of z on the domain $\mathbb{C} \backslash [z_c, \infty)$, where

$$z_c = \frac{2}{27}.$$

In the case $g = 0$, [8, Theorem 1.1] yields the exact formula

$$\frac{\vec{H}_{0,d}}{d!} = \frac{2^{d-1}}{d^2(d-1)} \binom{3d-3}{d-1},$$

so that, using Stirling's formula, we obtain $\frac{2}{27}$ as the radius of convergence of $\vec{\mathbf{S}}_0$. Moreover, since $\vec{\mathbf{S}}_0$ has positive coefficients, Pringsheim's theorem (see [28, §7.21]) guarantees that $z_c = \frac{2}{27}$ is a singularity of $\vec{\mathbf{S}}_0$.

We have thus shown that:

Theorem 3.2. *The series $\vec{\mathbf{S}}_g$, $g \geq 0$, have a common dominant singularity at the critical point $z_c = \frac{2}{27}$.*

Remark 3.3. The $g = 0$ case of Theorem 3.2 was obtained by Zinn-Justin in [29] using the Toda lattice equations. For a combinatorial solution of the Toda equations encompassing those of Okounkov and Zinn-Justin, see [10].

Remark 3.4. P. Di Francesco has pointed out to us that the critical value $z_c = \frac{2}{27}$ also appears in the enumeration of finite groups, see [26].

3.2. **Monotone double Hurwitz numbers.** Consider now the full genus-specific generating functions

$$\vec{\mathbf{H}}_g = \sum_{d=1}^{\infty} \frac{z^d}{d!} \sum_{\alpha,\beta \vdash d} \vec{H}_g(\alpha, \beta), \quad g \geq 0.$$

Obviously,

$$\sum_{\alpha,\beta\vdash d} \vec{H}_g(\alpha,\beta) \geq \vec{H}_{g,d},$$

so the radius of convergence of $\vec{\mathbf{H}}_g$ is at most the radius of convergence of $\vec{\mathbf{S}}_g$.

We now explain how a lower bound on the radius of convergence of $\vec{\mathbf{H}}_g$ follows from the study of a refinement of the monotone double Hurwitz number $\vec{H}_g(\alpha,\beta)$. Given a positive integer c, let $\vec{H}_g(\alpha,\beta;c)$ denote the number of walks counted by $\vec{H}_g(\alpha,\beta)$ whose steps have c distinct labels. The following inequality is obtained in [12]:

$$\sum_{c=2}^{d} 3^c \vec{H}_g(1^d,1^d;c) \leq \sum_{\alpha,\beta\vdash d} \vec{H}_g(\alpha,\beta) \leq \sum_{c=2}^{d} 4^c \vec{H}_g(1^d,1^d;c).$$

The proof is combinatorial, and makes use of an action of the symmetric group $S(r)$ on the set of walks counted by the classical double Hurwitz number $H_g(\alpha,\beta)$. From this inequality and the definition of $\vec{H}_g(\alpha,\beta;c)$, we obtain

$$\sum_{\alpha,\beta\vdash d} \vec{H}_g(\alpha,\beta) \leq 4^{d-1} \sum_{c=2}^{d} \vec{H}_g(1^d,1^d;c) = 4^{d-1}\vec{H}_{g,d},$$

which implies that the radius of convergence of $\vec{\mathbf{H}}_g$ is at least one quarter the radius of convergence of $\vec{\mathbf{S}}_g$. Combining this with Theorem 3.2, we have:

Theorem 3.5. *For each $g \geq 0$, the series $\vec{\mathbf{H}}_g$ is absolutely summable, and has radius of convergence at least $\frac{1}{54}$ and at most $\frac{2}{27}$.*

Remark 3.6. The results of [3] also imply the absolute summability of $\vec{\mathbf{H}}_0$, but without effective bounds on the radius of convergence.

3.3. **Convergence of the HCIZ free energy.** Let A_N, B_N be (M,h)-regular with limit moments ϕ_k, ψ_k, respectively. Then, by Theorem 3.5, the series $C_g(z)$ defined in Conjecture 3.1 is absolutely summable, with radius of convergence at least $\frac{1}{54M^2}$.

In this section, we will prove that $F_N(z) \to C_0(z)$ uniformly on compact subsets of a complex domain, under an assumption on the large N behaviour of the zeros of $I_N(z)$. Namely, we will assume the existence of a positive number R such that $I_N(z)$ is non-vanishing on the open disc $D(0,R)$ for all but finitely many N. This hypothesis guarantees that the domain of holomorphy of $F_N(z)$ contains $D(0,R)$ for N sufficiently large. Let

$$r = \min\left\{R, \frac{1}{54M^2}\right\}.$$

Theorem 3.7. *$F_N(z) \to C_0(z)$ uniformly on compact subsets of $D(0,r)$.*

Proof. First, we show that the claim holds pointwise. The proof is a two circles argument. Let $z_0 \in D(0,r)$ and $\varepsilon > 0$ be arbitrary. Choose r_1, r_2 such that

$$|z_0| < r_1 < r_2 < r.$$

For N sufficiently large, we have

$$|F_N(z_0) - C_0(z_0)| \leq \sum_{d=1}^{\infty} |F_N^{(d)}(0) - C_{0,d}| \frac{|z_0|^d}{d!}.$$

By Cauchy's inequality,

$$\frac{1}{d!}|F_N^{(d)}(0) - C_{0,d}| \leq \frac{\|F_N - C_0\|_{r_1}}{r_1^d},$$

where $\|\cdot\|_{r_1}$ denotes sup-norm over the circle of radius r_1. We thus have

$$|F_N(z_0) - C_0(z_0)| \leq \sum_{d=1}^{E} |F_N^{(d)}(0) - C_{0,d}| \frac{|z_0|^d}{d!} + \frac{\|F_N\|_{r_1} + \|C_0\|_{r_1}}{1 - \frac{|z_0|}{r_1}} \left(\frac{|z_0|}{r_1} \right)^{E+1}$$

for any positive integer E.

We will now bound the error term independently of N. Since

$$\begin{aligned} |I_N(z)| &\leq \int\limits_{U(N)} e^{|z|N|\operatorname{Tr}(A_N U B_N U^{-1})|} \mathrm{d}U \\ &\leq \int\limits_{U(N)} e^{|z|M^2 N \sum_{i,j=1}^N |u_{ij}|^2} \mathrm{d}U \\ &\leq e^{M^2N^2|z|} \end{aligned}$$

for all $z \in \mathbb{C}$, where the last line follows from the fact that $(|u_{ij}|^2)$ is a doubly stochastic matrix, the inequality

$$\Re F_N(z) \leq M^2|z|$$

holds for all $z \in \Omega_N$, the domain of holomorphy of $F_N(z)$. Combining this with the Borel-Carathéodory inequality, which bounds the sup-norm of an analytic function on a circle in terms of the maximum of its real part on a circle of larger radius (see e.g. [28, §5.5], we have

$$\|F_N\|_{r_1} \leq \frac{2r_1}{r_2 - r_1} \sup_{|z|=r_2} \Re F_N(z) \leq \frac{2M^2 r_1 r_2}{r_2 - r_1}.$$

Returning to our estimate on $|F_N(z_0) - C_0(z_0)|$, we now have the inequality

$$|F_N(z_0) - C_0(z_0)| \leq \sum_{d=1}^{E} |F_N^{(d)}(0) - C_{0,d}| \frac{|z_0|^d}{d!} + \frac{\frac{2M^2 r_1 r_2}{r_2 - r_1} + \|C_0\|_{r_1}}{1 - \frac{|z_0|}{r_1}} \left(\frac{|z_0|}{r_1} \right)^{E+1}$$

for all N sufficiently large and all $E \geq 1$. Choosing E_0 large enough so that

$$\frac{\frac{2M^2 r_1 r_2}{r_2 - r_1} + \|C_0\|_{r_1}}{1 - \frac{|z_0|}{r_1}} \left(\frac{|z_0|}{r_1} \right)^{E+1} < \frac{\varepsilon}{2},$$

and subsequently choosing N_0 large enough so that

$$\sum_{d=1}^{E_0} |F_{N_0}^{(d)}(0) - C_{0,d}| \frac{|z_0|^d}{d!} < \frac{\varepsilon}{2},$$

we obtain that

$$|F_N(z_0) - C_0(z_0)| < \varepsilon$$

for all $N \geq N_0$. This completes the proof.

We now explain how the mode of convergence can be boosted from pointwise to uniform on compact subsets of $D(0,r)$. It is easy to check that

$$\sup_N \|F_N\|_K < \infty$$

for each compact set $K \subset D(0,r)$; for example, one could prove this holds for closed discs by using the Borel-Carathédory inequality again. Thus $\{F_N\}$ is a locally uniformly bounded family on $D(0,r)$, and Vitali's theorem [28, §5.21] implies pointwise and uniform-on-compact convergence are equivalent. □

3.4. **Remarks on the non-vanishing hypothesis.** Our proof of Theorem 3.7 is conditional on knowing that, as $N \to \infty$, the zeros of $I_N(z)$ do not encroach on $z = 0$. Therefore it is of significant interest to determine sufficient hypotheses on the matrix sequences A_N and B_N which ensure that this condition holds.

Determining the exact locations of the zeros of $I_N(z)$ seems to be a difficult problem in general. Let us give one example where it can be solved explicitly. Suppose that A_N and B_N are real diagonal with distinct eigenvalues

$$a_1 < \cdots < a_N, \quad b_1 < \cdots < b_N,$$

and suppose further that the eigenvalues of A_N form an arithmetic progression with constant difference $\hbar > 0$. In this case, all determinants in the HCIZ formula (see [18, Equation 3.28]) are Vandermonde determinants, and one computes that the zeros of $I_N(z)$ are the pure imaginary points

$$z = \frac{2k\pi}{N\hbar(b_j - b_i)}\mathbf{i}, \quad 1 \leq i < j \leq N, \quad k \in \mathbb{Z}.$$

Thus one has a sort of Lee-Yang theorem in this special case. If the eigenvalues of A_N and B_N are confined to a fixed interval $[-M, M]$ for all N, then $\hbar$ is of order $1/N$ and the above formula shows that the zeros of $I_N(z)$ remain at bounded distance from the origin.

Concerning the verification of the non-vanishing hypothesis more generally, we do not know much at present. It might be that the Toda equations could be of help here, but we have not pursued this seriously.

References

1. P. Biane, *Parking functions of types A and B*, Electron. J. Combin. **9** (2002). #N7.
2. B. Collins, *Moments and cumulants of polynomial random variables on unitary groups, the Itzykson-Zuber integral, and free probability*, Int. Math. Res. Not. IMRN. **17** (2003), 954-982.
3. B. Collins, A. Guionnet, E. Maurel-Segala, *Asymptotics of unitary and orthogonal matrix integrals*, Adv. Math. **222** (2009), 172-215.
4. B. Collins, P. Śniady, *Integration with respect to the Haar measure on unitary, orthogonal and symplectic group*, Comm. Math. Phys, **264** (2006), 773-795.
5. T. Ekedahl, S. Lando, M. Shapiro, A. Vainshtein, *Hurwitz numbers and intersections on moduli spaces of curves*, Invent. Math. **146** (2001), 297-327.
6. L. Erdős, H.-T. Yau, *Universality of local spectral statistics of random matrices*, Bulletin (new series) of the American Mathematical Society **49**(3) (2012), 377-414.

7. D. A. Gewurz, F. Merola, *Some factorisations counted by Catalan numbers*, European J. Combin. **27** (2006), 990-994.
8. I. P. Goulden, M. Guay-Paquet, J. Novak, *Monotone Hurwitz numbers in genus zero*, Canad. J. Math. **65** (2013), 1020-1042.
9. I. P. Goulden, M. Guay-Paquet, J. Novak, *Polynomiality of monotone Hurwitz numbers in higher genera*, Adv. Math. **238** (2013), 1-23.
10. I. P. Goulden, M. Guay-Paquet, J. Novak, *Toda equations and piecewise polynomiality for mixed double Hurwitz numbers*, submitted.
11. I. P. Goulden, D. M. Jackson, *Combinatorial Enumeration*, John Wiley and Sons, New York, 1983 (reprinted by Dover, 2004).
12. M. Guay-Paquet, J. Novak, *A self-interacting random walk on the symmetric group*, in preparation.
13. A. Guionnet, *First order asymptotics of matrix integrals; a rigorous approach towards the understanding of matrix models*, Communications in Mathematical Physics **244** (2004), 527-569.
14. A. Guionnet, *Large deviations and stochastic calculus for large random matrices*, Probability Surveys **1** (2004), 72-172.
15. A. Guionnet, O. Zeitouni, *Large deviations asymptotics for spherical integrals*, Journal of Functional Analysis **188** (2002), 461-515.
16. Harish-Chandra, *Differential operators on a semisimple Lie algebra*, American Journal of Mathematics **79** (1957), 87-120.
17. A. Hurwitz, *Ueber Riemann'sche Flächen mit gegebenen Verzweigegungspunkten*, Mathematische Annalen **39** (1891), 1-60.
18. C. Itzykson, J.-B. Zuber, *The planar approximation. II*, Journal of Mathematical Physics **21**(3) (1980), 411-421.
19. M. E. Kazarian, S. K. Lando, *An algebro-geometric proof of Witten's conjecture*, Journal of the AMS **20**(4) (2007), 1079-1089.
20. S. Matsumoto, J. Novak, *Jucys-Murphy elements and unitary matrix integrals*, Int. Math. Res. Not. IMRN **2** (2013), 362-297.
21. A. Matytsin, *On the large-N limit of the Itzykson-Zuber integral*, Nuclear Physics B **411** (1994), 805-820.
22. J. Murray, *Generators for the centre of the group algebra of a symmetric group*, J. Algebra **271** (2004), 725-748.
23. J. Novak, *Jucys-Murphy elements and the unitary Weingarten function*, Banach Center Publications **89** (2010), 231-235.
24. A. Okounkov, *Toda equations for Hurwitz numbers*, Mathematical Research Letters **7** (2000), 447-453.
25. R. Pandharipande, *The Toda equations and the Gromov-Witten theory of the Riemann sphere*, Letters in Mathematical Physics **53** (2000), 59-74.
26. L. Pyber, *Enumerating finite groups of given order*, Annals of Mathematics **137**(1) (1993), 203-220.
27. R. P. Stanley, *Parking functions and noncrossing partitions*, Electron. J. Combin. **4** (1997), #R20.
28. E. C. Titchmarsh, *The Theory of Functions, Second Edition*, Oxford University Press, 1939.
29. P. Zinn-Justin, *HCIZ integral and 2D Toda lattice hierarchy*, Nuclear Physics B **634** [FS] (2002), 417-432.
30. P. Zinn-Justin, J.-B. Zuber, *On some integrals over the $U(N)$ unitary group and their large N limit*, Journal of Physics A: Mathematical and General **36** (2003), 3173-3193.

Department of Combinatorics & Optimization. University of Waterloo, Canada
E-mail address: `ipgoulden@uwaterloo.ca`

Department of Combinatorics & Optimization. University of Waterloo, Canada
E-mail address: `mguaypaq@uwaterloo.ca`

Department of Mathematics. Massachusetts Institute of Technology, USA
E-mail address: `jnovak@math.mit.edu`