J. Eur. Math. Soc. 26, 1851–1897 (2024) DOI 10.4171/JEMS/1315

© 2023 European Mathematical Society
Published by EMS Press and licensed under a CC BY 4.0 license

Benoît Collins · Razvan Gurau · Luca Lionni

The tensor Harish-Chandra–Itzykson–Zuber integral I: Weingarten calculus and a generalization of monotone Hurwitz numbers

Received November 3, 2020; revised October 26, 2021

Abstract. We study a generalization of the Harish-Chandra–Itzykson–Zuber integral to tensors and its expansion in terms of trace invariants of the two external tensors. This gives rise to natural generalizations of monotone double Hurwitz numbers, which count certain families of constellations. We find an expression of these numbers in terms of monotone simple Hurwitz numbers, thereby also providing expressions for monotone double Hurwitz numbers of arbitrary genus in terms of the single ones. We give an interpretation of the different combinatorial quantities at play in terms of enumeration of nodal surfaces. In particular, our generalization of Hurwitz numbers is shown to count certain isomorphism classes of branched coverings of a bouquet of D 2-spheres that touch at one common non-branch node.

Keywords. Harish-Chandra–Itzykson–Zuber integral, tensor HCIZ integral, Weingarten calculus, monotone Hurwitz numbers, enumerative geometry, branched coverings, ramified coverings, nodal surfaces, bouquet of spheres, constellations

1. Introduction

The problem. Our goal is to explore the following generalization of the HCIZ integral [30, 32]:

$$t \mapsto I_{D,N}(t, A, B) = \mathbb{E}_U\big[\exp(t\,\mathrm{Tr}(AUBU^*))\big]. \tag{1.1}$$

We will be particularly interested in the expansion of its logarithm,

$$C_{D,N}(t, A, B) = \log I_{D,N}(t, A, B),$$

Benoît Collins: Department of Mathematics, Graduate School of Science, Kyoto University, Sakyo, Kyoto 606-8502 Japan; collins@math.kyoto-u.ac.jp

Razvan Gurau: Institut für Theoretische Physik, Universität Heidelberg, 69120 Heidelberg, Germany; gurau@thphys.uni-heidelberg.de

Luca Lionni: IMAPP, Radboud University, 6525 AJ Nijmegen, The Netherlands; lionni@thphys.uni-heidelberg.de

Mathematics Subject Classification (2020): Primary 05A05; Secondary 05A17, 05C10, 15B52

as a power series in t and a Laurent series in N and in its behavior for large N. For $D = 1$ we take U to be a unitary $N \times N$ matrix in $U(N)$ ($U^* = U^{-1}$, where star denotes the adjoint), A and B self-adjoint $N \times N$ matrices and $\mathbb{E}_U$ the expectation with respect to the Haar measure; that is, for $D = 1$, (1.1) is the usual HCIZ integral [30, 32], here denoted by $I_{1,N}(t, A, B)$.

In this paper we are interested in the setup where U is a tensor product of $D \geq 2$ unitary matrices, $U = U^{(1)} \otimes \cdots \otimes U^{(D)}$ with $U^{(c)} \in U(N)$, $\mathbb{E}_U$ is the expectation with respect to the tensor product of D Haar measures and A, B are self-adjoint $N^D \times N^D$ matrices called the *external tensors*. In this case we will call (1.1) the *tensor HCIZ integral*. It is also written as

$$I_{D,N}(t, A, B) = e^{C_{D,N}(t,A,B)} = \int [dU]\, e^{t\,\mathrm{Tr}(AUBU^*)}. \tag{1.2}$$

Let us note that if the tensors A, B belong to the Lie algebra of $U(N)^{\otimes D}$ and are generic (i.e. in the interior of a Weyl chamber), the integral $I_{D,N}(t, A, B)$ admits an exact determinantal formula as per Harish-Chandra's general results. In the particular case of $D = 1$, this statement is important because it allows handling all multiplicity-free self-adjoint matrices A, B. For $D \geq 2$ however, the Lie algebra is smaller and the problem we are considering is much more general.

Motivations. The Kontsevich integral [33] for a self-adjoint $N \times N$ matrix X,

$$\int dX\, e^{-\frac{1}{2}\mathrm{Tr}(XBX)+\imath\frac{g}{3}\mathrm{Tr}(X^3)}, \quad dX = \prod_{a,b} dX_{ab},$$

with B a fixed $N \times N$ matrix is the prototype of a non-invariant probability distribution for a random matrix. This integral plays a crucial role in two-dimensional quantum gravity [16]. The Grosse–Wulkenhaar model [27] is obtained by replacing the cubic potential with a quartic one, and for a specific choice of B this model is a ϕ^4 quantum field theory on the non-commutative Moyal space [42] expressed in a matrix base [17, 19]. Such models can be generalized [3, 10, 20] to rank D complex tensors T (with components denoted by $T_{i^1\ldots i^D}$) transforming according to the D-fundamental representation of the unitary group ($T'_{i^1\ldots i^D} = \sum_j U^{(1)}_{i^1 j^1} \ldots U^{(D)}_{i^D j^D} T_{j^1\ldots j^D}$). One is then interested in partition functions of the type

$$Z = \int dT\, d\bar{T}\, e^{-TB\bar{T}+V(T,\bar{T})}, \quad TB\bar{T} = \sum_{i,j} T_{i^1\ldots i^D} B_{i^1\ldots i^D, j^1\ldots j^D} \overline{T_{j^1\ldots j^D}},$$

where $\bar{T}$ denotes the dual of T (with complex conjugate entries $\overline{T_{i^1\ldots i^D}}$ and transforming according to the conjugate representation) and B some $N^D \times N^D$ matrix. The crucial point is that the perturbation $V(T, \bar{T})$ is taken to be *invariant* under the action of $U(N)^{\otimes D}$.

A striking feature of both the Kontsevich integral and its generalizations involving random tensors is that one considers a non-invariant quadratic part and an invariant interaction. In order to study the interplay between these two, one can average over the unitary group:

$$Z = \int dT\, d\bar{T}\, e^{V(T,\bar{T})} \int [dU]\, e^{-TUBU^\star\bar{T}}.$$

The integral over U is then just a particular case of (1.2) for A the tensor product of T and its dual.

More generally, the HCIZ integral is extensively used in $D = 1$ in random matrix models with non-invariant probability distributions, such as the two-matrix models [21, 32, 38], and matrix models with an external source [4, 6–8, 44, 45], to cite just a few references. It is also central in studying the law of Gaussian Wishart matrices and non-centered Gaussian Wigner matrices [29]. The study of the tensor HCIZ integral (1.1) is justified by the generalization of these problems to random tensors. For example, it is natural in quantum information to study the sum of independent random tensors, so as soon as they have a $U(N)^{\otimes D}$-conjugation invariance, we expect that our results are a necessary preliminary to the study of such asymptotic models.

Another application of the HCIZ integral is as a generating function for the monotone double Hurwitz numbers [25, 26]. Hurwitz numbers count n-sheet coverings of the Riemann sphere by a Riemann surface of a certain fixed genus, where one branch point, for instance 0, is allowed arbitrary but fixed ramification, and all the other branch points are only allowed simple ramifications. Double Hurwitz numbers are such that not one but two points on the sphere, say 0 and ∞, are allowed non-simple ramifications [23–26]. Monotone single and double Hurwitz numbers are such that only a subset of the possible coverings are allowed. These numbers appear in $D = 1$ as coefficients when expanding the HCIZ integral in terms of the trace invariants of the two external matrices [25, 26].

It is this last aspect on which the emphasis is put in the present paper: we expand the logarithm of the tensor HCIZ integral in terms of the trace invariants of the two external tensors and study the $1/N$ expansion of the coefficients. This provides higher order generalizations of monotone double Hurwitz numbers. We provide a detailed study of the geometrical interpretation of these numbers and their relation to enumerations of branched coverings of a bouquet of D spheres.

Final comments. Before proceeding let us comment some more on our model:

Different N's. The generalization to the case of different dimensions, $U^{(c)} \in U(N_c)$, is straightforward.

The $D = 1$ case. For $D = 1$ we get the HCIZ integral which is a Fourier transform of the $U(N)$-invariant probability measure concentrated on the orbit of the matrix B. In the general case the same holds true, but for the smaller symmetry group $U(N)^{\otimes D}$ instead of $U(N^D)$.

Variants. Other models might be relevant, such as

$$t \to \log \mathbb{E}_U\big[\exp(t\,\mathrm{Re}(\bar{B}UA))\big],$$

where A, B are tensors. For $D = 1$ this model boils down to the Brézin–Gross–Witten (BGW) integral

$$t \to \log \mathbb{E}_U\big[\exp(t\,\mathrm{Tr}(AU + U^*B^*))\big].$$

which was widely studied in the literature. For an optimal analytic result both for HCIZ and BGW in the $D = 1$ setup, we refer to [40]. It turns out that the combina-

torics is a bit different and slightly more involved, so we will turn to this model in subsequent work.

Other groups. Similar results can be derived for the orthogonal or symplectic group. The initial theory does not change substantially, but the graphical interpretation does. We also leave this for future work.

Plan of the paper. The notations and prerequisites on Weingarten calculus, constellations and cumulants are gathered in Sec. 2, where the reader will also find, in Prop. 2.5, an expression for the moments of the tensor HCIZ integral which follows directly from the definitions.

The study of the cumulants of the tensor HCIZ integral is more involved. They can be expressed in terms of a *cumulant Weingarten function*, defined in Sec. 3.1 and expressed in Sec. 3.2 as a series in powers of $1/N$ whose coefficients p_C count certain transitive factorizations of D-uplets of permutations. The rest of the paper is dedicated to the study and interpretation of the coefficients p_C.

Sec. 4 contains our main theorem, Thm. 4.1. This theorem expresses the coefficients p_C as sums over partitions satisfying certain conditions. In this form we are able to compute p_C at leading order in $1/N$, that is, we identify the smallest exponent of $1/N$ with non-vanishing contribution to p_C and compute this contribution.

In $D = 1$, the coefficients p_C are related to *monotone double Hurwitz numbers*, as detailed in Sec. 4.3. These numbers are known to count certain isomorphism classes of connected branched coverings of the Riemann sphere. For $D > 1$, the coefficients p_C lead to a generalization of monotone double Hurwitz numbers, and one may wonder whether these numbers have a natural interpretation as counting certain branched coverings.

This question is addressed in Sec. 5. After introducing nodal surfaces in Sec. 5.1 we show in Sec. 5.2 that the generalized Hurwitz numbers of Sec. 4.3 count certain *connected branched coverings of D 2-spheres that "touch" at one common node* (*a bouquet of D 2-spheres*). This provides (see Sec. 5.3) a geometric interpretation for the combinatorial formulas of Thm. 4.1 and recasts the sums over partitions defining p_C as sums over certain nodal surfaces whose nodes are weighted with monotone single Hurwitz numbers.

2. Prerequisites and direct results

2.1. Notations

Indices ranging from 1 to N will be denoted by a_1, a_2, b_1, b_2 etc. Let S_n be the group of permutations of n elements and $S_n^* = S_n \setminus \{\mathrm{id}\}$. For $\sigma \in S_n$, $\#(\sigma)$ denotes the number of disjoint cycles of σ, and $\|\sigma\|$ the number of transpositions of σ (i.e. the minimal number of transpositions required to obtain σ).[1] These quantities satisfy

$$\#(\sigma) + \|\sigma\| = n. \tag{2.1}$$

[1] The common notation would be $|\sigma|$, but we want to avoid confusion with the number of blocks of a partition.

We denote by $(\rho_1, \dots, \rho_k) = (\rho_i)_{1 \le i \le k}$ with $\rho_i \in S_n$, or sometimes $\hat{\rho}$, an ordered sequence of k permutations, that is, a *constellation* (see Sec. 2.5).

In this paper, we will deal with indices, permutations, and sequences of permutations bearing a *color* $c \in \{1, \dots, D\}$. The color is indicated in superscript or subscript: a_1^c are indices, σ_c are permutations etc. D-uplets will be written in bold, for instance $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_D)$ with $\sigma_c \in S_n$ is a D-uple of permutations ($\boldsymbol{S}_n$ denotes the set of $(n!)^D$ such D-uplets), and $\hat{\boldsymbol{\rho}}$ is a D-uple of constellations. For $\boldsymbol{\sigma}, \boldsymbol{\tau} \in \boldsymbol{S}_n$, we denote by $\boldsymbol{\nu} = \boldsymbol{\sigma}\boldsymbol{\tau}^{-1}$ the D-uple of permutations $\boldsymbol{\nu} = (\sigma_1 \tau_1^{-1}, \dots, \sigma_D \tau_D^{-1})$.

We denote by π, π', π_1, π_2 etc. partitions of the set $\{1, \dots, n\}$, and by $\mathcal{P}(n)$ the set of all such partitions. The notation $|\pi|$ is used for the number of blocks of π, while $B \in \pi$ denotes the blocks, and $|B|$ the cardinality of the block B. We denote by $\le$ the refinement partial order: $\pi' \le \pi$ if each block of π' is a subset of some block of π. Furthermore, $\vee$ denotes joining of partitions: $\pi \vee \pi'$ is the finest partition which is coarser than both π and π'. Let 1_n be the one-block partition of $\{1, \dots, n\}$.

The partition induced by the transitivity classes of the permutation ν (i.e. the disjoint cycles of ν) is denoted by $\Pi(\nu)$, hence $|\Pi(\nu)| = \#(\nu)$. Further, $d_p(\nu)$ denotes the number of cycles of ν with p elements ($d_1(\nu)$ is the number of fixed points of ν) and we have

$$\sum_{p \ge 1} d_p(\nu) = \#(\nu) = |\Pi(\nu)| = n - \|\nu\|.$$

Note that if $\pi \ge \Pi(\nu)$, then ν stabilizes the blocks of π, that is, $\nu(B) = B$ for all $B \in \pi$.

Finally, $\Pi(\boldsymbol{\sigma}, \boldsymbol{\tau})$ denotes the partition induced by the transitivity classes of the group generated by $\{\sigma_c, \tau_c \mid c \in \{1, \dots, D\}\}$, that is, $\Pi(\boldsymbol{\sigma}, \boldsymbol{\tau}) = \bigvee_{c=1}^{D} (\Pi(\sigma_c) \vee \Pi(\tau_c))$ and $|\Pi(\boldsymbol{\sigma}, \boldsymbol{\tau})|$ is its number of blocks. Note that all the permutations in $\boldsymbol{\sigma}, \boldsymbol{\tau}$ stabilize the blocks of some partition π if and only if $\Pi(\boldsymbol{\sigma}, \boldsymbol{\tau}) \le \pi$.

2.2. *Trace invariants*

We are interested in the invariants that can be built starting from an $N^D \times N^D$ matrix A. We define the *trace invariant* associated to $\boldsymbol{\sigma} \in \boldsymbol{S}_n$ as

$$\mathrm{Tr}_{\boldsymbol{\sigma}}(A) = \sum_{\text{all indices}} \Big(\prod_{s=1}^{n} A_{b_s^1 \dots b_s^D, a_s^1 \dots a_s^D} \Big) \prod_{c \in \{1, \dots, D\}} \Big(\prod_{s=1}^{n} \delta_{a_s^c, b_{\sigma_c(s)}^c} \Big).$$

These quantities are obviously invariant under conjugation by $U(N)^{\otimes D}$, that is, under $A \mapsto UAU^\star$ with $U = U^{(1)} \otimes \dots \otimes U^{(D)}$, $U^{(c)} \in U(N)$. For example:

- For $D = 1$ any $\mathrm{Tr}_\sigma(A)$ is a product of traces of powers of A, and the powers are the lengths of the cycles of σ:

$$\mathrm{Tr}_\sigma(A) = \prod_{p \ge 1} [\mathrm{Tr}(A^p)]^{d_p(\sigma)}. \tag{2.2}$$

For $n = 5$, $D = 1$ and $\sigma = (123)(45)$ we get $\mathrm{Tr}_\sigma(A) = \mathrm{Tr}(A^2)\,\mathrm{Tr}(A^3)$.

- If all the σ's are equal, $\sigma_c = \sigma$, then $\mathrm{Tr}_{\boldsymbol{\sigma}}(A)$ is again a product of traces of powers of A, but this time the traces are over indices of size N^D. Taking as before $\sigma = (123)(45)$ ($n = 5$ and D arbitrary) we get $\mathrm{Tr}_{(\sigma,\dots,\sigma)}(A) = \mathrm{Tr}(A^2)\,\mathrm{Tr}(A^3)$.
- We finish with an example with different σ's. For $n = 2$, $D = 2$, $\sigma_1 = (12)$, and $\sigma_2 = (1)(2)$,
$$\mathrm{Tr}_{(\sigma_1,\sigma_2)}(A) = \mathrm{Tr}_1[\mathrm{Tr}_2(A)\,\mathrm{Tr}_2(A)],$$
where Tr_c denotes the partial trace on the index of color c.

2.3. *Weingarten calculus*

Weingarten calculus [43] allows one to integrate any polynomial function on the unitary group. There exists a function $W^{(N)} : S_n \to \mathbb{R}$ such that, denoting by dU the Haar measure on $U(N)$, we have (see [13])

$$\int dU\, U_{i_1a_1}\cdots U_{i_na_n}\overline{U_{j_1b_1}\cdots U_{j_nb_n}} = \sum_{\sigma,\tau\in S_n}\Big(\prod_{s=1}^{n}\delta_{i_s,j_{\sigma(s)}}\Big)\Big(\prod_{s=1}^{n}\delta_{a_s,b_{\tau(s)}}\Big)W^{(N)}(\sigma\tau^{-1}). \tag{2.3}$$

The function $W^{(N)}$ is uniquely defined if and only if $n \le N$, and it follows from obvious commutativity relations that $W^{(N)}(\sigma\tau^{-1})$ depends only on the conjugacy class of $\sigma\tau^{-1}$, that is, $W^{(N)}$ is a central function on the symmetric group S_n. The functions $W^{(N)}$ are called *Weingarten functions*.

The $1/N$ expansion of Weingarten functions. We start with a theorem that characterizes and defines the Weingarten functions. Multiplying (2.3) by $\prod_{s=1}^{n}\delta_{a_s,b_s}$ and summing the repeated indices, we get

Theorem 2.1 (Collins–Śniady [15]). *The Weingarten function $\nu \mapsto W^{(N)}(\nu)$ and the function $\nu \mapsto \Phi(\nu) = N^{\#(\nu)} = N^n\delta_{\mathrm{id},\nu} + N^n\sum_{\rho\neq\mathrm{id}} N^{-\|\rho\|}\delta_{\rho;\nu}$ are pseudo-inverses under convolution. In particular, for $N \ge n$ one has*

$$\sum_{\nu\in S_n} W^{(N)}(\nu)\Phi(\nu^{-1}\sigma) = \delta_{\sigma,\mathrm{id}}.$$

This theorem can be used to compute the Weingarten functions:

$$W^{(N)}(\nu) = N^{-n}\sum_{k\ge 0}\sum_{\substack{\rho_1,\dots,\rho_k\in S_n^*\\ \rho_1\cdots\rho_k=\nu}}(-1)^k N^{-\sum_{i=1}^{k}\|\rho_i\|}, \tag{2.4}$$

with the convention that empty products are 1 and empty sums are 0. For $k = 0$, ν is an empty product in S_n, hence $\nu = \mathrm{id}$, and the empty sum $\sum_{i=1}^{k}\|\rho_i\|$ is zero. This expansion is convergent for $N \ge n$.

The coefficient of N^{-n-l} in the $1/N$ expansion of $W^{(N)}(\nu)$ is identified as in [13]:

$$W^{(N)}(\nu) = N^{-n}\sum_{l\ge 0}(-1)^l p(\nu;l)N^{-l}, \quad (-1)^l p(\nu;l) = \sum_{k\ge 0}(-1)^k m(\nu;l,k),$$

where

$$m(\nu; l, k) = \mathrm{Card}\left\{ (\rho_i)_{1\le i\le k} \;\middle|\; \rho_i \in S_n^* \text{ with } \rho_1 \cdots \rho_k = \nu \text{ and } \sum_{i=1}^{k} \|\rho_i\| = l \right\}, \quad (2.5)$$

and for $l = 0$ or 1, we have respectively $m(\nu; 0, k) = \delta_{\nu,\mathrm{id}}\delta_{k,0}$ and $m(\nu, 1, k) = \delta_{k,1}\delta_{\|\nu\|,1}$, while for $k = 0$ or 1, we have respectively $m(\nu; l, 0) = \delta_{\nu,\mathrm{id}}\delta_{l,0}$ and $m(\nu, l, 1) = \delta_{l,\|\nu\|}(1 - \delta_{\nu,\mathrm{id}})$. We conclude that $p(\nu; 0) = \delta_{\nu,\mathrm{id}}$ and $p(\nu, 1) = \delta_{\|\nu\|,1}$.

This expression for the coefficient of N^{-n-l} as an alternating sum does not render explicit its sign. Another expression [14] (see also [37]) solves this issue.

Definition 2.2. Let (pq) be the elementary transposition of p and q (that is, we use cycle notation, but we omit the cycles with one element). An ordered l-tuple of transpositions $\mu_1 = (p_1 q_1), \ldots, \mu_l = (p_l q_l)$ is said to have *weakly monotone maxima* if $p_k < q_k$ for each $k \in \{1, \ldots, l\}$ and $q_k \le q_{k+1}$ for each $k \in \{1, \ldots, l-1\}$.

Theorem 2.3 ([14, 37]). *Denote by $P(\nu; l)$ the set of solutions of $\nu = \mu_1 \ldots \mu_l$ with $\mu_1, \ldots, \mu_l$ elementary transpositions with weakly monotone maxima. Then $p(\nu; l) = |P(\nu; l)|$.*

Asymptotics of Weingarten functions. Classical theorems in combinatorics allow one to obtain the asymptotics of Weingarten functions [13, Theorem 2.15 (ii)].

Corollary 2.4. *For $\nu \in S_n$, we have the asymptotic expansion*

$$W^{(N)}(\nu) = N^{-n-\|\nu\|}\mathsf{M}(\nu)(1 + O(N^{-2})), \quad (2.6)$$

where $\mathsf{M}(\nu)$ is Biane–Speicher's Möbius function on the lattice of non-crossing partitions [39, Lecture 10]*, which is a central function which can be written in terms of the Catalan numbers:*

$$\mathsf{M}(\nu) = \prod_{p\ge 1}\left[\frac{(-1)^{p-1}}{p}\binom{2p-2}{p-1}\right]^{d_p(\nu)}. \quad (2.7)$$

2.4. Moments of the tensor HCIZ integral

The moments of the tensor HCIZ integral (1.1) can be written in terms of Weingarten functions.

Proposition 2.5. *The moments of the tensor HCIZ integral* (1.1) *are*

$$\begin{aligned}\mathbb{E}_U\left([\mathrm{Tr}(AUBU^*)]^n\right) &= \int [dU]\,[\mathrm{Tr}(AUBU^*)]^n \\ &= \sum_{\boldsymbol{\sigma},\boldsymbol{\tau}\in \boldsymbol{S}_n} \mathrm{Tr}_{\boldsymbol{\sigma}}(A)\,\mathrm{Tr}_{\boldsymbol{\tau}^{-1}}(B)\prod_{c=1}^{D} W^{(N)}(\sigma_c\tau_c^{-1}), \end{aligned} \quad (2.8)$$

where $\boldsymbol{\tau}^{-1} = (\tau_1^{-1}, \ldots, \tau_D^{-1})$ and $W^{(N)}(\cdot)$ are the Weingarten functions.

Proof. The proof is straightforward. Starting from

$$[\mathrm{Tr}(AUBU^*)]^n = \sum_{\text{all indices}} \prod_{s=1}^{n} A_{j_s^1 \dots j_s^D, i_s^1 \dots i_s^D} \Big(\prod_{c=1}^{D} U^{(c)}_{i_s^c a_s^c} \Big) B_{a_s^1 \dots a_s^D, b_s^1 \dots b_s^D} \Big(\prod_{c=1}^{D} \bar{U}^{(c)}_{j_s^c b_s^c} \Big),$$

and using D times the Weingarten formula (2.3), the expectation amounts to

$$\sum_{\boldsymbol{\sigma},\boldsymbol{\tau} \in \boldsymbol{S}_n} \sum_{\text{all indices}} \Big(\prod_{s=1}^{n} A_{j_s^1 \dots j_s^D, i_s^1 \dots i_s^D} B_{a_s^1 \dots a_s^D, b_s^1 \dots b_s^D} \Big) \Big(\prod_{c=1}^{D} \prod_{s=1}^{n} \delta_{i_s^c, j^c_{\sigma_c(s)}} \delta_{a_s^c, b^c_{\tau_c(s)}} \Big) \cdot W^{(N)}(\sigma_c \tau_c^{-1}),$$

where we recognize the definition of the trace invariants (Sec. 2.2). Observe that the first and the second indices of the U's in (2.3) play slightly different roles, leading to the fact that the permutations for the invariant of B are inverted. ∎

From Corollary 2.4, we obtain the asymptotic expression of the moments:

$$\mathbb{E}\big([\mathrm{Tr}(AUBU^*)]^n\big) = \sum_{\boldsymbol{\sigma},\boldsymbol{\tau} \in \boldsymbol{S}_n} \mathrm{Tr}_{\boldsymbol{\sigma}}(A)\, \mathrm{Tr}_{\boldsymbol{\tau}^{-1}}(B) \prod_{c=1}^{D} \mathsf{M}(\sigma_c \tau_c^{-1}) N^{-n-\|\sigma_c \tau_c^{-1}\|}(1 + O(N^{-2})). \tag{2.9}$$

2.5. Constellations

We now review some results on the number of constellations. Constellations are central to the combinatorial interpretation of our main results.

Definition and graphical representation. Intuitively, a combinatorial map (fatgraph, or ribbon graph in the physics literature) is a graph embedded in a closed surface[2] in which each edge is subdivided into two *half*-edges. The map is bipartite if its vertices have one of two flavors (say 1 and 2) and every edge connects two vertices of different flavors.

Formally a bipartite combinatorial map, or a 2-*constellation*, is an ordered pair of permutations, $\hat{\rho} = (\rho_1, \rho_2)$, $\rho_1, \rho_2 \in S_n$. It is represented canonically as an embedded graph as follows:

- We let the *flavor* be $i = 1, 2$. To each cycle of ρ_i we draw a vertex embedded in the plane (a disc). To each $s \in \{1, \dots, n\}$ we attach a *half-edge*, i.e. an outgoing segment to one of the vertices, labeled $\langle s \rangle_i$. Every s belongs to a cycle of ρ_i and we draw the half-edges $\langle s \rangle_i$, $\langle \rho_i(s) \rangle_i$, $\langle \rho_i(\rho_i(s)) \rangle_i$ etc. ordered cyclically *counterclockwise* around the vertex corresponding to this cycle.

[2]More precisely, the graph is drawn on the surface without edge-crossing and so that the complement of the graph in the surface is homeomorphic to a collection of discs. The graph is then considered up to orientation preserving homeomorphism of the surface whose restriction to the embedded graph is an isomorphism.

- For every $s \in \{1, \ldots, n\}$ we join the two half-edges $\langle s \rangle_1$ and $\langle s \rangle_2$ into an edge labeled s.

The permutations ρ_1 and ρ_2 encode the "successor" half-edge: $\langle \rho_i(s) \rangle_i$ is the first half-edge encountered after $\langle s \rangle_i$ when turning counterclockwise around the vertex of flavor i to which $\langle s \rangle_i$ belongs. The permutation $\rho_1\rho_2$ maps the edge s onto the edge $\rho_1(\rho_2(s))$ obtained by first stepping from s to $\rho_2(s)$, the successor of s on the vertex of flavor 2 to which s is hooked, and then stepping from $\rho_2(s)$ to $\rho_1(\rho_2(s))$, the successor of $\rho_2(s)$ on the vertex of flavor 1 to which $\rho_2(s)$ is hooked. The cycles of $\rho_1\rho_2$ are the *faces* of the map.

This can be generalized to $k \geq 2$ flavors. A labeled k-*constellation* is an ordered k-uple of permutations, $\hat{\rho} = (\rho_1, \ldots, \rho_k)$, $\rho_1, \ldots, \rho_k \in S_n$. The construction below is exemplified in Fig. 1, which we will be using extensively:

- To each permutation ρ_i, $i \in \{1, \ldots, k\}$ we associate a set of embedded *vertices of flavor* i corresponding to its cycles. In Fig. 1 the flavored vertices are represented as blue (for flavor 1), red (for flavor 2) and yellow (for flavor 3). As $\rho_1 = (12)(354)$ there are two blue vertices in the figure, one bi-valent corresponding to the cycle (12) and one tri-valent corresponding to the cycle (354).

  The vertices of flavor i have a total of n outgoing, cyclically ordered *counterclockwise*, half-edges $\langle s \rangle_i$ for $s \in \{1, \ldots, n\}$. For instance, in Fig. 1, the three half-edges incident to the tri-valent blue vertex are labeled $\langle 3 \rangle_1$, $\langle 5 \rangle_1$ and $\langle 4 \rangle_1$.
- To each $s \in \{1, \ldots, n\}$ we associate an embedded *white vertex* $[s]$.[3] We connect the half-edges $\langle s \rangle_i$ for all the flavors i to the vertex $[s]$ via edges such that the flavors are

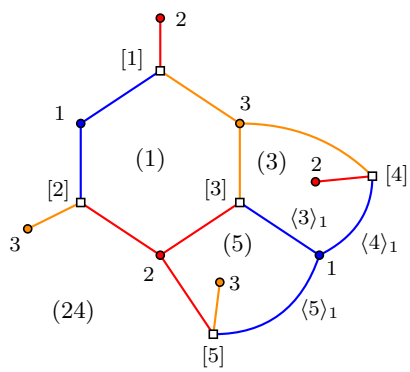

Fig. 1. A constellation with $k = 3$ flavors and $n = 5$ with $\rho_1 = (12)(354)$, $\rho_2 = (1)(253)(4)$ and $\rho_3 = (134)(2)(5)$. Its faces are the cycles of $\rho_1\rho_2\rho_3 = (1)(24)(3)(5)$, corresponding to three hexagons and one dodecagon. The labels in the figure are respectively the flavors of the vertices (no parenthesis), the labels of some of the half-edges (angle brackets), the labels of the white vertices (square brackets), and the cycles of $\rho_1\rho_2\rho_3$ corresponding to the faces (parentheses).

[3]The labeled white vertices generalize the labeled edges of the bipartite maps and can be viewed as hyper-edges.

encountered in the order $1, \dots, k$ when turning around the vertex $[s]$ *clockwise*; see Fig. 1. We label the edges by their end vertices as $([s], i)$ or $(i, [s])$.

- The *faces* of the constellation are the disjoint cycles of the product $\rho_1 \cdots \rho_k$. A cycle of length p corresponds to a face with $2pk$ corners (bounded by $2pk$ edges) that are alternately white vertices $[s]$ and flavored vertices corresponding to ρ_i. The flavored vertices are encountered cyclically in the order $k, k-1, \dots, 2, 1$ when going around the face while keeping the boundary edges to the left.

  The point is that the composition of the permutations ρ_i encodes the walk around the faces of the constellation. To see this, let us consider the face to the right of the edge $([4], 3)$ in Fig. 1. Its perimeter consists of the edges

$$([4], 3), (3, [1]), ([1], 2), (2, [1]), ([1], 1), (1, [2]),$$
$$([2], 3), (3, [2]), ([2], 2), (2, [5]), ([5], 1), (1, [4]).$$

  The first two edges $([4], 3), (3, [1])$ encode the fact that one passes from the white vertex $[4]$ to the white vertex $[1]$ by passing through a vertex of flavor 3. This translates into the fact that $\rho_3(4) = 1$. The next couple of edges $([1], 2), (2, [1])$ leads to the fact that $\rho_2(1) = 1$, hence the first four edges together read, in terms of permutations, $\rho_2\rho_3(4) = 1$. Finally, $([1], 1), (1, [2])$ signifies that $\rho_1(1) = 2$, which combined with the previous edges reads $\rho_1\rho_2\rho_3(4) = 2$. Continuing the walk along the face encodes the action with ρ_3, ρ_2 and ρ_1 again. The face closes when we come back to the vertex $[4]$, at which point the tour around the face reads $\rho_1\rho_2\rho_3\rho_1\rho_2\rho_3(4) = 4$.

- The connected components of the resulting graph correspond to the transitivity classes of the group generated by $\{\rho_1, \dots, \rho_k\}$. Indeed, for $s, s' \in \{1, \dots, n\}$ such that $\rho_{i_1} \cdots \rho_{i_l}(s) = s'$, the vertices $[s]$ and $[s']$ are connected in the constellation by the path

$$([s], i_l)(i_l, [\rho_{i_l}(s)]) \dots ([\rho_{i_2} \cdots \rho_{i_l}(s)] i_1)(i_1 [s']).$$

  The partition $\Pi(\rho_1, \dots, \rho_k) = \bigvee_{i=1}^k \Pi(\rho_i) \equiv \Pi(\hat{\rho})$ is reconstructed by collecting all the white vertices $[s]$ belonging to the same connected component of the constellation into a block. The number of connected components of a constellation is $|\Pi(\hat{\rho})|$. The constellation is said to be *connected* if $|\Pi(\hat{\rho})| = 1$, that is, the group generated by $\{\rho_1, \dots, \rho_k\}$ acts transitively on $\{1, \dots, n\}$.

For $k = 2$ the white vertices $[s]$ have valency 2 and therefore can be viewed as decorations (labels) on edges, and we recover the bipartite maps described at the beginning of the section.

Euler characteristic. A k-constellation $\hat{\rho}$ has $\sum_i \#(\rho_i) + n$ vertices, kn edges, $\#(\rho_1 \cdots \rho_k)$ faces and $|\Pi(\hat{\rho})|$ connected components. Being a combinatorial map, it has a non-negative genus, denoted by $g(\hat{\rho}) \geq 0$, and Euler characteristic

$$\sum_{i=1}^k \#(\rho_i) - n(k-1) + \#(\rho_1 \cdots \rho_k) = 2|\Pi(\hat{\rho})| - 2g(\hat{\rho}). \tag{2.10}$$

A constellation (seen as a combinatorial map) is *planar*, $g(\hat{\rho}) = 0$, if and only if it can be drawn on the 2-sphere without edge-crossings so that each region of the complement of the graph on the sphere is homeomorphic to a disc.

In terms of the length of permutations, (2.10) becomes

$$\sum_{i=1}^{k} \|\rho_i\| = 2n - \|\rho_1 \cdots \rho_k\| + 2g(\hat{\rho}) - 2|\Pi(\hat{\rho})|. \tag{2.11}$$

Among the connected constellations, the planar ones are such that $\sum_{i=1}^{k} \|\rho_i\|$ is minimal at fixed $\nu = \rho_1 \cdots \rho_k$.

Counting planar constellations. The main result we will need is due to [5] and concerns the number of planar constellations. We fix $\nu \in S_n$. For $k \geq 2$, the number of *connected planar* k-constellations $(\rho_1, \ldots, \rho_k)$ in S_n with faces corresponding to disjoint cycles of $\rho_1 \cdots \rho_k = \nu$ is

$$\tilde{\gamma}(\nu; k) = k \frac{[(k-1)n-1]!}{[(k-1)n - \#(\nu) + 2]!} \prod_{p \geq 1} \left[p \binom{kp-1}{p} \right]^{d_p(\nu)} = \sum_{\substack{\rho_1, \ldots, \rho_k \in S_n,\, \rho_1 \cdots \rho_k = \nu \\ \Pi(\hat{\rho}) = 1_n,\, \sum_i \|\rho_i\| = 2n-2-\|\nu\|}} 1. \tag{2.12}$$

For the boundary values, we get

$$\tilde{\gamma}(\nu; 0) = \delta_{\nu, \mathrm{id}}, \quad \tilde{\gamma}(\nu; 1) = \delta_{\#(\nu), 1}. \tag{2.13}$$

This can be adapted to constellations $(\rho_1, \ldots, \rho_k)$ satisfying the same assumptions, but for which none of the permutations ρ_i involved is the identity [5] (such constellations are said to be *proper*), whose number is given by

$$\gamma(\nu; k) = \sum_{j=0}^{k} \binom{k}{j} \tilde{\gamma}(\nu; j)(-1)^{k+j}, \tag{2.14}$$

and $\gamma(\nu; 0) = \delta_{\nu, \mathrm{id}}$. One can furthermore compute [13] the following alternating sum:

$$\begin{aligned} \gamma(\nu) &= \sum_{k \geq 0} (-1)^k \gamma(\nu, k) = (-1)^{\|\nu\|} \frac{(3n - \|\nu\| - 3)!}{(2n)!} \prod_{p \geq 1} \left[\frac{(2p)!}{p!(p-1)!} \right]^{d_p(\nu)} \\ &= \mathsf{M}(\nu) \frac{(3n - \|\nu\| - 3)!}{(2n)!} \prod_{p \geq 1} [2p(2p-1)]^{d_p(\nu)}, \end{aligned} \tag{2.15}$$

with $\mathsf{M}(\nu)$ the Möbius function (2.7) on non-crossing partitions. Note that this is a class function.

More generally, we denote by $\tilde{\gamma}_l(\nu; k)$ and $\gamma_l(\nu; k)$ the numbers of generic (i.e. not necessarily proper) and respectively proper connected k-constellations with faces corresponding to the disjoint cycles of $\rho_1 \cdots \rho_k = \nu$ with $\sum_{i=1}^{k} \|\rho_i\| = l$ (hence with genus

$2g(\hat{\rho}) = l + 2 - 2n + \|\nu\|$):

$$\tilde{\gamma}_l(\nu;k) = \sum_{\substack{\rho_1,\dots,\rho_k \in S_n,\, \rho_1\cdots\rho_k=\nu \\ \sum_i \|\rho_i\|=l,\, \Pi(\hat{\rho})=1_n}} 1,$$

$$\gamma_l(\nu;k) = \sum_{\substack{\rho_1,\dots,\rho_k \in S_n^*,\, \rho_1\cdots\rho_k=\nu \\ \sum_i \|\rho_i\|=l,\, \Pi(\hat{\rho})=1_n}} 1 = \sum_{j=0}^{k} \binom{k}{j} \tilde{\gamma}_l(\nu;j)(-1)^{k+j},$$

where the last equality follows by inverting the relation $\tilde{\gamma}_l(\nu;k) = \sum_{j=0}^{k} \binom{k}{j} \gamma_l(\nu;j)$.

Finally, we denote by $\gamma_l(\nu)$ the alternating sum of the numbers of connected proper constellations with faces corresponding to the disjoint cycles of $\rho_1 \cdots \rho_k = \nu$ with $\sum_{i=1}^{k} \|\rho_i\| = l$:

$$\gamma_l(\nu) = \sum_{k\geq 0} (-1)^k \gamma_l(\nu;k). \tag{2.16}$$

Equations (2.12), (2.14), and (2.15) correspond to the minimal possible value $l = 2n - 2 - \|\nu\|$.

2.6. *Cumulants*

For X some random variable, the *cumulant* $C(X^n)$, also sometimes called connected correlation, is defined by

$$C(t) = \log \mathbb{E}(\exp tX) = \sum_{p\geq 1} t^n \frac{C(X^n)}{n!}.$$

For instance, the second cumulant $C(X^2) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$ is the variance of the probability distribution of X. The cumulants can be written in terms of the moments of the distribution and vice versa. In order to write down the relation between the two in a convenient form, we introduce some notation.

It is convenient to distinguish between the different factors X in the monomial X^n. We do this by introducing a fictitious label $i = 1, \dots, n$ and writing $X^n = X_1 \cdots X_n$ where $X_i = X$ for all i. Then the nth cumulant can be written as $C(X^n) \equiv C(\prod_{i=1}^{n} X_i)$.

For any partition $\pi \in \mathcal{P}(n)$ with blocks $B \in \pi$, we define $C_\pi = \prod_{B\in\pi} C(\prod_{i\in B} X_i)$. We are now in a position to write expectations in terms of cumulants:

$$\mathbb{E}\Big(\prod_{i=1}^{n} X_i\Big) = \sum_{\pi\in\mathcal{P}(n)} C_\pi. \tag{2.17}$$

Equation (2.17) can be inverted through the Möbius inversion formula to express cumulants in terms of expectations. Defining $\mathbb{E}_\pi = \prod_{B\in\pi} \mathbb{E}(\prod_{i\in B} X_i)$, we have

$$\mathbb{E}_\pi = \sum_{\pi'\leq\pi} C_{\pi'}, \quad C_\pi = \sum_{\pi'\leq\pi} \lambda_{\pi',\pi} \mathbb{E}_{\pi'}, \tag{2.18}$$

where $\lambda_{\pi',\pi}$ is the Möbius function (2.7) for the lattice of partitions [41]:[4]

$$\lambda_{\pi',\pi} = \prod_{B\in\pi} (-1)^{|\pi'_{|B}|-1}(|\pi'_{|B}|-1)!,$$

where $\pi'_{|B}$ is the restriction of the partition π' to the block $B \in \pi$. This restriction is well-defined because $\pi' \leq \pi$. In particular, recalling that 1_n denotes the one-block partition, we have

$$C(X^n) = C_{1_n} = \sum_{\pi} \lambda_\pi \mathbb{E}_\pi, \quad \lambda_\pi \equiv \lambda_{\pi,1_n} = (-1)^{|\pi|-1}(|\pi|-1)!. \tag{2.19}$$

3. Cumulants of the tensor HCIZ integral

The study of cumulants of the tensor HCIZ integral is the core of this paper. They expand in terms of trace invariants of A and B times *cumulant Weingarten functions* defined in Sec. 3.1. An expression of the latter as a series in $1/N$ is derived in Sec. 3.2. The coefficients of this expansion are shown to count certain transitive factorizations of D-uplets of permutations.

3.1. Cumulant Weingarten functions

We denote by $\chi(\cdot)$ the indicator function which is 1 if the condition $\cdot$ is true and 0 otherwise.

Definition 3.1 (Cumulant Weingarten functions). For any partition π, let $W_\pi^{(N)}[\boldsymbol{\sigma},\boldsymbol{\tau}]$ be

- zero if at least one of the permutations involved in $\boldsymbol{\sigma}$ or $\boldsymbol{\tau}$ does not stabilize the blocks of π, that is, $W_\pi^{(N)}[\boldsymbol{\sigma},\boldsymbol{\tau}]$ is zero unless $\Pi(\boldsymbol{\sigma},\boldsymbol{\tau}) \leq \pi$;
- the product over the blocks of π of Weingarten functions involving permutations restricted to these blocks if all the permutations in $\boldsymbol{\sigma},\boldsymbol{\tau}$ stabilize the blocks of π.

Denoting by $\sigma_{c|B}$ the restriction of σ_c to the block $B \in \pi$ (which is well-defined whenever σ_c stabilizes the blocks of π), we have

$$W_\pi^{(N)}[\boldsymbol{\sigma},\boldsymbol{\tau}] = \chi(\Pi(\boldsymbol{\sigma},\boldsymbol{\tau}) \leq \pi) \prod_{B\in\pi} \prod_{c=1}^{D} W^{(N)}(\sigma_{c|B}\tau_{c|B}^{-1}).$$

The *cumulant Weingarten function* $W_C^{(N)}[\boldsymbol{\sigma},\boldsymbol{\tau}]$ is

$$W_C^{(N)}[\boldsymbol{\sigma},\boldsymbol{\tau}] = \sum_{\pi} \lambda_\pi W_\pi^{(N)}[\boldsymbol{\sigma},\boldsymbol{\tau}], \tag{3.1}$$

where $\lambda_\pi = \lambda_{\pi,1_n}$ is the Möbius function with the second argument set to the one-block partition.

[4]This should not be confused with the Möbius function for the lattice of *non-crossing* partitions.

Observe that, due to the indicator function, both $W_\pi^{(N)}[\boldsymbol{\sigma},\boldsymbol{\tau}]$ and $W_C^{(N)}[\boldsymbol{\sigma},\boldsymbol{\tau}]$ depend on $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$ and not only on the product $\boldsymbol{\sigma}\boldsymbol{\tau}^{-1}$. The cumulant Weingarten functions arise naturally in the expansion of the cumulants of the tensor HCIZ integral in terms of trace invariants.

Proposition 3.2. *The cumulants of the tensor HCIZ integral* (1.1) *are*

$$C\left([\mathrm{Tr}(AUBU^*)]^n\right) = \sum_{\boldsymbol{\sigma},\boldsymbol{\tau}\in \boldsymbol{S}_n} \mathrm{Tr}_{\boldsymbol{\sigma}}(A)\,\mathrm{Tr}_{\boldsymbol{\tau}^{-1}}(B) W_C^{(N)}[\boldsymbol{\sigma},\boldsymbol{\tau}], \tag{3.2}$$

where $W_C^{(N)}[\boldsymbol{\sigma},\boldsymbol{\tau}]$ is the cumulant Weingarten function, uniquely defined for $N \geq n$.

Proof. To any partition π, we associate the expectation

$$\mathbb{E}_\pi = \prod_{B\in\pi} \mathbb{E}\Big(\prod_{i\in B} \mathrm{Tr}(AU_i BU_i^*)\Big) = \sum_{\boldsymbol{\sigma},\boldsymbol{\tau}\in \boldsymbol{S}_n} \mathrm{Tr}_{\boldsymbol{\sigma}}(A)\,\mathrm{Tr}_{\boldsymbol{\tau}^{-1}}(B) W_\pi^{(N)}[\boldsymbol{\sigma},\boldsymbol{\tau}],$$

where the second equality follows from Prop. 2.5. It then follows from (2.19) that

$$C\left([\mathrm{Tr}(AUBU^*)]^n\right) = \sum_{\boldsymbol{\sigma},\boldsymbol{\tau}\in \boldsymbol{S}_n} \mathrm{Tr}_{\boldsymbol{\sigma}}(A)\,\mathrm{Tr}_{\boldsymbol{\tau}^{-1}}(B) \sum_\pi \lambda_\pi W_\pi^{(N)}[\boldsymbol{\sigma},\boldsymbol{\tau}],$$

which leads to (3.2) using the definition (3.1). ∎

3.2. Exact expression of cumulant Weingarten functions

Theorem 3.3. *The cumulant Weingarten functions are*

$$W_C^{(N)}[\boldsymbol{\sigma},\boldsymbol{\tau}] = N^{-nD} \sum_{l\geq 0} (-1)^l p_C[\boldsymbol{\sigma},\boldsymbol{\tau};l] N^{-l},$$

where

$$(-1)^l p_C[\boldsymbol{\sigma},\boldsymbol{\tau};l] = \sum_{k\geq 0} (-1)^k m_C(\boldsymbol{\sigma},\boldsymbol{\tau};l,k),$$

and $m_C(\boldsymbol{\sigma},\boldsymbol{\tau};l,k)$ is the number of D-uplets of constellations $(\rho_{i_c}^c)_{1\leq i_c\leq k_c}$, $c \in \{1,\dots,D\}$, with the following properties:

- *all the permutations $\rho_{i_c}^c$ are different from the identity permutation,*
- *for all $c \in \{1,\dots,D\}$, $\sigma_c\tau_c^{-1} = \rho_1^c \cdots \rho_{k_c}^c$,*
- *$\sum_{c=1}^D k_c = k$ with $k_c \geq 0$, and $k_c = 0$ implies $\sigma_c = \tau_c$,*
- *$\sum_{c=1}^D \sum_{i_c=1}^{k_c} \|\rho_{i_c}^c\| = l$,*
- *the collection of all $\{\sigma_c, \tau_c, (\rho_{i_c}^c)_{1\leq i_c\leq k_c}\}_{1\leq c\leq D}$ acts transitively on $\{1,\dots,n\}$.*

This expansion is convergent for $N \geq n$.

The boundary values are $m_C(\boldsymbol{\sigma},\boldsymbol{\tau};0,k) = \delta_{|\Pi(\boldsymbol{\sigma},\boldsymbol{\tau})|,1}\delta_{\boldsymbol{\sigma},\boldsymbol{\tau}}\delta_{k,0}$ and $p_C[\boldsymbol{\sigma},\boldsymbol{\tau};0] = \delta_{|\Pi(\boldsymbol{\sigma},\boldsymbol{\tau})|,1}\delta_{\boldsymbol{\sigma},\boldsymbol{\tau}}$.

Proof. The functions $W_\pi^{(N)}[\boldsymbol{\sigma},\boldsymbol{\tau}]$ in Def. 3.1 are non-trivial only if $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$ stabilize the blocks of π. We let $\boldsymbol{\nu}=\boldsymbol{\sigma}\boldsymbol{\tau}^{-1}$ and denote by $\nu_{c|B}$ the restriction of ν_c to the block B. Then (2.4) leads to

$$\begin{aligned}
W_\pi^{(N)}[\boldsymbol{\sigma},\boldsymbol{\tau}] &= \chi(\Pi(\boldsymbol{\sigma},\boldsymbol{\tau})\le\pi)\\
&\quad\cdot\prod_{c=1}^{D}\prod_{B\in\pi}N^{-|B|}\Big(\sum_{k_c^B\ge 0}\sum_{\substack{\rho_1^{c,B},\dots,\rho_{k_c^B}^{c,B}\in S^*_{|B|}\\ \rho_1^{c,B}\cdots\rho_{k_c^B}^{c,B}=\nu_{c|B}}}(-1)^{k_c^B}N^{-\sum_{i_c^B=1}^{k_c^B}\|\rho_{i_c^B}^{c,B}\|}\Big)\\
&= \chi(\Pi(\boldsymbol{\sigma},\boldsymbol{\tau})\le\pi)\sum_{\{k_c^B\}_{c,B}\ge 0}\sum_{\substack{\{\rho_1^{c,B},\dots,\rho_{k_c^B}^{c,B}\in S^*_{|B|}\}_{c,B}\\ \{\rho_1^{c,B}\cdots\rho_{k_c^B}^{c,B}=\nu_{c|B}\}_{c,B}}}(-1)^{\sum_{c,B}k_c^B}\\
&\qquad\cdot N^{-nD-\sum_{c,B}\sum_{i_c^B=1}^{k_c^B}\|\rho_{i_c^B}^{c,B}\|}.
\end{aligned}\tag{3.3}$$

where we have exchanged the sums and the products. Note that if $k_c^B=0$ for some c and B, then there are no permutations $\rho_{i_c^B}^{c,B}$ and the rightmost sum becomes $\delta_{\nu_{c|B},\mathrm{id}}$.

The permutations $\rho_{i_c^B}^{c,B}$ can be trivially lifted to permutations on $\{1,\dots,n\}$ by supplementing them with the identity on the complement of B. We denote the set of all the (lifted) permutations ρ by

$$\hat{\boldsymbol{\rho}}=\big\{\rho_{i_c^B}^{c,B}\;\big|\;1\le i_c^B\le k_c^B,\ B\in\pi,\ c\in\{1,\dots,D\}\big\},$$

and by $\Pi(\hat{\boldsymbol{\rho}})$ the partition induced by the transitivity classes of the group generated by all the permutations in $\hat{\boldsymbol{\rho}}$. As $\rho_{i_c^B}^{c,B}$ acts non-trivially only on the block $B\in\pi$, it follows that all the permutations in $\hat{\boldsymbol{\rho}}$ stabilize the partition π, hence $\Pi(\hat{\boldsymbol{\rho}})\le\pi$.

Now comes a subtle point. We would like to rewrite $W_\pi^{(N)}[\boldsymbol{\sigma},\boldsymbol{\tau}]$ via a moment-cumulant formula such as (2.18), that is, as a sum of "cumulants" over $\pi'\le\pi$. The obvious idea to rearrange the sum according to the partition $\Pi(\hat{\boldsymbol{\rho}})$ of a summand which in turn sums all the $\hat{\boldsymbol{\rho}}$s with the same $\Pi(\hat{\boldsymbol{\rho}})$ fails due to the global factor $\chi(\Pi(\boldsymbol{\sigma},\boldsymbol{\tau})\le\pi)$. Another idea works: we rearrange the sum according to the partition $\Pi(\boldsymbol{\sigma},\boldsymbol{\tau})\vee\Pi(\hat{\boldsymbol{\rho}})\ge\Pi(\boldsymbol{\sigma},\boldsymbol{\tau})$, that is, we note that

$$W_\pi^{(N)}[\boldsymbol{\sigma},\boldsymbol{\tau}]=\sum_{\Pi(\boldsymbol{\sigma},\boldsymbol{\tau})\le\pi'\le\pi}W_{C,\pi'}^{(N)}[\boldsymbol{\sigma},\boldsymbol{\tau}],$$

with the cumulant

$$W_{C,\pi'}^{(N)}[\boldsymbol{\sigma},\boldsymbol{\tau}]=\sum_{\{k_c^B\}_{c,B}\ge 0}\sum_{\substack{\{\rho_1^{c,B},\dots,\rho_{k_c^B}^{c,B}\in S^*_{|B|}\}_{c,B}\\ \{\rho_1^{c,B}\cdots\rho_{k_c^B}^{c,B}=\nu_{c|B}\}_{c,B}\\ \Pi(\boldsymbol{\sigma},\boldsymbol{\tau})\vee\Pi(\hat{\boldsymbol{\rho}})=\pi'}}(-1)^{\sum_{c,B}k_c^B}N^{-nD-\sum_{c,B}\sum_{i_c^B=1}^{k_c^B}\|\rho_{i_c^B}^{c,B}\|}.\tag{3.4}$$

This expression is inverted using (2.18) to yield

$$W^{(N)}_{C,\pi}[\boldsymbol{\sigma},\boldsymbol{\tau}] = \sum_{\Pi(\boldsymbol{\sigma},\boldsymbol{\tau})\le\pi'\le\pi} \lambda_{\pi',\pi} W^{(N)}_{\pi'}[\boldsymbol{\sigma},\boldsymbol{\tau}].$$

Choosing $\pi = 1_n$, we recover the right hand side of (3.1), thus $W^{(N)}_C[\boldsymbol{\sigma},\boldsymbol{\tau}] = W^{(N)}_{C,1_n}[\boldsymbol{\sigma},\boldsymbol{\tau}]$, i.e.

$$\begin{aligned} W^{(N)}_C[\boldsymbol{\sigma},\boldsymbol{\tau}] &= \sum_{\{k_c\}_c\ge 0} \sum_{\substack{\{\rho^c_1,\dots,\rho^c_{k_c}\in S^*_n,\, \rho^c_1\cdots\rho^c_{k_c}=\nu_c\}_c \\ \Pi(\boldsymbol{\sigma},\boldsymbol{\tau})\vee\Pi(\hat{\boldsymbol{\rho}})=1_n}} (-1)^{\sum_c k_c} N^{-nD-\sum_c\sum_{i_c=1}^{k_c}\|\rho^c_{i_c}\|}, \\ &= \sum_{k,l\ge 0} (-1)^k N^{-nD-l} \sum_{\substack{\{k_c\}_c\ge 0 \\ \sum_c k_c = k}} \sum_{\substack{\{\rho^c_1,\dots,\rho^c_{k_c}\in S^*_n,\, \rho^c_1\cdots\rho^c_{k_c}=\nu_c\}_c \\ \Pi(\boldsymbol{\sigma},\boldsymbol{\tau})\vee\Pi(\hat{\boldsymbol{\rho}})=1_n,\, \sum_{c=1}^D\sum_{i_c=1}^{k_c}\|\rho^c_{i_c}\|=l}} 1, \end{aligned} \tag{3.5}$$

and we recognize the coefficient of N^{-nD-l} in this expansion to be the alternating sum defining $m_C(\boldsymbol{\sigma},\boldsymbol{\tau};l,k)$. ∎

One drawback of (3.5) is that analytic bounds are difficult to obtain because the sum is signed. On the other hand, it renders obvious the invariance by relabeling of $\{1,\dots,n\}$.

Corollary 3.4. *If $\{\boldsymbol{\sigma},\boldsymbol{\tau}\}$ acts transitively on $\{1,\dots,n\}$, that is, $|\Pi(\boldsymbol{\sigma},\boldsymbol{\tau})| = 1$, then*

$$W^{(N)}_C[\boldsymbol{\sigma},\boldsymbol{\tau}]|_{|\Pi(\boldsymbol{\sigma},\boldsymbol{\tau})|=1} = \prod_{c=1}^D W^{(N)}[\sigma_c\tau_c^{-1}] = \prod_{c=1}^D N^{-n-\|\sigma_c\tau_c^{-1}\|}\mathsf{M}(\sigma_c\tau_c^{-1})(1+O(N^{-2})).$$

If moreover $\boldsymbol{\sigma} = \boldsymbol{\tau}$, then

$$W^{(N)}_C[\boldsymbol{\sigma},\boldsymbol{\sigma}]|_{|\Pi(\boldsymbol{\sigma})|=1} = N^{-nD}(1+O(N^{-2})).$$

Proof. If $\Pi(\boldsymbol{\sigma},\boldsymbol{\tau}) = 1_n$, the condition $\Pi(\boldsymbol{\sigma},\boldsymbol{\tau})\vee\Pi(\hat{\boldsymbol{\rho}}) = 1_n$ in (3.5) is lifted, and we recover the product of (2.4) for each color. ∎

Remark 3.5. Let $D = 1$ and fix $\sigma,\tau\in S_n$. We observe that $\Pi(\sigma,\tau)\ge\Pi(\sigma\tau^{-1})$ and $\Pi(\hat\rho)\ge\Pi(\sigma\tau^{-1})$ for any $\hat\rho$ such that $\sigma\tau^{-1} = \rho_1\cdots\rho_k$. This is because the group generated by $\sigma\tau^{-1}$ is a subgroup of both the group generated by (σ,τ) and the one generated by $(\rho_1,\dots,\rho_k)$.

We have $\Pi(\sigma,\tau)\vee\Pi(\hat\rho) = \Pi(\hat\rho)$ for any $\hat\rho$ such that $\sigma\tau^{-1} = \rho_1\cdots\rho_k$ if and only if $\Pi(\sigma,\tau) = \Pi(\sigma\tau^{-1})$.[5] This comes about as follows:

- If $\Pi(\sigma,\tau) = \Pi(\sigma\tau^{-1})$ then taking into account that $\Pi(\hat\rho)\ge\Pi(\sigma\tau^{-1})$, we conclude that $\Pi(\hat\rho)\ge\Pi(\sigma,\tau)$; therefore $\Pi(\sigma,\tau)\vee\Pi(\hat\rho) = \Pi(\hat\rho)$.

[5]The condition $\Pi(\sigma,\tau) = \Pi(\sigma\tau^{-1})$ means that the bipartite map (σ,τ^{-1}) has one face per each connected component: the faces of this map are exactly the cycles of $\sigma\tau^{-1}$, hence correspond to the blocks of $\Pi(\sigma\tau^{-1})$, while the connected components correspond to the blocks of $\Pi(\sigma,\tau)$.

- Conversely, assume that $\Pi(\sigma,\tau) > \Pi(\sigma\tau^{-1})$. We will exhibit three permutations ρ_1, ρ_2, ρ_3 such that $\rho_1\rho_2\rho_3 = \sigma\tau^{-1}$ and $\Pi(\sigma,\tau) \vee \Pi(\hat{\rho}) > \Pi(\hat{\rho})$.

  As $\Pi(\sigma,\tau) > \Pi(\sigma\tau^{-1})$ there exists a block of $\Pi(\sigma,\tau)$ containing at least two blocks B_1 and B_2 of $\Pi(\sigma\tau^{-1})$. We choose ρ_1, ρ_2, ρ_3 as follows:
  - $\rho_{1|B_1} = (\sigma\tau^{-1})_{|B_1}$ and ρ_1 is the identity on the complement of B_1,
  - $\rho_{2|B_2} = (\sigma\tau^{-1})_{|B_2}$ and ρ_2 is the identity on the complement of B_2,
  - ρ_3 is the identity on B_1 and B_2 and coincides with $\sigma\tau^{-1}$ on their complement:

$$\rho_{3|\{1,\dots n\}\setminus(B_1\cup B_2)} = (\sigma\tau^{-1})_{|\{1,\dots n\}\setminus(B_1\setminus B_2)}.$$

  Obviously $\sigma\tau^{-1} = \rho_1\rho_2\rho_3$. At the same time $\Pi(\sigma,\tau) \vee \Pi(\hat{\rho}) > \Pi(\hat{\rho})$ because the blocks B_1 and B_2 which are distinct blocks in $\Pi(\hat{\rho})$ are collapsed into one block of $\Pi(\sigma,\tau)$.

Now, at $D = 1$, for any σ and τ we have

$$m_C(\sigma,\tau;l,k) = \sum_{\substack{\rho_1,\dots,\rho_k \in S_n^*,\, \rho_1\dots\rho_k = \sigma\tau^{-1} \\ \sum_{i=1}^k \|\rho_i\| = l,\, \Pi(\sigma,\tau)\vee\Pi(\hat{\rho}) = 1_n}} 1, \tag{3.6}$$

and using on the one hand the remark above and on the other noting that $\Pi(\sigma\tau^{-1}) \le \Pi(\hat{\rho})$ we get

$$\begin{aligned} m_C(\sigma,\tau;l,k)|_{\Pi(\sigma,\tau)=\Pi(\sigma\tau^{-1})} &= m_C(\sigma\tau^{-1},\mathrm{id};l,k) = \gamma_l(\sigma\tau^{-1};k), \\ (-1)^l p_C[\sigma,\tau;l]|_{\Pi(\sigma,\tau)=\Pi(\sigma\tau^{-1})} &= (-1)^l p_C[\sigma\tau^{-1},\mathrm{id};l] = \gamma_l(\sigma\tau^{-1}), \end{aligned} \tag{3.7}$$

as in both cases the condition $\Pi(\sigma,\tau) \vee \Pi(\hat{\rho}) = 1_n$ in (3.6) reduces to $\Pi(\hat{\rho}) = 1_n$.

There is a non-signed version in terms of a generalization of monotone double Hurwitz numbers, which we describe now (see also Sec. 4.3).

Proposition 3.6. *The number $p_C[\boldsymbol{\sigma},\boldsymbol{\tau};l]$ in Thm. 3.3 is also the number of D ordered sequences of* transpositions $(\mu_1^1,\dots,\mu_{l_1}^1),\dots,(\mu_1^D,\dots,\mu_{l_D}^D)$ *such that*

- *for every $c \in \{1,\dots,D\}$, the l_c transpositions $\mu_1^c,\dots,\mu_{l_c}^c$ have weakly monotone maxima (Def. 2.2) and satisfy $\sigma_c = \mu_1^c\cdots\mu_{l_c}^c\tau_c$,*
- *$l_1 + \cdots + l_D = l$,*
- *the group generated by all the transpositions μ and all σ_c and τ_c is transitive on $\{1,\dots,n\}$.*

In particular, $p_C[\boldsymbol{\sigma},\boldsymbol{\tau};l]$ is a non-negative integer.

Proof. Note that

$$W_\pi^{(N)}[\boldsymbol{\sigma},\boldsymbol{\tau}] = \chi(\Pi(\boldsymbol{\sigma},\boldsymbol{\tau}) \le \pi) \sum_{\{l_c^B\}_{c,B}\ge 0} \sum_{\{\mu_1^{c,B},\dots,\mu_{l_c^B}^{c,B} \in P(\nu_{c|B},l_c^B)\}_{c,B}} (-N^{-1})^{\sum_{c,B} l_c^B}, \tag{3.8}$$

where $P(\nu,l)$ denotes the set of transpositions with weakly monotone maxima which factorize ν (Def. 2.2). From this point on, the proof is *mutatis mutandis* the same as that of Thm. 3.3. ∎

4. Asymptotics of cumulant Weingarten functions

In many applications, such as random tensor models, one is mostly interested in computing the large-N contribution to the logarithm $C_{D,N}(t, A, B)$ of the tensor HCIZ integral (1.1). For some given $\boldsymbol{\sigma}, \boldsymbol{\tau} \in \boldsymbol{S}_n$, one thus needs to identify the smallest integer l such that $p_C[\boldsymbol{\sigma}, \boldsymbol{\tau}; l]$ does not vanish and, if possible, to obtain an explicit expression for the corresponding p_C. We provide this in Thm. 4.1. A general combinatorial formula, (4.6), is furthermore derived for $p_C[\boldsymbol{\sigma}, \boldsymbol{\tau}; l]$ for any l. To our knowledge this expression for the sub-leading contributions to the cumulant Weingarten functions is new also for $D = 1$.

4.1. Main result

The large-N behavior of cumulant Weingarten functions is captured by the following theorem.

Theorem 4.1. *For any l, the coefficient $p_C[\boldsymbol{\sigma}, \boldsymbol{\tau}; l]$ is given by*

$$(-1)^l p_C[\boldsymbol{\sigma}, \boldsymbol{\tau}; l] = \sum_{\substack{\pi_1 \geq \Pi(\nu_1), \dots, \pi_D \geq \Pi(\nu_D) \\ \Pi(\boldsymbol{\sigma}, \boldsymbol{\tau}) \vee \pi_1 \vee \dots \vee \pi_D = 1_n}} \sum_{\substack{l_1, \dots, l_D \geq 0 \\ \sum l_c = l}} \prod_{c=1}^{D} \Big[\sum_{\substack{\{l_{B_c}\}_{B_c \in \pi_c} \geq 0 \\ \sum_{B_c} l_{B_c} = l_c}} \prod_{B_c \in \pi_c} \gamma_{l_{B_c}}(\nu_{c|B_c}) \Big], \tag{4.1}$$

with $\nu_c = \sigma_c \tau_c^{-1}$ and $\gamma_l(\nu)$ defined in Sec. 2.5. The smallest value of l such that $p_C[\boldsymbol{\sigma}, \boldsymbol{\tau}; l]$ does not vanish is

$$\ell(\boldsymbol{\sigma}, \boldsymbol{\tau}) = \sum_{c=1}^{D} \|\sigma_c \tau_c^{-1}\| + 2(|\Pi(\boldsymbol{\sigma}, \boldsymbol{\tau})| - 1). \tag{4.2}$$

In order to simplify the notation we sometimes denote $\ell \equiv \ell(\boldsymbol{\sigma}, \boldsymbol{\tau})$. The cumulant Weingarten functions thus have the asymptotic expression

$$W_C^{(N)}[\boldsymbol{\sigma}, \boldsymbol{\tau}] = N^{-nD-\ell}(-1)^\ell p_C[\boldsymbol{\sigma}, \boldsymbol{\tau}; \ell](1 + O(N^{-2})), \tag{4.3}$$

where the leading coefficient is

$$(-1)^\ell p_C[\boldsymbol{\sigma}, \boldsymbol{\tau}, \ell] = \sum_{\substack{\pi_1 \geq \Pi(\nu_1), \dots, \pi_D \geq \Pi(\nu_D) \\ \Pi(\boldsymbol{\sigma}, \boldsymbol{\tau}) \vee \pi_1 \vee \dots \vee \pi_D = 1_n \\ \sum_c (|\Pi(\nu_c)| - |\pi_c|) = |\Pi(\boldsymbol{\sigma}, \boldsymbol{\tau})| - 1}} \prod_{c=1}^{D} \prod_{B_c \in \pi_c} \gamma(\nu_{c|B_c}). \tag{4.4}$$

Note that $\nu_{c|B_c}$ is well-defined as $\pi_c \geq \Pi(\nu_c)$. In detail,

$$\gamma(\nu_{c|B_c}) = \mathsf{M}(\nu_{c|B_c}) \frac{(3|B_c| - \|\nu_{c|B_c}\| - 3)!}{(2|B_c|)!} \prod_{p \geq 1} [2p(2p-1)]^{d_p(\nu_{c|B_c})}, \tag{4.5}$$

with the non-crossing Möbius function M defined in (2.7).

In general, for $l > \ell(\boldsymbol{\sigma}, \boldsymbol{\tau})$, (4.1) can be rewritten as

$$(-1)^l p_C[\boldsymbol{\sigma}, \boldsymbol{\tau}; l] = \sum_{L=0}^{\frac{l-\ell}{2}} \sum_{\substack{\pi_1 \geq \Pi(\nu_1), \ldots, \pi_D \geq \Pi(\nu_D) \\ \Pi(\boldsymbol{\sigma},\boldsymbol{\tau}) \vee \pi_1 \vee \cdots \vee \pi_D = 1_n \\ \sum_c (|\Pi(\nu_c)| - |\pi_c|) = |\Pi(\boldsymbol{\sigma},\boldsymbol{\tau})| + L - 1}} \sum_{\substack{g_1, \ldots, g_D \geq 0 \\ \sum_c g_c = \frac{l-\ell}{2} - L}} \prod_{c=1}^{D} \Big[\sum_{\substack{\{g_{B_c}\}_{B_c \in \pi_c} \geq 0 \\ \sum_{B_c} g_{B_c} = g_c}} \prod_{B_c \in \pi_c} \gamma_{l(g_{B_c})}(\nu_{c|B_c}) \Big], \tag{4.6}$$

where $l(g_{B_c}) = |B_c| + \Pi(\nu_{c|B_c}) + 2g_{B_c} - 2$.

Proof. See Sec. 4.2. ■

The sum over partitions appears rather complicated, but it has a simple graphical interpretation in terms of sums of trees. This interpretation was developed in [46] for $D = 1$ and $l = \ell(\sigma, \tau)$ and is generalized in this paper to larger l and larger D in Sec. 5.3.2.

Corollary 3.4 implies that if $|\Pi(\boldsymbol{\sigma}, \boldsymbol{\tau})| = 1$ then $(-1)^\ell p_C[\boldsymbol{\sigma}, \boldsymbol{\tau}, \ell] = \prod_{c=1}^{D} \mathsf{M}(\nu_c)$. We have chosen to factor the Möbius functions to render this explicit. This can be obtained directly from (4.4): as $\pi_c \geq \Pi(\nu_c)$, we have $|\Pi(\nu_c)| \geq |\pi_c|$, and from the condition in the sum, $|\Pi(\nu_c)| = |\pi_c|$ so that only $\pi_c = \Pi(\nu_c)$ contributes, and $\nu_{c|B_c}$ is the restriction of ν_c to one of its cycles:

$$\frac{(3|B_c| - \|\nu_{c|B_c}\| - 3)!}{(2|B_c|)!} \prod_{p \geq 1} [2p(2p-1)]^{d_p(\nu_{c|B_c})} = 1.$$

The leading contribution at large N to the cumulant is

$$C\left([\mathrm{Tr}(AUBU^*)]^n\right) = N^{-nD} \sum_{\boldsymbol{\sigma}, \boldsymbol{\tau} \in S_n} \mathrm{Tr}_{\boldsymbol{\sigma}}(A)\, \mathrm{Tr}_{\boldsymbol{\tau}^{-1}}(B) N^{-\sum_{c=1}^{D} \|\sigma_c \tau_c^{-1}\| - 2(|\Pi(\boldsymbol{\sigma},\boldsymbol{\tau})| - 1)} \cdot (-1)^\ell p_C[\boldsymbol{\sigma}, \boldsymbol{\tau}, \ell]\, (1 + O(N^{-2})). \tag{4.7}$$

As a result of the scaling behavior of the trace invariants $\mathrm{Tr}_{\boldsymbol{\sigma}}(A)$ and $\mathrm{Tr}_{\boldsymbol{\tau}^{-1}}(B)$ in terms of N, the sum in (4.7) is dominated by a subset of its terms. For instance if $\mathrm{Tr}_{\boldsymbol{\sigma}}(A) \sim \mathrm{Tr}_{\boldsymbol{\tau}^{-1}}(B) \sim O(1)$ in the limit of large N, then the term with $\boldsymbol{\sigma} = \boldsymbol{\tau}$, $\|\Pi(\boldsymbol{\sigma})\| = 1$ will dominate. If $\mathrm{Tr}_{\boldsymbol{\sigma}}(A) \sim 1$ but $\mathrm{Tr}_{\boldsymbol{\tau}^{-1}}(B) \sim N^{\sum_c \#(\tau_c)}$, more terms dominate at large N. A detailed study of the possible behaviors of the cumulant, relevant for different applications to physics, will be conducted in future work.

4.2. *Proof of Theorem 4.1*

The proof of Theorem 4.1 is divided into four parts:

- *Derivation of* (4.1). Although lengthy, this part is straightforward: we compute the sum over k in Thm. 3.3. We obtain (4.1) where the right hand side is an alternating sum (with constraints) over constellations.

- *Reinterpretation of* $\Pi(\boldsymbol{\sigma},\boldsymbol{\tau}) \vee \pi_1 \vee \cdots \vee \pi_D = 1_n$. We show that the condition

$$\Pi(\boldsymbol{\sigma},\boldsymbol{\tau}) \vee \pi_1 \vee \cdots \vee \pi_D = 1_n,$$

  which constrains the sum over $\{\pi_c\}_c$ in (4.1), is equivalent to requiring that a certain abstract graph, denoted by $G[\Pi(\boldsymbol{\sigma},\boldsymbol{\tau}), \{\pi_c\}_c; \Pi(\nu_c)]$, is connected.
- *The lower bound* (4.2) *on* l. We show that the leading order at large N (minimal l) in (4.1) fulfills two conditions:
  - the abstract graph $G[\Pi(\boldsymbol{\sigma},\boldsymbol{\tau}), \{\pi_c\}_c; \Pi(\nu_c)]$ has minimal number of edges; as it is connected, this means it is a tree;
  - the constellations are planar.

  We will show in Section 5 that these two conditions translate in fact into the planarity of a certain nodal surface.
- *Proof of* (4.6). We rearrange the terms in (4.1) according to the number of excess edges of the abstract graph $G[\Pi(\boldsymbol{\sigma},\boldsymbol{\tau}), \{\pi_c\}_c; \Pi(\nu_c)]$ (i.e. the number of independent cycles, or loop edges in the physics literature) and to the genera of the constellations to get (4.6).

Derivation of (4.1). Our starting point is Thm. 3.3, which states that

$$W_C^{(N)}[\boldsymbol{\sigma},\boldsymbol{\tau}] = N^{-nD} \sum_{l\geq 0} (-1)^l p_C[\boldsymbol{\sigma},\boldsymbol{\tau}; l] N^{-l},$$
$$(-1)^l p_C[\boldsymbol{\sigma},\boldsymbol{\tau}, l] = \sum_{k\geq 0} (-1)^k m_C(\boldsymbol{\sigma},\boldsymbol{\tau}; l, k),$$

where, denoting $\nu_c = \sigma_c \tau_c^{-1}$, we have

$$m_C(\boldsymbol{\sigma},\boldsymbol{\tau}; l, k) = \sum_{\{k_c\}_c \geq 0} \sum_{\substack{\{\rho_1^c,\ldots,\rho_{k_c}^c \in S_n^*, \rho_1^c \cdots \rho_{k_c}^c = \nu_c\}_c \\ \sum_c \sum_{i_c=1}^{k_c} \|\rho_{i_c}^c\| = l,\, \sum_c k_c = k \\ \Pi(\boldsymbol{\sigma},\boldsymbol{\tau}) \vee \Pi(\hat{\boldsymbol{\rho}}) = 1_n}} 1,$$

and $\hat{\boldsymbol{\rho}}$ denotes the D-uple of constellations $(\hat{\rho}^1, \ldots, \hat{\rho}^D)$ where, for $c \in \{1, \ldots, D\}$, the constellation $\hat{\rho}^c$ is $(\rho_1^c, \ldots, \rho_{k_c}^c)$. We aim to derive the asymptotic behavior of $W_C^{(N)}[\boldsymbol{\sigma},\boldsymbol{\tau}]$ using the results in Sec. 2.5. Let us classify the terms in the above formula by the values $l_c = \sum_{i_c=1}^{k_c} \|\rho_{i_c}^c\|$ and by the partitions $\pi_c = \Pi(\hat{\rho}^c) \geq \Pi(\nu_c)$:

$$m_C(\boldsymbol{\sigma},\boldsymbol{\tau}; l, k) = \sum_{\substack{\pi_1 \geq \Pi(\nu_1), \ldots, \pi_D \geq \Pi(\nu_D) \\ \Pi(\boldsymbol{\sigma},\boldsymbol{\tau}) \vee \pi_1 \vee \cdots \vee \pi_D = 1_n}} \sum_{\substack{l_1,\ldots,l_D \geq 0 \\ \sum_c l_c = l}} \sum_{\substack{\{k_c\}_c \geq 0 \\ \sum_c k_c = k}} \prod_{c=1}^{D} M(\pi_c, \nu_c; l_c, k_c),$$
$$M(\pi_c, \nu_c; l_c, k_c) = \sum_{\substack{\rho_1^c,\ldots,\rho_{k_c}^c \in S_n^*, \rho_1^c \cdots \rho_{k_c}^c = \nu_c \\ \Pi(\hat{\rho}^c) = \pi_c,\, \sum_{i_c=1}^{k_c} \|\rho_{i_c}^c\| = l_c}} 1. \tag{4.8}$$

We wish to compute

$$(-1)^l p_C[\boldsymbol{\sigma},\boldsymbol{\tau},l] = \sum_{k\geq 0}(-1)^k m_C(\boldsymbol{\sigma},\boldsymbol{\tau};l,k)$$
$$= \sum_{\substack{\pi_1\geq\Pi(\nu_1),\dots,\pi_D\geq\Pi(\nu_D)\\ \Pi(\boldsymbol{\sigma},\boldsymbol{\tau})\vee\pi_1\vee\cdots\vee\pi_D=1_n}} \sum_{\substack{l_1,\dots,l_D\geq 0\\ \sum_c l_c=l}} \prod_{c=1}^{D}\sum_{\{k_c\}_c\geq 0}(-1)^{k_c}M(\pi_c,\nu_c;l_c,k_c). \tag{4.9}$$

The sum $\sum_{k\geq 0}(-1)^k M(\pi,\nu;l,k)$. Let us focus on the rightmost sum of (4.9). We define

$$S[\pi,\nu;l] = \sum_{k\geq 0}(-1)^k M(\pi,\nu;l,k), \quad M(\pi,\nu;l,k) = \sum_{\substack{\rho_1,\dots,\rho_k\in S_n^*,\,\rho_1\cdots\rho_k=\nu\\ \sum_i\|\rho_i\|=l,\,\Pi(\hat\rho)=\pi}} 1,$$
$$\tilde M(\pi,\nu;l,k) = \sum_{\substack{\rho_1,\dots,\rho_k\in S_n,\,\rho_1\cdots\rho_k=\nu\\ \sum_i\|\rho_i\|=l,\,\Pi(\hat\rho)=\pi}} 1,$$

where we emphasize that, unlike M, the sum defining $\tilde M$ includes the case when some of the permutations ρ_i are the identity. M and $\tilde M$ are related by

$$\tilde M(\pi,\nu;l,q) = \sum_{k=0}^{q}\binom{q}{k}M(\pi,\nu;l,k), \quad M(\pi,\nu;l,k) = \sum_{q=0}^{k}\binom{k}{q}(-1)^{k-q}\tilde M(\pi,\nu;l,q).$$

$\tilde M(\pi,\nu;l,q)$ is a sum over q permutations ρ_i. The first equation follows by noting that if exactly $q-k$ of these permutations are the identity then $\tilde M(\pi,\nu;l,q)$ reduces to $M(\pi,\nu;l,k)$; the second equation is obtained by inverting the first one.

The point is that $\tilde M$ factors over the blocks of π. As ν (respectively ρ_i) stabilizes any block $B\in\pi$, it can be decomposed as the product of $|\pi|$ permutations $\nu_{|B}$ (respectively $\rho_{i|B}$), $\nu=\prod_B\nu_{|B}$ (respectively $\rho_i=\prod_B\rho_{i|B}$), where we trivially lift $\nu_{|B}$ (respectively $\rho_{i|B}$) to the whole set $\{1,\dots,n\}$. The number of transpositions of ρ_i is distributed among the blocks of π, $\|\rho_i\|=\sum_B\|\rho_{i|B}\|$, and we get

$$\sum_{\substack{\rho_1,\dots,\rho_q\in S_n,\,\rho_1\cdots\rho_q=\nu\\ \sum_i\|\rho_i\|=l,\,\Pi(\hat\rho)=\pi}} 1 = \sum_{\substack{\{l_B\}_B\geq 0\\ \sum_B l_B=l}}\prod_{B\in\pi}\sum_{\substack{\rho_{1|B},\dots,\rho_{q|B}\in S_{|B|}\\ \rho_{1|B}\cdots\rho_{q|B}=\nu_{|B}\\ \sum_i\|\rho_{i|B}\|=l_B,\,\Pi(\hat\rho_{|B})=1_{|B|}}} 1 = \sum_{\substack{\{l_B\}_B\geq 0\\ \sum_B l_B=l}}\prod_{B\in\pi}\tilde\gamma_{l_B}(\nu_{|B};q),$$

where we recognize the number of connected q-constellations with fixed l_B, $\tilde\gamma_{l_B}(\nu_{|B};q)$ of Sec. 2.5. We emphasize that constellations are not necessarily proper (i.e. the sums run over $S_{|B|}$ and not $S^*_{|B|}$). The number of arbitrary (i.e. not necessarily proper) constellations is written in terms of the number of proper ones as

$$\tilde\gamma_{l_B}(\nu_{|B};q) = \sum_{k_B=0}^{q}\binom{q}{k_B}\gamma_{l_B}(\nu_{|B};k_B),$$

and substituting, we find

$$\begin{aligned}
S[\pi,\nu;l] &= \sum_{k\geq 0}\sum_{q=0}^{k}\binom{k}{q}(-1)^q\tilde{M}(\pi,\nu;l,q)\\
&= \sum_{\substack{\{l_B\}_B\geq 0\\ \sum_B l_B=l}}\sum_{k\geq 0}\sum_{q=0}^{k}\binom{k}{q}(-1)^q\prod_{B\in\pi}\tilde{\gamma}_{l_B}(\nu_{|B};q)\\
&= \sum_{\substack{\{l_B\}_B\geq 0\\ \sum_B l_B=l}}\sum_{\{k_B\}_B\geq 0}\Big(\prod_{B\in\pi}\gamma_{l_B}(\nu_{|B};k_B)\Big)\Big[\sum_{k\geq 0}\sum_{q=0}^{k}\binom{k}{q}(-1)^q\prod_{B\in\pi}\binom{q}{k_B}\Big],
\end{aligned}$$

with the convention that ill-defined binomial coefficients (e.g. with $k_B > q$) are zero. At fixed k_B, the sum over q and k can be computed as it is the coefficient of the monomial $\prod_B x_B^{k_B}$ in the generating function:

$$\begin{aligned}
\sum_{k\geq 0}\sum_{q=0}^{k}\binom{k}{q}(-1)^q\prod_B\Big[\sum_{j_B=0}^{q}\binom{q}{j_B}x_B^{j_B}\Big] &= \sum_{k\geq 0}\Big[1-\prod_B(1+x_B)\Big]^k\\
&= \sum_{\{k_B\}_B\geq 0}\prod_B(-1)^{k_B}x_B^{k_B},
\end{aligned}$$

which ultimately leads to

$$S[\pi,\nu;l] = \sum_{\substack{\{l_B\}_B\geq 0\\ \sum_B l_B=l}}\sum_{\{k_B\}_B\geq 0}\prod_{B\in\pi}(-1)^{k_B}\gamma_{l_B}(\nu_{|B};k_B) = \sum_{\substack{\{l_B\}_B\geq 0\\ \sum_B l_B=l}}\prod_{B\in\pi}\gamma_{l_B}(\nu_{|B}).$$

Inserting this expression in (4.9) completes the proof of (4.1):

$$(-1)^l p_C[\boldsymbol{\sigma},\boldsymbol{\tau};l] = \sum_{\substack{\pi_1\geq\Pi(\nu_1),\dots,\pi_D\geq\Pi(\nu_D)\\ \Pi(\boldsymbol{\sigma},\boldsymbol{\tau})\vee\pi_1\vee\cdots\vee\pi_D=1_n}}\sum_{\substack{l_1,\dots,l_D\geq 0\\ \sum l_c=l}}\prod_{c=1}^{D}\Big[\sum_{\substack{\{l_{B_c}\}_{B_c\in\pi_c}\geq 0\\ \sum_{B_c} l_{B_c}=l_c}}\prod_{B_c\in\pi_c}\gamma_{l_{B_c}}(\nu_{c|B_c})\Big].$$

This formula can be analyzed further.

The graph $G[\Pi,\{\pi_c\}_c;\{\Pi_c\}_c]$. Consider $2D+1$ partitions $\Pi,\{\pi_c\}_c,\{\Pi_c\}_c$ on $\{1,\dots,n\}$ such that

$$\forall c,\quad \Pi\geq\Pi_c \quad\text{and}\quad \pi_c\geq\Pi_c.$$

The relation between these partitions can be encoded in a convenient graphical representation. Let $G[\Pi,\{\pi_c\}_c;\{\Pi_c\}_c]$ be the abstract bipartite graph consisting of

- white vertices associated to the blocks B of the partition Π,
- c-colored vertices associated to the blocks B_c of the partitions π_c,
- c-colored edges associated to the blocks b_c of Π_c linking a white and a c-colored vertex; the block b_c is at the same time

– contained in a block of Π, which we denote by $B(b_c)$, as $\Pi \geq \Pi_c$,

– contained in a block of π_c, which we denote by $B_c(b_c)$, as $\pi_c \geq \Pi_c$.

The edge corresponding to b_c links $B(b_c)$ to $B_c(b_c)$.

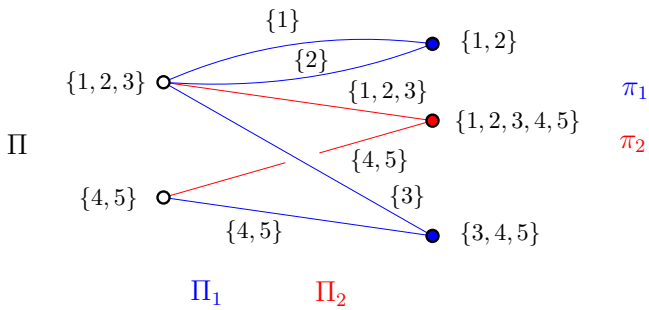

Fig. 2. The graph $G[\Pi, \{\pi_c\}_c; \{\Pi_c\}_c]$ for $\Pi = \{\{1,2,3\},\{4,5\}\}$, $\pi_1 = \{\{1,2\},\{3,4,5\}\}$, $\pi_2 = \{\{1,2,3,4,5\}\}$ and $\Pi_1 = \{\{1\},\{2\},\{3\},\{4,5\}\}$, $\Pi_2 = \{\{1,2,3\},\{4,5\}\}$. Here $D = 2$ (two colors), $n = 5$.

Lemma 4.2. *The graph $G[\Pi, \{\pi_c\}_c; \{\Pi_c\}_c]$ has $|\Pi \vee \pi_1 \vee \cdots \vee \pi_D|$ connected components.*

Proof. Two blocks $B \in \Pi$ and $B_c \in \pi_c$ are connected by an edge in $G[\Pi, \{\pi_c\}_c; \{\Pi_c\}_c]$ if and only if there exists a $b_c \in \Pi_c$ such that $b_c \subset B, B_c$, hence both B and B_c belong to the block of $\Pi \vee \pi_1 \vee \cdots \vee \pi_D$ which contains b_c. The lemma follows by noting that

- belonging to the same connected component of $G[\Pi, \{\pi_c\}_c; \{\Pi_c\}_c]$ and
- belonging to the same block of the partition $\Pi \vee \pi_1 \vee \cdots \vee \pi_D$

are both transitive relations between the blocks of Π and $\{\pi_c\}_c$. ∎

Lower bound on l. In order to find a lower bound on l, we first rewrite $\sum_c |\pi_c|$. Observe that in (4.1) we sum over partitions $\{\pi_c\}_c$ such that

$$\forall c \quad \pi_c \geq \Pi(\nu_c), \quad \text{and} \quad \Pi(\boldsymbol{\sigma}, \boldsymbol{\tau}) \vee \pi_1 \vee \cdots \vee \pi_D = 1_n.$$

Since $\Pi(\boldsymbol{\sigma}, \boldsymbol{\tau}) \geq \Pi(\nu_c)$, from Lemma 4.2 we conclude that the sum runs over partitions $\{\pi_c\}_c$ such that the graph $G[\Pi(\boldsymbol{\sigma}, \boldsymbol{\tau}), \{\pi_c\}_c; \Pi(\nu_c)]$ is connected.

The graph $G[\Pi(\boldsymbol{\sigma}, \boldsymbol{\tau}), \{\pi_c\}_c; \Pi(\nu_c)]$ has $\sum_c |\Pi(\nu_c)|$ edges and $|\Pi(\boldsymbol{\sigma}, \boldsymbol{\tau})| + \sum_c |\pi_c|$ vertices. If it is connected then any tree spanning this graph will have exactly $|\Pi(\boldsymbol{\sigma}, \boldsymbol{\tau})| + \sum_c |\pi_c| - 1$ edges. We denote the number of *excess edges* of $G[\Pi(\boldsymbol{\sigma}, \boldsymbol{\tau}), \{\pi_c\}_c; \{\Pi(\nu_c)\}_c]$, that is, the number of edges in the complement of a spanning tree in the graph, by

$$L[\Pi(\boldsymbol{\sigma}, \boldsymbol{\tau}), \{\pi_c\}_c; \{\Pi(\nu_c)\}_c] = \sum_c |\Pi(\nu_c)| - |\Pi(\boldsymbol{\sigma}, \boldsymbol{\tau})| - \sum_c |\pi_c| + 1 \geq 0. \tag{4.10}$$

Let us consider a term in (4.1). At fixed l_{B_c}, each block B_c of π_c contains a sum over constellations with fixed genus because, from (2.11),

$$l_{B_c} = 2|B_c| - \|\nu_{c|B_c}\| + 2g_{B_c} - 2, \quad l_c = 2n - \|\nu_c\| + 2\sum_{B_c\in\pi_c} g_{B_c} - 2|\pi_c|. \tag{4.11}$$

Summing over c and using $n = \|\nu_c\| + |\Pi(\nu_c)|$ we get

$$l = \sum_c \|\nu_c\| + 2(|\Pi(\boldsymbol{\sigma},\boldsymbol{\tau})| - 1) + 2\sum_{c=1}^{D}\sum_{B_c\in\pi_c} g_{B_c} + 2L[\Pi(\boldsymbol{\sigma},\boldsymbol{\tau}), \{\pi_c\}_c; \{\Pi(\nu_c)\}_c]. \tag{4.12}$$

Thus $l \geq \ell(\boldsymbol{\sigma},\boldsymbol{\tau})$, leading to (4.2), which proves (4.3). We see that the bound $l = \ell(\boldsymbol{\sigma},\boldsymbol{\tau})$ is attained if and only if $g_{B_c} = 0$ for all B_c, and

$$L[\Pi(\boldsymbol{\sigma},\boldsymbol{\tau}), \{\pi_c\}_c; \{\Pi(\nu_c)\}_c] = 0 \implies \sum_c |\Pi(\nu_c)| - \sum_c |\pi_c| = |\Pi(\boldsymbol{\sigma},\boldsymbol{\tau})| - 1,$$

which proves (4.4).

Remark 4.3. The condition $g_{B_c} = 0$ for all B_c already suggests that the large-N limit corresponds to some type of planarity. It turns out that the additional condition $L[\Pi(\boldsymbol{\sigma},\boldsymbol{\tau}), \{\pi_c\}_c; \{\Pi(\nu_c)\}_c] = 0$ is also a "minimal genus" condition. In Section 5 we will show that the sum in (4.1) can be reinterpreted as a sum over a class of *nodal* surfaces and at leading order at large-N only nodal surfaces of minimal arithmetic genus contribute.

Formula for $l > \ell \equiv \ell(\boldsymbol{\sigma},\boldsymbol{\tau})$. More generally for non-minimal l, we can rearrange the sum in (4.1) according to the number of excess edges of the graph $G[\Pi(\boldsymbol{\sigma},\boldsymbol{\tau}), \{\pi_c\}_c; \{\Pi(\nu_c)\}_c]$. Writing (4.12) as

$$\frac{l-\ell}{2} = \sum_{c=1}^{D}\sum_{B_c\in\pi_c} g_{B_c} + L[\Pi(\boldsymbol{\sigma},\boldsymbol{\tau}), \{\pi_c\}_c; \{\Pi(\nu_c)\}_c], \tag{4.13}$$

we rearrange (4.1) as

$$(-1)^l p_C[\boldsymbol{\sigma},\boldsymbol{\tau}; l] = \sum_{\substack{\pi_1\geq\Pi(\nu_1),\dots,\pi_D\geq\Pi(\nu_D)\\ \Pi(\boldsymbol{\sigma},\boldsymbol{\tau})\vee\pi_1\vee\dots\vee\pi_D=1_n}} \sum_{\substack{l_1,\dots,l_D\geq 0\\ \sum l_c = l}} \prod_{c=1}^{D}\Big[\sum_{\substack{\{l_{B_c}\}_{B_c\in\pi_c}\geq 0\\ \sum_{B_c} l_{B_c} = l_c}} \prod_{B_c\in\pi_c} \gamma_{l_{B_c}}(\nu_{c|B_c})\Big],$$

by replacing the sum over the partitions $\{\pi_c\}_c$ with a sum over $\{\pi_c\}_c$ such that $L = \sum_c(|\Pi(\nu_c)| - |\pi_c|) - \Pi(\boldsymbol{\sigma},\boldsymbol{\tau}) + 1$ is fixed and a sum over L from its minimal allowed value 0 to the maximal allowed value $(l-\ell)/2$ fixed by (4.13). Then at fixed L, we use (4.11) and replace the sums over $\{l_c\}_c$ with $\sum_c l_c = l$ with sums over the genera $g_{B_c} \geq 0$ constrained by (4.11) to obey

$$\sum_{c=1}^{D}\sum_{B_c\in\pi_c} g_{B_c} = \frac{l-\ell}{2} - L.$$

At the end, we get

$$(-1)^l p_C[\boldsymbol{\sigma}, \boldsymbol{\tau}; l] =$$
$$\sum_{L=0}^{\frac{l-\ell}{2}} \sum_{\substack{\pi_1 \geq \Pi(\nu_1), \ldots, \pi_D \geq \Pi(\nu_D) \\ \Pi(\boldsymbol{\sigma},\boldsymbol{\tau}) \vee \pi_1 \vee \cdots \vee \pi_D = 1_n \\ \sum_c (|\Pi(\nu_c)| - |\pi_c|) = |\Pi(\boldsymbol{\sigma},\boldsymbol{\tau})| + L - 1}} \sum_{\substack{g_1, \ldots, g_D \geq 0 \\ \sum_c g_c = \frac{l-\ell}{2} - L}} \prod_{c=1}^{D} \Big[\sum_{\substack{\{g_{B_c}\}_{B_c \in \pi_c} \geq 0 \\ \sum_{B_c} g_{B_c} = g_c}} \prod_{B_c \in \pi_c} \gamma_{l(g_{B_c})}(\nu_{c|B_c}) \Big],$$

where $l(g_{B_c}) = |B_c| + \Pi(\nu_{c|B_c}) + 2g_{B_c} - 2$, which is (4.6).

This completes the proof of Thm. 4.1. ∎

Corollary 4.4. *For $D = 1$, consider all the possible σ, τ with fixed numbers of cycles $\#(\sigma)$ and $\#(\tau)$. Consider an arbitrary $k \geq 0$ and a k-constellation as in Thm. 3.3, with $l = \#(\sigma) + \#(\tau) - 2$. Then $g(\sigma, \tau^{-1}) = 0$ (as a bipartite map), and $l = \ell(\sigma, \tau)$.*

Proof. For $D = 1$, denoting $\pi = \Pi(\hat{\rho})$, (4.12) becomes

$$l = \ell(\sigma, \tau) + 2 \sum_{B \in \pi} g_B + 2L[\Pi(\sigma, \tau), \pi; \Pi(\nu)], \ell(\sigma, \tau) = \|\sigma\tau^{-1}\| + 2(|\Pi(\sigma, \tau)| - 1).$$

The Euler relation of the 2-constellation (σ, τ^{-1}) is $\|\sigma\tau^{-1}\| = \#(\sigma) + \#(\tau) + 2g(\sigma, \tau^{-1}) - 2|\Pi(\sigma, \tau)|$, so that the equation above becomes

$$l = \#(\sigma) + \#(\tau) - 2 + 2(g(\sigma, \tau^{-1}) + \sum_{B \in \pi} g_B + L[\Pi(\sigma, \tau), \pi; \Pi(\nu)]). \tag{4.14}$$

∎

4.3. Monotone Hurwitz numbers and their generalization

For $D = 1$, monotone double Hurwitz numbers are obtained by summing the coefficients $p_C[\sigma, \tau; l]$ for σ, τ of fixed cycle types. These numbers have an interpretation in enumerative geometry. We detail some known facts about monotone Hurwitz numbers, and review the results of our paper in this context. A generalization of monotone double Hurwitz numbers is then introduced, using the coefficients p_C for $D > 1$.

Monotone double Hurwitz numbers. The definition of the number $p_C[\boldsymbol{\sigma}, \boldsymbol{\tau}; l]$ in Prop. 3.6 resembles for $D = 1$ the combinatorial definition of monotone Hurwitz numbers. It is known that for $D = 1$ and A and B having asymptotic traces of order N, the HCIZ integral has the expansion [25, 26]

$$\frac{1}{N^2} \log I_{1,N}(t, A, B)$$
$$= \sum_{n=1}^{N} \frac{t^n}{n!} \sum_{h \geq 0} \frac{1}{N^{2h}} \sum_{\alpha, \beta \vdash n} (-1)^{\#(\alpha)+\#(\beta)} \vec{H}_h(\alpha, \beta) \frac{\mathrm{Tr}_\alpha(A)}{N^{\#(\alpha)}} \frac{\mathrm{Tr}_\beta(B)}{N^{\#(\beta)}} + O(t^{N+1}), \tag{4.15}$$

where α, β are partitions of n, $n = \sum_p p \cdot d_p(\alpha)$, where $d_p(\alpha)$ is the number of parts of α of size p, and $\#(\alpha) = \sum_p d_p(\alpha)$ denotes the total number of parts of α.

Denoting by C_α the set of permutations having fixed cycle type α, the coefficients

$$\vec{H}_h(\alpha,\beta) = \sum_{\sigma\in C_\alpha}\sum_{\tau\in C_\beta} p_C[\sigma,\tau,l], \quad l = \#(\alpha)+\#(\beta)+2h-2, \tag{4.16}$$

are the genus-h *monotone double Hurwitz numbers*. In detail these numbers count the number of $l+2$-uplets of permutations $\sigma,\tau,\mu_1,\dots,\mu_l$ with $\sigma\in C_\alpha$, $\tau\in C_\beta$, and μ_i transpositions with weakly monotone maxima (Def. 2.2) such that $\sigma=\mu_1\cdots\mu_l\tau$ and furthermore the group generated by σ and all the transpositions acts transitively on $\{1,\dots,n\}$.

Divided by $n!$, these numbers also count weighted branched coverings of the Riemann sphere by a surface of genus h with $l+2$ branch points, l of which have simple ramifications (that is, they have $n-1$ preimages), and the ramifications profiles at 0 and infinity are given respectively by the partitions α and β (for more details, see Sec. 5.2). The condition that the transpositions have weakly monotone maxima restricts the admissible coverings. The formula relating l (the number of simple branch points) and h is the well-known Riemann–Hurwitz formula. Note that if $\sigma=\mu_1\cdots\mu_l\tau$ and the group generated by σ and all the transpositions acts transitively on $\{1,\dots,n\}$, then applying the Euler characteristic formula (2.10) to the constellations (σ,τ^{-1}), $(\mu_1,\dots,\mu_l)$, and $(\sigma,\mu_1,\dots,\mu_l)$ (the last one is connected) yields

$$h = g(\sigma,\mu_1,\dots,\mu_l) = g(\sigma,\tau^{-1}) + g(\mu_1,\dots,\mu_l) + L\big[\Pi(\sigma,\tau),\pi;\Pi(\sigma\tau^{-1})\big], \tag{4.17}$$

where $\pi=|\Pi(\mu_1,\dots,\mu_l)|$, and we recall that L is given by (4.10).

The value of l for $h=0$ is fixed by the Riemann–Hurwitz formula as $l=\#(\alpha)+\#(\beta)-2=\#(\sigma)+\#(\tau)-2$ for any $(\sigma,\tau)\in C_\alpha\times C_\beta$. From Corollary 4.4, this restricts the sum in (4.16) to σ,τ satisfying $g(\sigma,\tau^{-1})=0$ and $l=\ell(\sigma,\tau)$ so that

$$\vec{H}_0(\alpha,\beta) = \sum_{\substack{(\sigma,\tau)\in C_\alpha\times C_\beta\\ g(\sigma,\tau^{-1})=0}} p_C[\sigma,\tau,\ell], \quad \ell=\#(\alpha)+\#(\beta)-2, \tag{4.18}$$

and we can use Thm. 4.1 in $D=1$ (see also [13]) to express $\vec{H}_0(\alpha,\beta)$ as

$$\begin{aligned}\vec{H}_0(\alpha,\beta)&\\ &= \sum_{\substack{(\sigma,\tau)\in C_\alpha\times C_\beta\\ g(\sigma,\tau^{-1})=0}} \prod_{p=1}^{n}\left(\frac{(2p)!}{p!(p-1)!}\right)^{d_p(\sigma\tau^{-1})} \sum_{\substack{\pi\geq\Pi(\nu)\\ \Pi(\sigma,\tau)\vee\pi=1_n\\ \#(\nu)-|\pi|=|\Pi(\sigma,\tau)|-1}} \prod_{B\in\pi}\frac{(2|B|+\#(\nu_{|B})-3)!}{(2|B|)!}.\end{aligned} \tag{4.19}$$

In contrast to single Hurwitz numbers (see (4.20) below), one cannot eliminate the sum over permutations, since both the number of connected components and the genus of (σ,τ^{-1}) depend on the specific representatives $(\sigma,\tau)\in C_\alpha\times C_\beta$ and not only on their conjugacy classes α and β.[6]

[6]For instance, ((12)(3)(4), (12)(3)(4)) and ((1)(2)(34), (12)(3)(4)) have respectively 3 and 2 connected components, while ((123), (321)) and ((123), (312)) have respectively genus 0 and 1.

Monotone single Hurwitz numbers. Single Hurwitz numbers are obtained when taking $\tau = \mathrm{id}_n$ (or similarly for σ) above:

$$\vec{H}_h(\alpha) = \sum_{\sigma \in C_\alpha} p_C[\sigma, \mathrm{id}_n; l] = |C_\alpha| \times p_C[\alpha, \mathrm{id}_n; l], \quad l = \#(\alpha) + n + 2h - 2, \tag{4.20}$$

where we have used the fact that $p_C[\nu, \mathrm{id}; l]$ depends only on the partition α induced by the cycles of ν and not on the specific representative $\nu \in C_\alpha$, since $m_C(\nu, \mathrm{id}, l, k)$ is invariant under conjugation (note that we have introduced the somewhat abusive notation $p_C[\alpha, \mathrm{id}_n; l]$). The cardinality of C_α is $|C_\alpha| = n!/(\prod_{p\geq 1} p^{d_p(\alpha)} d_p(\alpha)!)$. Similarly to (4.15), these numbers are obtained from the HCIZ integral for $D = 1$, but when the asymptotic moments of the matrix B are degenerate, $\lim_{N\to\infty} \frac{1}{N} \mathrm{Tr}(B^k) = \delta_{1,k}$ [26].

Combining (3.7) and (4.20) we find that for any $h \geq 0$ and $\alpha \vdash n$, *the genus-h monotone single Hurwitz numbers are, up to signs, the numbers of connected proper constellations with faces $\rho_1 \cdots \rho_k = \nu \in C_\alpha$, and $\sum_{i=1}^k \|\rho_i\| = l$ for some $\nu \in C_\alpha$:*

$$\vec{H}_h(\alpha) = (-1)^{\#(\nu)+n} |C_\alpha| \gamma_{\#(\nu)+n+2h-2}(\nu), \tag{4.21}$$

and this is independent of the representative $\nu \in C_\alpha$ chosen. Like the double numbers, they also count branched coverings of the sphere, and requiring the covering to be of genus 0 amounts to requiring that $l = \#(\sigma) + n - 2$, which is its minimal possible value, $\ell(\sigma, \mathrm{id})$. Fixing $\tau = \mathrm{id}_n$ in (4.19), as $\Pi(\nu) = \Pi(\sigma, \mathrm{id}) = \Pi(\sigma)$, the sum restricts to $\pi = 1_n$ and

$$\vec{H}_0(\alpha) = \frac{n!}{\prod_{p\geq 1} d_p(\alpha)!} \frac{(2n + \#(\alpha) - 3)!}{(2n)!} \prod_{p=1}^{n} \binom{2p}{p}^{d_p(\alpha)}, \tag{4.22}$$

recovering the value found in [23, 26].

The expressions for single Hurwitz numbers (4.22) were first obtained in [13] as sums over permutations, with p_C expressed[7] as in Thm. 3.3 and then evaluated for zero genus using the counting of planar constellations of Bousquet-Mélou–Schaeffer [5]. Higher genus monotone single Hurwitz numbers count higher genus constellations, but for now there is no simple closed formula for them. Single and double monotone Hurwitz numbers were later studied in [23–26]. To our knowledge, the result of [13] leading to (4.19) is the only explicit expression for monotone double Hurwitz numbers of genus 0.

Double numbers in terms of single ones. Note that (4.19) expresses monotone double Hurwitz numbers of genus 0 in terms of monotone single Hurwitz numbers:

$$\vec{H}_0(\alpha, \beta) = \sum_{\substack{(\sigma,\tau)\in C_\alpha \times C_\beta \\ g(\sigma,\tau)=0}} \sum_{\substack{\pi \geq \Pi(\nu) \\ \Pi(\sigma,\tau)\vee\pi = 1_n \\ \#(\nu)-|\pi| = |\Pi(\sigma,\tau)|-1}} \prod_{B\in\pi} \frac{\vec{H}_0(c(\nu_B))}{|C_{c(\nu_B)}|}, \tag{4.23}$$

[7] The expressions for $p_C[\sigma, \tau; \ell]$ and $p_C[\sigma, \mathrm{id}; \ell] = \vec{H}_0(\alpha)/|C_\alpha|$, where α is the partition of n induced by the disjoint cycles of σ, are found in (iii) and (i) of [13, Thm. 2.15]. See also [31, (5.41)].

where $\nu = \sigma\tau^{-1}$, and we denote by $c(\nu_B)$ the cycle type of $\sigma_B \tau_B^{-1}$ (the associated partition of the integer $|B|$). More generally, using (4.6) for $D = 1$ and (4.21), as well as the fact that if l is given by the Riemann–Hurwitz formula (4.16), then $\frac{l-\ell(\sigma,\tau)}{2} = h - g(\sigma, \tau^{-1})$, we get

Theorem 4.5. *For any $h \geq 0$ and $\alpha, \beta \vdash n$, the genus-h monotone double Hurwitz numbers are expressed in terms of the single ones as follows:*

$$\vec{H}_h(\alpha, \beta) = \sum_{(\sigma,\tau)\in C_\alpha \times C_\beta} \sum_{L=0}^{h-g(\sigma,\tau^{-1})} \sum_{\substack{\pi \geq \Pi(\nu) \\ \Pi(\sigma,\tau)\vee\pi = 1_n \\ \#(\nu)-|\pi| = |\Pi(\sigma,\tau)|+L-1}} \sum_{\substack{\{g_B\}_{B\in\pi}\geq 0 \\ \sum_B g_B = h-g(\sigma,\tau^{-1})-L}} \prod_{B\in\pi} \frac{\vec{H}_{g_B}(c(\nu_B))}{|C_{c(\nu_B)}|}.$$

The sum over partitions can be interpreted as a sum over all ways to add nodes to the 2-constellation (σ, τ^{-1}) with a weight per node given by monotone Hurwitz numbers, as explained in [46] for $h = 0$ and in Sec. 5 below for $h > 0$.

We thus get an expression for monotone double Hurwitz numbers of genus h for partitions of n in terms of monotone single Hurwitz numbers of genus $\leq h$, for partitions of integers $\leq n$ (see e.g. [1, 18, 28, 31]). We do not know if this relation is already known in the literature.

Higher order monotone Hurwitz numbers. The tensor generalization of the HCIZ integral naturally gives rise to the following generalization of monotone Hurwitz numbers:

$$\vec{H}^l(\alpha_1, \beta_1, \ldots, \alpha_D, \beta_D) = \sum_{\{(\sigma_c,\tau_c)\in C_{\alpha_c}\times C_{\beta_c}\}_{c=1,\ldots D}} p_C[\boldsymbol{\sigma}, \boldsymbol{\tau}; l], \tag{4.24}$$

where the combinatorial definition of $p_C[\boldsymbol{\sigma}, \boldsymbol{\tau}; l]$ in terms of factorizations of a D-uplet of permutations (Prop. 3.6) generalizes the combinatorial definition of monotone double Hurwitz numbers. Just as in Thm. 4.5, $p_C[\boldsymbol{\sigma}, \boldsymbol{\tau}, \ell]$ can be expressed as a sum over partitions of products of monotone single Hurwitz numbers by using Thm. 4.1 and (4.21).

For each color $c = 1, \ldots, D$ in Prop. 3.6, $p_C[\boldsymbol{\sigma}, \boldsymbol{\tau}; l]$ counts factorizations of permutations, that is, constellations, and thereby certain branched coverings of genera $h_c = g(\sigma_c, \mu_1, \ldots, \mu_{l_c})$, with $|\Pi(\sigma_c, \mu_1, \ldots, \mu_{l_c})|$ connected components which satisfy the Riemann–Hurwitz formula for disconnected coverings,

$$l_c = \#(\alpha_c) + \#(\beta_c) + 2h_c - 2|\Pi(\sigma_c, \mu_1, \ldots, \mu_{l_c})|.$$

However, the transitivity condition in Prop. 3.6, which involves all the permutations, for all c, imposes a global constraint on the coverings. We show in Sec. 5.2 that $\frac{1}{n!}H^l(\alpha_1, \beta_1, \ldots, \alpha_D, \beta_D)$ counts certain weighted *connected* branched coverings of D distinguishable 2-spheres *that "touch" at a single point*. The covering spaces are nodal surfaces, for which a generalization of the genus – the arithmetic genus – is kept fixed. A generalization of the Riemann–Hurwitz formula relates the arithmetic genus to the number of preimages of branch points.

5. Interpretation of the combinatorial quantities in terms of nodal surfaces

In this section we give a geometrical picture for combinatorial quantities at play:

- In Sec. 5.2, we interpret transitive factorizations of multiplets of permutations with conditions on the length of the permutations – as counted for instance by m_C (Thm. 3.3) or p_C (Prop. 3.6) – as branched coverings of D spheres touching at one common point.
- In Sec. 5.3 we develop step-by-step a geometric understanding of m_C, p_C, and of the combinatorial formula (4.6) that gives an expression for p_C in terms of single Hurwitz numbers.

Both geometric descriptions involve nodal surfaces, which are collections of surfaces that "touch" in groups at certain points called nodes. Let us provide more formal definitions.

5.1. Nodal surfaces and nodal topological constellations

Nodal surfaces. Given $p \geq 2$ topological spaces X_i, $1 \leq i \leq p$, each with a distinguished point x_i, the *wedge sum* of the spaces X_i at the points x_i is the quotient space of the disjoint union of the spaces X_i under the identifications $x_i \sim x_j$ for all $i < j \in \{1, \dots, p\}$. A wedge sum of 2-spheres is often poetically called a *bouquet* of 2-spheres.

A surface is an orientable manifold of dimension 2, together with an orientation. Given a surface $X = \bigsqcup_i X_i$ with p connected components X_i, as well as r sets of points P_j, $1 \leq j \leq r$, such that all the elements in all of the sets P_j are distinct points that may belong to any of the connected surfaces X_i (a set P_j may contain several different points from the same X_i), a *nodal surface with* $r \geq 1$ *nodes* is the quotient space of X under the identifications $x \sim y$ for all $j \in \{1, \dots, r\}$ and all $x, y \in P_j$. For each $j \in \{1, \dots, r\}$, the identification of all the points in P_j defines a *node* or *nodal point*. Such a nodal surface is then denoted by

$$X^\bullet = X/\{P_j\}_{1 \leq j \leq r},$$

and the X_i are said to be its *irreducible components*. A wedge sum of surfaces is a nodal surface with one nodal point but the converse is not always true: a surface may have two or more distinct points in a node.

A nodal surface is said to be *connected* if for any two points, there exists a path between them, the path being allowed to jump from one surface to another through a nodal point at which they touch.

A map $F : X^\bullet \to Y^\bullet$ between two nodal surfaces is said to be a *homeomorphism* if

- the restriction of F to each irreducible component of $X^\bullet$ is well-defined and is a homeomorphism between surfaces,
- F preserves the identifications for each node (that is, the points in the irreducible components of the codomain $Y^\bullet$ that are identified in a given node are exactly the images of the points that belong to the irreducible components of the domain $X^\bullet$ that are identified in a node of $X^\bullet$).

Arithmetic genus. The *arithmetic genus* of a connected nodal surface $X^\bullet = \bigsqcup_{i=1}^p X_i/\{P_j\}_{1\le j\le r}$ with p irreducible components X_i and r nodes is

$$\mathcal{G}(X^\bullet) = \sum_{i=1}^p g(X_i) + \mathcal{L}(X^\bullet), \quad \mathcal{L}(X^\bullet) = \sum_{j=1}^r [\mathrm{Card}(P_j) - 1] - p + 1, \tag{5.1}$$

where $g(X_i)$ is the genus of X_i and $\mathcal{L}(X^\bullet)$ is the rank of the first homology group of $X^\bullet$. This is also the number of excess edges of the abstract graph that has a point vertex for each node P_j, a square vertex for each X_i, and an edge between a point vertex and a square vertex if the corresponding node belongs to the corresponding irreducible component X_i, and $P_j \cap X_i \neq \emptyset$. The arithmetic genus is the genus obtained by "smoothing" the nodes, whereas $\sum_{i=1}^p g(X_i)$ is sometimes called the *geometric genus* of the nodal surface. See the example in Fig. 3, which has geometric genus 2 but arithmetic genus 4 ($\mathcal{L} = 2$).

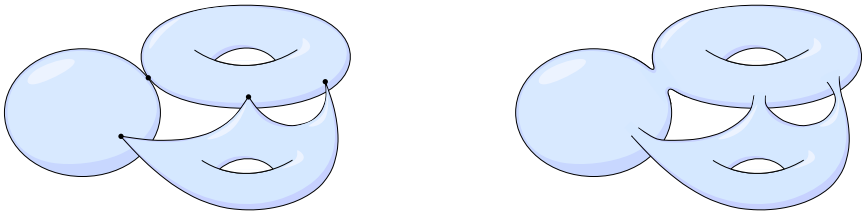

Fig. 3. A nodal surface of arithmetic genus 4 (left), and a surface of genus 4 obtained by "smoothing" the nodes (right).

Nodal topological constellations. Consider a D-uplet $\hat{\boldsymbol{\eta}} = (\hat{\eta}^1, \dots, \hat{\eta}^D)$ of constellations defined on the same set of n elements, where $\hat{\eta}^c = (\eta_1^c, \dots, \eta_{k_c}^c)$ is a k_c-constellation, $k_c \ge 1$. As detailed in Sec. 2.5, a graph embedded in a connected surface X is the drawing of a connected graph on X so that the vertices correspond to distinct points on the surface, the images of the edges are paths that may only intersect at the vertices, and the complement of the graph in X is homeomorphic to a disjoint union of discs. To simplify the discussion below, we say that a disconnected graph with p components is embedded in a surface with p connected components if each connected component of the graph is embedded in a connected component of the surface.

For two connected surfaces X_1, X_2, two embedded graphs $\Gamma_1 \subset X_1$ and $\Gamma_2 \subset X_2$ are said to be *isomorphic* if there exists a homeomorphism $\phi : X_1 \to X_2$ of surfaces whose restriction to Γ_1 is a graph isomorphism between Γ_1 and Γ_2.

We consider each constellation $\hat{\eta}^c$ as an isomorphism class of (not necessarily connected) embedded graphs (for more details, see [34]). For each $1 \le c \le D$, and for every choice of graph embeddings $\Gamma_c \subset X_c$ in the isomorphism class, the white vertices are points on the (not necessarily connected) surface X_c, which we denote by v_i^c, $1 \le i \le n$, and setting $P_j = \{v_j^1, \dots, v_j^D\}$, we consider the nodal surface $\bigsqcup_{c=1}^D X_c/\{P_j\}_{1\le j\le n}$, together with the graph Γ_c embedded in each surface X_c. Two such objects, called here

nodal embedded graphs, are said to be isomorphic if there exists a homeomorphism between the nodal surfaces as defined above such that the restriction to each domain irreducible component is an isomorphism between embedded graphs.

We define a *nodal topological constellation* to be a resulting isomorphism class of nodal embedded graphs. It is uniquely encoded by an ordered multiplet of constellations on the same n elements.

Example: the nodal topological constellation $\mathcal{S}(\boldsymbol{\sigma}, \boldsymbol{\tau})$. We may for instance view $(\boldsymbol{\sigma}, \boldsymbol{\tau}^{-1})$ as a nodal topological constellation, which we denote by $\mathcal{S}(\boldsymbol{\sigma}, \boldsymbol{\tau})$, where the role of $\hat{\eta}^c$ is played by the 2-constellation (σ_c, τ_c^{-1}). We represent for each $i \in \{1, \dots, n\}$ the identification of the white vertices labeled i by introducing a new *triangular* vertex, linked by dotted edges to the white vertices labeled i in each of the D bipartite maps (σ_c, τ_c^{-1}). This is illustrated in Fig. 4.

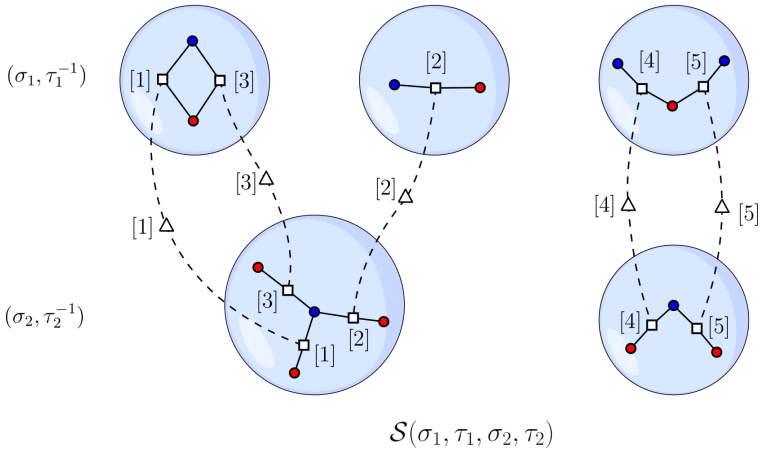

Fig. 4. Graphical representation of $\mathcal{S}(\boldsymbol{\sigma}, \boldsymbol{\tau})$ for an example in $D = 2$, $n = 5$, where $\sigma_1 = (13)(2)(4)(5)$, $\tau_1^{-1} = (13)(2)(45)$, $\sigma_1\tau_1^{-1} = (1)(2)(3)(45)$, $\sigma_2 = (123)(45)$, $\tau_2^{-1} = \mathrm{id}_5$, $\sigma_2\tau_2^{-1} = (123)(45)$. The blue vertices (flavor 1) represent the σs and the red vertices (flavor 2) represent the τs. Here we have represented the bipartite maps as 2-constellations and added a D-valent triangular node for every $i \in \{1, \dots, n\}$, between the corresponding white vertices (edges of the bipartite map). In this example, $|\Pi(\boldsymbol{\sigma}, \boldsymbol{\tau})| = 2$, and the arithmetic genus is 2.

Isomorphisms and relabeling. Two (topological) k-constellations $(\eta_1, \dots, \eta_k)$ and $(\rho_1, \dots, \rho_k)$ are said to be isomorphic if they differ by a relabeling of $1, \dots, n$, that is, if there exists $\nu \in S_n$ such that $\eta_i = \nu\rho_i\nu^{-1}$ for all i. Two nodal topological constellations encoded respectively by $\hat{\boldsymbol{\eta}}$ and $\hat{\boldsymbol{\rho}}$ where $\hat{\eta}^c = (\eta_1^c, \dots, \eta_{k_c}^c)$ and $\hat{\rho}^c = (\rho_1^c, \dots, \rho_{k_c}^c)$ are k_c-constellations are said to be isomorphic if there exists $\nu \in S_n$ such that for all $1 \le c \le D$ and all $1 \le i \le k_c$, $\eta_i^c = \nu\rho_i^c\nu^{-1}$. Note that ν must be the same for all colors: an isomorphism between nodal constellations is a *simultaneous* relabeling of $1, \dots, n$ for all colors.

Transitivity and connectivity. The introduction of nodal surfaces and nodal topological constellations is motivated by the following lemma:

Lemma 5.1. *For $c \in \{1, \dots, D\}$, let $\hat{\eta}^c = (\eta_1^c, \dots, \eta_{k_c}^c)$ be a k_c-constellation, $k_c \geq 1$. The number $|\Pi(\hat{\boldsymbol{\eta}})|$ of transitivity classes of the group generated by all $\hat{\eta}_i^c$ on $\{1, \dots, n\}$ is the number of connected components of the corresponding nodal topological constellation.*

For instance, $\mathcal{S}(\boldsymbol{\sigma}, \boldsymbol{\tau})$ is connected if and only if the group generated by all σ_c, τ_c acts transitively on $\{1, \dots, n\}$, and more generally, the number of connected components of this nodal topological constellation is $|\Pi(\boldsymbol{\sigma}, \boldsymbol{\tau})|$.

Proof of Lemma 5.1. Consider a representative in the isomorphism class of nodal embedded graphs, that is, a nodal surface $X^\bullet = \bigsqcup_c X_c / \{P_j\}_{1 \leq j \leq n}$ together with the graph Γ_c corresponding to $\hat{\eta}^c$ embedded in the surface X_c for every c. From the definition of an embedded graph, we know that X_c has $|\Pi(\hat{\eta}^c)|$ connected components. It is therefore enough to show that two elements $a, b \in \{1, \dots, n\}$ are in the same transitivity class of the group generated by $\{\eta_1^c, \dots, \eta_{k_c}^c\}_{1 \leq c \leq D}$ if and only if there exists a path between the corresponding nodes on the graph $\Gamma^\bullet$ obtained from $\Gamma_1, \dots, \Gamma_D$ by identifying the D white vertices of flavor k for each $k \in \{1, \dots, n\}$.

Two elements $a, b \in \{1, \dots, n\}$ are in the same transitivity class of the group generated by $\{\eta_1^c, \dots, \eta_{k_c}^c\}_{1 \leq c \leq D}$ if and only if there exists a word w in these permutations and their inverses so that $w(a) = b$. Assuming that this is the case, we may build a path between the points corresponding to the two nodes labeled a and b in $\{1, \dots, n\}$ in $\Gamma^\bullet$ as follows: we read the word w from right to left; when encountering a permutation $\eta_i^c(d)$, $1 \leq d \leq n$, the path follows the two edges of flavor i from the node labeled d to the node labeled $\eta_i^c(d)$ on the embedded graph $\Gamma_c \subset X_c$, and similarly for $\eta_i^c(d)^{-1}$. *The important point is that there is no problem in successively applying permutations of different colors, since the path may go between any two $\Gamma_c \subset X_c$ and $\Gamma_{c'} \subset X_{c'}$ at any node.* Conversely, a path in $\Gamma^\bullet$ from a node a to a node b is composed of successive steps from a node to a vertex of flavor i via an edge e and on to another node j via an edge e' in some Γ_c for some c. To each such step we associate the permutation $(\eta_i^c)^{(d+1)}$, d being the number of edges encountered when turning from e to e' around the flavored vertex clockwise. A word w such that $w(a) = b$ is then obtained by composing these permutations from right to left. ∎

Now, as the number of connected components of the nodal topological constellation is $|\Pi(\hat{\boldsymbol{\eta}})|$, the number of its irreducible components is $\sum_c |\Pi(\hat{\eta}^c)|$, its geometric genus is $\sum_c g(\hat{\eta}^c)$, and the surface has n nodal points each with cardinal D, we obtain the arithmetic genus (5.1) of the nodal topological constellation:

$$\begin{aligned}\mathcal{G}(\hat{\boldsymbol{\eta}}) &= \sum_{c=1}^{D} g(\hat{\eta}^c) + Dn - \Big(\sum_{c=1}^{D} |\Pi(\hat{\eta}^c)| + n - |\Pi(\hat{\boldsymbol{\eta}})|\Big) \\ &= \sum_{c=1}^{D} (g(\hat{\eta}^c) - |\Pi(\hat{\eta}^c)|) + n(D-1) + |\Pi(\hat{\boldsymbol{\eta}})|. \end{aligned} \tag{5.2}$$

For instance, combining this formula for $\mathcal{S}(\boldsymbol{\sigma},\boldsymbol{\tau})$ with (2.11) for the Euler characteristics of (σ_c,τ_c^{-1}) we get

$$2g(\sigma_c,\tau_c^{-1})-2|\Pi(\sigma_c,\tau_c^{-1})| = \|\sigma_c\tau_c^{-1}\|-\#(\sigma_c)-\#(\tau_c). \tag{5.3}$$

Supplementing this with the definition (4.2), $\ell(\boldsymbol{\sigma},\boldsymbol{\tau})=\sum_{c=1}^{D}\|\sigma_c\tau_c^{-1}\|+2(|\Pi(\boldsymbol{\sigma},\boldsymbol{\tau})|-1)$, leads to the following lemma.

Lemma 5.2. *The arithmetic genus $\mathcal{G}(\boldsymbol{\sigma},\boldsymbol{\tau})$ of $\mathcal{S}(\boldsymbol{\sigma},\boldsymbol{\tau})$ is related to $\ell(\boldsymbol{\sigma},\boldsymbol{\tau})$ by*

$$\ell(\boldsymbol{\sigma},\boldsymbol{\tau})=\sum_{c=1}^{D}[\#(\sigma_c)+\#(\tau_c)]+2\mathcal{G}(\boldsymbol{\sigma},\boldsymbol{\tau})-2-2n(D-1). \tag{5.4}$$

5.2. *Transitive factorizations of multiplets of permutations and branched coverings of a bouquet of 2-spheres*

Given two topological spaces X and Y, and a subset L of Y, a map $f:X\to Y$ is said to be an *n-sheeted branched covering of Y branched over L* if f restricted to the complement of the preimage of L in X is continuous, and for every $y\in Y\setminus L$ there exists an open neighborhood U such that $f^{-1}(U)$ is homeomorphic to $U\times\{1,\dots,n\}$. Two branched coverings $f_1:X_1\to Y$ and $f_2:X_2\to Y$ are said to be *isomorphic* if there exists an orientation preserving homeomorphism $u:X_1\to X_2$ such that $f_1=f_2\circ u$. The set L is called the *branch locus*, Y the *target space*, and X the *covering space*. The number of connected components of a covering is that of the covering space. For X,Y two nodal surfaces, L consists of points called *branch points*, and their preimages are called *singular points*.

It is well known (see [34]) that n-sheeted coverings of the oriented 2-sphere branched over k ordered points are up to isomorphism in one-to-one correspondence with k-constellations $\hat{\eta}=(\eta_1,\dots,\eta_k)$, that is, k-uplets of permutations of n elements such that $\eta_1\cdots\eta_k=\mathrm{id}$, up to isomorphism. Given such an unlabeled constellation, an isomorphism class of branched coverings is obtained by sending each face of the corresponding topological constellation (for every surface in the isomorphism class) to the face of the unique constellation with one white vertex. Each "star" in the constellation formed by a white vertex and its incident edges thus corresponds to the preimage of the only "star" in the target space. The vertices with flavors of the constellation correspond to the singular points, and the partitions of n that label the conjugacy classes of the permutations $\eta_1,\dots,\eta_k$, called *ramification profiles*, describe the way in which the n sheets meet in groups at the singular points. The covering space is a collection of $K=|\Pi(\hat{\eta})|$ connected surfaces seen up to isomorphism, whose genera sum up to $h=g(\hat{\eta})$. The *Riemann–Hurwitz formula* relates these two numbers:

$$\sum_{i=1}^{k}\|\eta_i\|=2(n+h-K), \tag{5.5}$$

where for the branch point labeled i, $\|\eta_i\| = n - \#(\eta_i)$ is the difference between the number n of preimages that the point would have if it was not in the branch locus, and the number of preimages it actually has.

A D-uplet $\hat{\boldsymbol{\eta}} = (\hat{\eta}^1, \dots, \hat{\eta}^D)$ of constellations, where $\hat{\eta}^c = (\eta_1^c, \dots, \eta_{k_c}^c)$ is a k_c-constellation on n elements, $k_c \geq 1$, up to isomorphism is therefore in bijection with D branched coverings $f_1, \dots, f_D$ of the 2-sphere S, up to isomorphism, $f_c : X_c \to S$ being branched over k_c points. Unlike for nodal constellations, here the isomorphisms are for each color independently, that is, independent relabelings of $1, \dots, n$ for different colors are allowed. There is no direct interpretation in this context for the quantity $|\Pi(\hat{\boldsymbol{\eta}})|$, which moreover is not invariant under relabelings of $\{1, \dots, n\}$ for each color independently; it is only invariant under simultaneous relabelings for all colors. On the other hand, $|\Pi(\hat{\boldsymbol{\eta}})|$ has a natural interpretation in the context of nodal constellations, as stated in the following theorem.

Theorem 5.3. *Isomorphism classes of connected branched coverings of a bouquet of D distinguishable 2-spheres S_c, $c \in \{1, \dots, D\}$, branched over a set of $k + 2D$ ordered points that do not belong to the nodal point, $k_c + 2$ of which belong to S_c for each c ($\sum_{c=1}^D k_c = k$), are in one-to-one correspondence with systems of permutations of the type:*

- *for $c \in \{1, \dots, D\}$, $\eta_0^c, \dots, \eta_{k_c+1}^c \in S_n$ such that* $\mathrm{id} = \eta_0^c \cdots \eta_{k_c+1}^c$,
- $|\Pi(\hat{\boldsymbol{\eta}})| = 1$, *that is, the group generated by all the permutations is transitive on $\{1, \dots, n\}$,*

up to isomorphism of nodal constellations (*up to simultaneous relabeling of* $1, \dots, n$ *for all* $1 \leq c \leq D$).

Proof. We prove the correspondence between topological objects, knowing the correspondence between nodal topological constellations and systems of permutations. Consider a branched covering $f : X^\bullet \to Y^\bullet$ where $Y^\bullet$ is a bouquet of D distinguishable 2-spheres S_c, $c \in \{1, \dots, D\}$. On each 2-sphere S_c of the target space $Y^\bullet$, one can draw a star-graph γ_c by adding non-crossing arcs between the k_c branch points and the nodal point so that the arcs around the nodal point are numbered from 1 to k_c clockwise (see the right of Fig. 5). Doing this for all c, we get a nodal embedded graph $\gamma^\bullet \subset Y^\bullet$, whose preimage $\Gamma^\bullet \subset X^\bullet$ is a representative of a nodal topological constellation in the sense that it is a representative in the corresponding isomorphism class of nodal embedded graphs. There is no labeling of the n preimages of the nodal point, so that the nodal constellation can be seen up to isomorphism (up to simultaneous relabeling of $1, \dots, n$ for all $1 \leq c \leq D$).

For two isomorphic branched coverings $f_1 : X_1^\bullet \to Y^\bullet$ and $f_2 : X_2^\bullet \to Y^\bullet$, there exists by definition an orientation preserving homeomorphism $u : X_1^\bullet \to X_2^\bullet$ of nodal surfaces such that $f_1 = f_2 \circ u$. Denoting by $\Gamma_1^\bullet \subset X_1^\bullet$ and $\Gamma_2^\bullet \subset X_2^\bullet$ the preimages of $\gamma^\bullet \subset Y^\bullet$ by f_1 and f_2 respectively, it is clear that u restricted to each irreducible component of $X_1^\bullet$ is an isomorphism between embedded graphs: it is a homeomorphism of surfaces by definition, and it is clear that the restriction of u to $\Gamma_1^\bullet$ on each irreducible component is

a graph isomorphism. Therefore, the nodal embedded graphs $\Gamma_1^\bullet \subset X_1^\bullet$ and $\Gamma_2^\bullet \subset X_2^\bullet$ are isomorphic, and are two representatives of the unlabeled nodal topological constellation.

This defines a map from isomorphism classes of branched coverings of $Y^\bullet$ to isomorphism classes of nodal topological constellations, and we now verify that this map is invertible. Indeed, consider a representative $\Gamma^\bullet \subset X^\bullet$ of a nodal topological constellation $\hat{\boldsymbol{\eta}} = (\hat{\eta}^1, \dots, \hat{\eta}^D)$, where $\hat{\eta}^c = (\eta_1^c, \dots, \eta_{k_c}^c)$ is a k_c-constellation. Then $\Gamma^\bullet \subset X^\bullet$ is a nodal embedded graph, and for every c, we denote by X_c the disjoint union of the irreducible components of $X^\bullet$ that contain vertices associated with η_0^c. A branched covering $f : X^\bullet \to Y^\bullet$ is then obtained by choosing homeomorphisms sending each connected component of the complement of the graph $\Gamma^\bullet$ in X_c to the complement of the star-graph γ_c in the irreducible component S_c of $Y^\bullet$.

Given two representatives $\Gamma_1^\bullet \subset X_1^\bullet$ and $\Gamma_2^\bullet \subset X_2^\bullet$ of a nodal topological constellation, there exists a homeomorphism $u : X_1^\bullet \to X_2^\bullet$ of nodal surfaces that induces an isomorphism of embedded graphs on every irreducible component of $X_1^\bullet$. Considering the branched coverings $f_1 : X_1^\bullet \to Y^\bullet$ and $f_2 : X_2^\bullet \to Y^\bullet$ constructed as in the previous paragraph, we see that $f_1 = f_2 \circ u$ so that f_1 and f_2 are isomorphic.

The construction described above that associates a covering f to a representative $\Gamma^\bullet \subset X^\bullet$ is independent of the labeling of the nodal points of $\Gamma^\bullet$, so that we have defined the converse map from isomorphism classes of nodal topological constellations to isomorphism classes of branched coverings of $Y^\bullet$.

The statement regarding the number of connected components is a direct consequence of Lemma 5.1. ∎

We illustrate this with the following example with $n = 5$, $D = 2$, $k_1 = 2$, $k_2 = 3$, for which the nodal topological constellation is represented graphically on the left of Fig. 5: for the permutations of $\hat{\eta}^1$: $\eta_0^1 = (12)(3)(4)(5)$, $\eta_1^1 = (12)(34)(5)$, $\eta_2^1 = (12)(345)$, $\eta_1^3 = (12)(3)(45)$; for the permutations of $\hat{\eta}^2$: $\eta_0^2 = (132)(45)$, $\eta_1^2 = (15)(24)(3)$, $\eta_2^2 = (1)(23)(4)(5)$, $\eta_3^2 = (14)(2)(35)$, $\eta_2^4 = \mathrm{id}_5$. In the figure, the colors representing the flavors

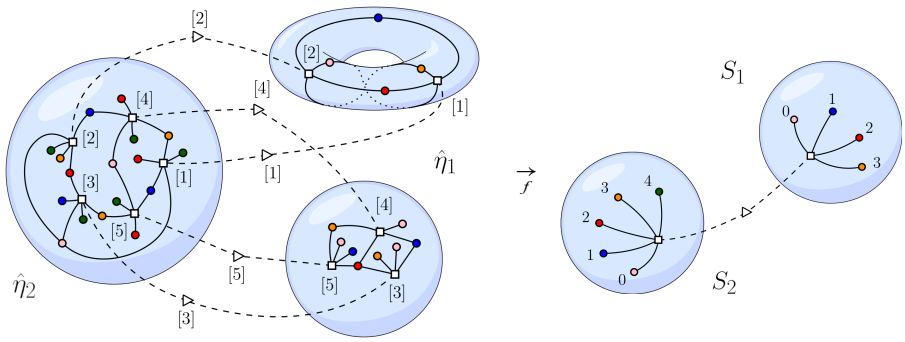

Fig. 5. A nodal topological constellation (left) can be interpreted, up to relabeling the nodes, as an isomorphism class of branched coverings f of a bouquet of D distinguishable 2-spheres (right). In this example, $D = 2$, $n = 5$, $k_1 = 2$, and $k_2 = 3$.

0,1,2,3,4 are in that order pink, blue, red, orange, green. The nodal topological constellation is connected and has arithmetic genus 4. The fact that $\eta_2^4 = \mathrm{id}_5$ (all green vertices are leaves in the nodal constellation) means that when interpreted as a branched covering of a bouquet of two 2-spheres, the green vertex in the target space is actually not a branch point since it has five preimages.

Fixing $D, k \geq 1$, $\mathcal{H} \geq 0$ and for all $c \in \{1, \dots, D\}$, $\alpha_c, \beta_c \vdash n$ non-trivial, we let $\mathfrak{C}_{\mathcal{H}}[\{\alpha_c, \beta_c\}_c, k]$ be the set of isomorphism classes of connected n-sheeted branched coverings of a bouquet of D distinguishable 2-spheres S_c, $c \in \{1, \dots, D\}$, branched over a set of precisely[8] $k + 2D$ ordered points that do not belong to the nodal point, at least two of which belong to S_c for each c, so that the first and last points for each c have ramification profiles α_c and β_c,[9] and so that the arithmetic genus of the covering space as defined in (5.1) is $\mathcal{H}$.

Note that for an element X of $\mathfrak{C}_{\mathcal{H}}[\{\alpha_c, \beta_c\}_c, k]$, the nodal surfaces in the isomorphism class have n nodal points (the node of the bouquet of spheres does not belong to the branch locus and its n preimages are the only nodes of X), so that $\mathcal{L}(X) = n(D-1) - p + 1$ in (5.1).

We let $\mathfrak{B}_{\mathcal{H}}[\{\alpha_c, \beta_c\}_c, k]$ be the subset of $\mathfrak{C}_{\mathcal{H}}[\{\alpha_c, \beta_c\}_c, k]$ of the elements X for which the branch points whose ramification profiles are not fixed to be one of the α_c or β_c have simple ramification (they have $n-1$ preimages) and satisfy an additional monotonicity condition: Consider any set of permutations encoding X (a choice of labeling of $1, \dots, n$ in Thm. 5.3). For $c \in \{1, \dots, D\}$, the transpositions encoding the ramification profiles of the branch points in S_c whose ramification profiles are not fixed to be α_c or β_c inherit an ordering from the global ordering of the branch points. With this ordering, these transpositions must have weakly monotone maxima.

We recall that for $\alpha \vdash n$, the conjugacy class C_α gathers the permutations in S_n whose cycle type is α, that m_C and p_C were respectively defined in Thm. 3.3 and Prop. 3.6, as well as the definition of higher order monotone double Hurwitz numbers (4.24):

$$\vec{H}^l(\{\alpha_c, \beta_c\}_c) = \sum_{c=1}^{D} \sum_{(\sigma_c, \tau_c) \in C_{\alpha_c} \times C_{\beta_c}} p_C[\boldsymbol{\sigma}, \boldsymbol{\tau}; l]. \tag{5.6}$$

We also define the following generalization of the Bousquet-Mélou–Schaeffer numbers [5]:

$$\vec{B}S_k^l(\{\alpha_c, \beta_c\}_c) = \sum_{c=1}^{D} \sum_{(\sigma_c, \tau_c) \in C_{\alpha_c} \times C_{\beta_c}} m_C(\boldsymbol{\sigma}, \boldsymbol{\tau}; l, k). \tag{5.7}$$

[8] That is, each of these points has less than n preimages. Since α_c, β_c are non-trivial and since the other permutations involved in the definitions of m_C and p_C are different from the identity permutation (proper), all the $k + 2D$ points have non-trivial ramifications.

[9] For the Riemann sphere, these points are usually taken to be zero and infinity.

Corollary 5.4. *Fix $D, k \geq 1$, $\mathcal{H} \geq 0$, and for all $c \in \{1, \dots, D\}$, $\alpha_c, \beta_c \vdash n$ non-trivial and $\sigma_c \in C_{\alpha_c}$, $\tau_c^{-1} \in C_{\beta_c}$, and define*

$$l = \sum_c [\#(\alpha_c) + \#(\beta_c)] + 2\mathcal{H} - 2 - 2n(D-1). \tag{5.8}$$

(1) *The cardinality of $\mathfrak{C}_{\mathcal{H}}[\{\alpha_c, \beta_c\}_c, k]$ is $\frac{1}{n!}\vec{B}S_k^l(\{\alpha_c, \beta_c\}_c)$.*

(2) *The cardinality of $\mathfrak{B}_{\mathcal{H}}[\{\alpha_c, \beta_c\}_c, k]$ is $\frac{1}{n!}\vec{H}^k(\{\alpha_c, \beta_c\}_c)$ and $k = l$ as defined in (5.8).*

In both cases, the total number of singular points is

$$nk - l + \sum_c \#(\alpha_c) + \#(\beta_c) = 2 - 2\mathcal{H} + n(k + D - 1). \tag{5.9}$$

The relation (5.8) should be compared with the Riemann–Hurwitz formula (4.16).

Proof of Corollary 5.4. Let $f : X^\bullet \to Y^\bullet$ be an element of $\mathfrak{C}_{\mathcal{H}}[\{\alpha_c, \beta_c\}_c, k]$, where $Y^\bullet$ is a bouquet of D distinguishable 2-spheres S_c, $c \in \{1, \dots, D\}$. Then there exist $k_1, \dots, k_D \geq 0$ such that $\sum_c k_c = k$ and for each c, $k_c + 2$ of the (ordered) branch points belong to S_c, and the first and last respectively have ramification profiles α_c and β_c. From Thm. 5.3 and its proof, f is bijectively mapped to D ordered sequences of permutations

$$\hat{\eta}^c = (\sigma_c^{-1}, \eta_1^c, \dots, \eta_{k_c}^c, \tau_c) \in S_n^{k_c+2} \quad \text{with} \quad \mathrm{id} = \sigma_c^{-1}\eta_1^c \cdots \eta_{k_c}^c \tau_c,$$

for $c \in \{1, \dots, D\}$, such that $\sigma_c \in C_{\alpha_c}$ and $\tau_c \in C_{\beta_c}$, and the group generated by all the permutations is transitive on $\{1, \dots, n\}$, up to simultaneous relabelings of $1, \dots, n$ for all c, and so that the arithmetic genus of the nodal topological constellation encoded by this system of permutations is $\mathcal{H}$. This explains that $\mathrm{Card}\,\mathfrak{C}_{\mathcal{H}}[\{\alpha_c, \beta_c\}_c, k] = \frac{1}{n!}\vec{B}S_k^l(\{\alpha_c, \beta_c\}_c)$ if we show that $\sum_{c=1}^D \sum_{i=1}^{k_c} \|\eta_i^c\|$ is given by the right hand side of (5.8).

From (5.2), $\mathcal{H} = \sum_{c=1}^D (g(\hat{\eta}^c) - |\Pi(\hat{\eta}^c)|) + n(D-1) + 1$, and from the Euler characteristic (2.11) of $\hat{\eta}^c$,

$$2(g(\hat{\eta}^c) - |\Pi(\hat{\eta}^c)|) = \sum_{i=1}^{k_c} \|\eta_i^c\| - \#(\sigma_c) - \#(\tau_c),$$

so that

$$\sum_{c=1}^D \sum_{i=1}^{k_c} \|\eta_i^c\| = \sum_{c=1}^D [\#(\sigma_c) + \#(\tau_c)] + 2(\mathcal{H} - n(D-1) - 1),$$

which proves the first point of the corollary. For the elements of $\mathfrak{B}_{\mathcal{H}}[\{\alpha_c, \beta_c\}_c, k]$, $\eta_1^c, \dots, \eta_{k_c}^c$ are transpositions with weakly monotone maxima, and the total number k of these transpositions is also $\sum_{c=1}^D \sum_{i=1}^{k_c} \|\eta_i^c\|$. This concludes the proof. ∎

Remark 5.5. For $D = 2$, the number of isomorphism classes of branched covers of a bouquet of two 2-spheres should be relevant in the context of compactifications of the moduli spaces of curves such as the Deligne–Mumford compactification, where the necessity to include degenerate cycles implies considering nodal surfaces where at each node only two surfaces meet [11, 22, 35, 47].

Remark 5.6. We have presented a geometrical interpretation based on nodal surfaces. From the colored structure, the reader familiar with the literature on colored triangulations and random tensor models will recognize a combinatorial encoding that resembles that of colored triangulations in dimension 2 and higher. This begs for an interpretation in terms of higher-dimensional objects, instead of nodal surfaces, but we leave this for future work.

5.3. The $1/N$ expansions as topological expansions

The aim of this subsection is to provide a combinatorial and geometric interpretation to the formulas of Theorem 4.1. The transpositions μ_i^c in the combinatorial definition of p_C (Prop. 3.6) do not appear for instance in (4.6): the intuition is that we should try to keep all the σ_c, τ_c fixed on one the hand, and "resum" the contributions of all the μ_i^c on the other hand, in some way. To this end, given D sequences of permutations $\sigma_c, \tau_c, \eta_1^c, \dots, \eta_{k_c}^c$ for $c \in \{1, \dots, D\}$ such that $\mathrm{id} = \sigma_c^{-1}\eta_1^c \cdots \eta_{k_c}^c \tau_c$, instead of considering the nodal topological constellation encoded by the $(\sigma_c^{-1}, \eta_1^c, \dots, \eta_{k_c}^c, \tau_c)$ for all c as in Corollary 5.4, we will rather consider a new kind of isomorphism class of nodal surfaces from the nodal topological constellation $\mathcal{S}(\boldsymbol{\sigma}, \boldsymbol{\tau})$ on the one hand, and the D topological constellations $\hat{\eta}^c = (\eta_1^c, \dots, \eta_{k_c}^c)$ on the other hand.

5.3.1. Nodal surfaces for $(\boldsymbol{\sigma}, \boldsymbol{\tau}^{-1}, \hat{\boldsymbol{\eta}})$. We fix $\boldsymbol{\sigma}, \boldsymbol{\tau} \in \boldsymbol{S}_n$ as well as $k, l \geq 0$, and for $c \in \{1, \dots, D\}$, we let $\hat{\eta}^c = (\eta_1^c, \dots, \eta_{k_c}^c)$ be a k_c-constellation, $k_c \geq 0$, such that $\sum_{c=1}^D k_c = k$ and

$$\forall c, \quad \sigma_c \tau_c^{-1} = \eta_1^c \cdots \eta_{k_c}^c, \tag{5.10}$$

subject to the conditions

(C1) the collection of all $\{\hat{\eta}^c, \tau_c\}_c$ acts transitively on $\{1, \dots, n\}$,

(C2) $\sum_{c=1}^D \sum_{i_c=1}^{k_c} \|\eta_{i_c}^c\| = l$.

This data defines

- a (not necessarily connected) nodal topological constellation $\mathcal{S}(\boldsymbol{\sigma}, \boldsymbol{\tau})$ as defined in Sec. 5.1 (see Fig. 4),
- a (not necessarily connected) topological k_c-constellation $\hat{\eta}^c$ for each $c \in \{1, \dots, D\}$.

Since $\nu_c = \sigma_c \tau_c^{-1} = \eta_1^c \cdots \eta_{k_c}^c$, the disjoint cycles of $\sigma_c \tau_c^{-1}$ and the disjoint cycles of $\eta_1^c \cdots \eta_{k_c}^c$ match, so that for every nodal embedded graph $\Gamma^\bullet \subset X^\bullet$ in the isomorphism class $\mathcal{S}(\boldsymbol{\sigma}, \boldsymbol{\tau})$ and any embedded graphs $\Gamma_1 \subset Y_1, \dots, \Gamma_D \subset Y_D$ in the isomorphism classes $\hat{\eta}^1, \dots, \hat{\eta}^D$, there is a one-to-one correspondence Ψ between the faces $F_1, \dots, F_{nk-l}$ of $\Gamma^\bullet \subset X^\bullet$ (the connected components of the complement of the graph $\Gamma^\bullet$ in the nodal surface $X^\bullet$) and the faces $F'_1, \dots, F'_{nk-l}$ of the $\Gamma_c \subset Y_c$ for $c \in \{1, \dots, D\}$, where the labelings are chosen so that $\Psi(F_j) = F'_j$. To render this pairwise identification obvious, we choose for each $j \in \{1, \dots, nk - l\}$ two points v_j and v'_j in the interiors of F_j and F'_j respectively, and we consider the nodal surface $Z^\bullet = (X^\bullet \sqcup \bigsqcup_{c=1}^D Y_c)/\{P_j\}_j$ (together with the graphs $\Gamma^\bullet$ and Γ_c drawn on $Z^\bullet$).

We then denote by $\mathcal{S}(\boldsymbol{\sigma}, \boldsymbol{\tau}, \hat{\boldsymbol{\eta}})$ the isomorphism class of such objects, where by isomorphism we mean homeomorphism of nodal surfaces that induces an isomorphism of the embedded graph on each irreducible component, and preserves the incidence between the nodes P_j and the faces F_j and F'_j, in the sense that if a node P_j belongs to the interior of the faces F_j and F'_j, then the image of the node also belongs to the interior of the images of the faces.

An example is shown in Fig. 6, where the nodal points in the interior of the faces are represented by dotted edges (whereas we recall that the nodes of $\mathcal{S}(\boldsymbol{\sigma}, \boldsymbol{\tau})$ are represented by dashed edges linking triangular vertices).

In this context, the graph $G[\Pi(\boldsymbol{\sigma}, \boldsymbol{\tau}), \{\pi_c\}_c; \{\Pi(\nu_c)\}_c]$ for $\pi_c = \Pi(\hat{\eta}^c)$ and $\nu_c = \sigma_c \tau_c^{-1}$ introduced in Sec. 4.2 is simply obtained by *contracting* the connected components of the nodal surface $\mathcal{S}(\boldsymbol{\sigma}, \boldsymbol{\tau})$ (not its irreducible components!) and those of each constellation $\hat{\eta}^c$ to points. This retains the information on which faces F_j and F'_j identified by Ψ are in the same connected component of $\mathcal{S}(\boldsymbol{\sigma}, \boldsymbol{\tau})$ on the one hand, and which are in the same connected component of $\hat{\eta}^c$ on the other hand (this is why it only depends on the associated *partitions*). The graph $G[\Pi(\boldsymbol{\sigma}, \boldsymbol{\tau}), \{\pi_c\}_c; \{\Pi(\nu_c)\}_c]$ for the example in Fig. 6 is the one in Fig. 2.

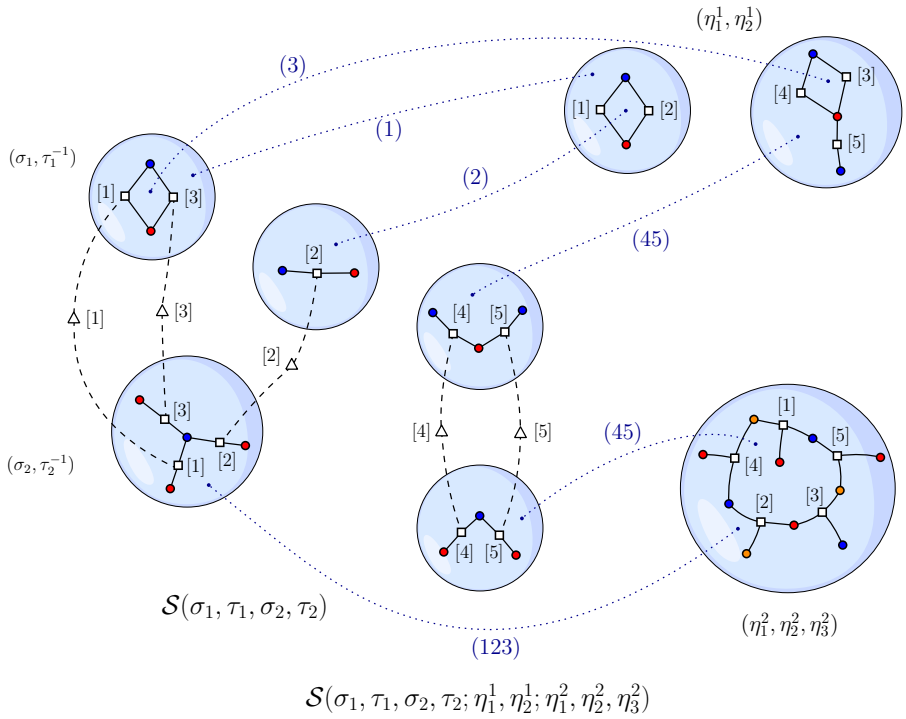

Fig. 6. Graphical representation of $\mathcal{S}(\boldsymbol{\sigma}, \boldsymbol{\tau}; \hat{\boldsymbol{\eta}})$ for an example in $D = 2$, $n = 5$, where $\mathcal{S}(\boldsymbol{\sigma}, \boldsymbol{\tau})$ is as in Fig. 4, and $\eta_1^1 = (12)(34)(5)$, $\eta_2^1 = (12)(345)$, $\eta_1^2 = (15)(24)(3)$, $\eta_2^2 = (1)(23)(4)(5)$, and $\eta_3^2 = (14)(2)(35)$. The colors for the constellations are blue for 1, red for 2, orange for 3. The dotted lines representing the pairwise identification of faces are labeled with the corresponding cycles.

Lemma 5.7. *The number of connected components of $\mathcal{S}(\boldsymbol{\sigma},\boldsymbol{\tau},\hat{\boldsymbol{\eta}})$ is the number of connected components of $G[\Pi(\boldsymbol{\sigma},\boldsymbol{\tau}),\{\Pi(\hat{\eta}_c)\}_c;\{\Pi(\nu_c)\}_c]$, which is also the number of transitivity classes of the group generated by all $\boldsymbol{\sigma},\boldsymbol{\tau},\hat{\boldsymbol{\eta}}$, namely $|\Pi(\boldsymbol{\sigma},\boldsymbol{\tau},\hat{\boldsymbol{\eta}})|$. Imposing transitivity in* (C1) *is imposing that $\mathcal{S}(\boldsymbol{\sigma},\boldsymbol{\tau},\hat{\boldsymbol{\eta}})$ is connected.*

Lemma 5.8. *With the notations* (5.10)*, l is related to the arithmetic genus $\mathcal{G}(\boldsymbol{\sigma},\boldsymbol{\tau},\hat{\boldsymbol{\eta}})$ of $\mathcal{S}(\boldsymbol{\sigma},\boldsymbol{\tau},\hat{\boldsymbol{\eta}})$ by*

$$l=\sum_{c=1}^{D}\sum_{i=1}^{k_c}\|\eta_i^c\|=\sum_{c=1}^{D}[\#(\sigma_c)+\#(\tau_c)]+2\mathcal{G}(\boldsymbol{\sigma},\boldsymbol{\tau},\hat{\boldsymbol{\eta}})-2-2n(D-1). \tag{5.11}$$

In particular, by (5.4)*,*

$$l-\ell(\boldsymbol{\sigma},\boldsymbol{\tau})=2(\mathcal{G}(\boldsymbol{\sigma},\boldsymbol{\tau},\hat{\boldsymbol{\eta}})-\mathcal{G}(\boldsymbol{\sigma},\boldsymbol{\tau})). \tag{5.12}$$

Fixing l in (C2) *therefore amounts to fixing the arithmetic genus of $\mathcal{S}(\boldsymbol{\sigma},\boldsymbol{\tau},\hat{\boldsymbol{\eta}})$.*

Proof. From the definition (5.1),

$$\begin{aligned}\mathcal{G}(\boldsymbol{\sigma},\boldsymbol{\tau},\hat{\boldsymbol{\eta}})=&\sum_{c=1}^{D}\big(g(\sigma_c,\tau_c^{-1})+g(\hat{\eta}^c)\big)+nD+\sum_{c=1}^{D}\#(\sigma_c\tau_c^{-1})-n\\&-\sum_{c=1}^{D}\big(|\Pi(\sigma_c,\tau_c^{-1})|+|\Pi(\hat{\eta}^c)|\big)+1.\end{aligned}$$

Summing the Euler characteristics (2.10) of $\hat{\eta}^c$ yields

$$\sum_{c=1}^{D}(g(\hat{\eta}^c)-|\Pi(\hat{\eta}^c)|)=\frac{1}{2}(l-nD)-\frac{1}{2}\sum_{c=1}^{D}\#(\sigma_c\tau_c^{-1}).$$

The result follows using the Euler characteristics of (σ_c,τ_c^{-1}) (5.3). ∎

From these lemmas (for item (1) of the proposition) as well as the results of Sec. 5.2 (for item (2)) we get

Proposition 5.9. *The $1/N$ expansion of the cumulant Weingarten functions in Thm.* 3.3 *can be seen as a topological expansion:*

$$\begin{aligned}W_C^{(N)}[\boldsymbol{\sigma},\boldsymbol{\tau}]=&(-1)^{\ell(\boldsymbol{\sigma},\boldsymbol{\tau})}N^{2\mathcal{G}(\boldsymbol{\sigma},\boldsymbol{\tau})-\ell(\boldsymbol{\sigma},\boldsymbol{\tau})-nD}\\&\cdot\sum_{\mathcal{G}\geq\mathcal{G}(\boldsymbol{\sigma},\boldsymbol{\tau})}\frac{1}{N^{2\mathcal{G}}}p_C[\boldsymbol{\sigma},\boldsymbol{\tau};\ell(\boldsymbol{\sigma},\boldsymbol{\tau})+2(\mathcal{G}-\mathcal{G}(\boldsymbol{\sigma},\boldsymbol{\tau}))],\end{aligned}$$

where

$$(-1)^{\ell(\boldsymbol{\sigma},\boldsymbol{\tau})}p_C[\boldsymbol{\sigma},\boldsymbol{\tau};l]=\sum_{k\geq 0}(-1)^k m_C(\boldsymbol{\sigma},\boldsymbol{\tau};l,k),$$

and $m_C(\boldsymbol{\sigma},\boldsymbol{\tau};\ell(\boldsymbol{\sigma},\boldsymbol{\tau})+2(\mathcal{G}-\mathcal{G}(\boldsymbol{\sigma},\boldsymbol{\tau})),k)$ counts both

(1) *the isomorphism classes of nodal surfaces* $\mathcal{S}(\boldsymbol{\sigma}, \boldsymbol{\tau}, \hat{\boldsymbol{\eta}})$ *with* $\hat{\boldsymbol{\eta}} = (\hat{\eta}^1, \dots, \hat{\eta}^D)$ *as in* (5.10),

(2) *the nodal topological constellations encoded by* $\hat{\boldsymbol{\eta}}' = (\hat{\eta}'^1, \dots, \hat{\eta}'^D)$ *where* $\hat{\eta}'^c = (\sigma_c^{-1}, \hat{\eta}^c, \tau_c)$,

and in both cases, the spaces are ***connected*** *and of* ***fixed arithmetic genus*** $\mathcal{G} \geq \mathcal{G}(\boldsymbol{\sigma}, \boldsymbol{\tau})$, *and so that the vertices of flavor* i *of the* $\hat{\eta}^c$ *are not all of valency* 1. *The number* $p_C[\boldsymbol{\sigma}, \boldsymbol{\tau}; l]$ *counts those spaces listed above for which the* $\hat{\eta}^c$ *consist of sequences of transpositions with weakly monotone maxima* (*this also translates into a condition on the flavored vertices*).

The arithmetic genus $\mathcal{G}(\boldsymbol{\sigma}, \boldsymbol{\tau}, \hat{\boldsymbol{\eta}})$, or equivalently the exponent $l = \sum_{c=1}^{D} \sum_{i_c=1}^{k_c} \|\eta_{i_c}^c\|$, can then be expressed as (Sec. 4.2)

$$\mathcal{G}(\boldsymbol{\sigma}, \boldsymbol{\tau}, \hat{\boldsymbol{\eta}}) - \mathcal{G}(\boldsymbol{\sigma}, \boldsymbol{\tau}) = \frac{l - \ell(\boldsymbol{\sigma}, \boldsymbol{\tau})}{2} = \sum_{c=1}^{D} g(\hat{\eta}^c) + L[\Pi(\boldsymbol{\sigma}, \boldsymbol{\tau}), \{\pi_c\}_c; \{\Pi(\nu_c)\}_c], \tag{5.13}$$

where $g(\hat{\eta}^c)$ is the genus of the k_c-constellation $\hat{\eta}^c$ (the sum of the genera of its connected components) and $L[\Pi(\boldsymbol{\sigma}, \boldsymbol{\tau}), \{\pi_c\}_c; \{\Pi(\nu_c)\}_c]$ is the number of excess edges (4.10) of $G[\Pi(\boldsymbol{\sigma}, \boldsymbol{\tau}), \{\pi_c\}_c; \{\Pi(\nu_c)\}_c]$:

$$L[\Pi(\boldsymbol{\sigma}, \boldsymbol{\tau}), \{\pi_c\}_c; \{\Pi(\nu_c)\}_c] = \sum_c (|\Pi(\nu_c)| - |\pi_c|) - |\Pi(\boldsymbol{\sigma}, \boldsymbol{\tau})| + 1. \tag{5.14}$$

This provides a better understanding of how to characterize and count the contributions to m_C and p_C in Prop. 5.9: counting connected $\mathcal{S}(\boldsymbol{\sigma}, \boldsymbol{\tau}, \hat{\boldsymbol{\eta}})$ of fixed arithmetic genus $\mathcal{G}$ amounts to counting those $\mathcal{S}(\boldsymbol{\sigma}, \boldsymbol{\tau}, \hat{\boldsymbol{\eta}})$ for which the graph $G[\dots]$ is connected and has excess L between 0 and $\mathcal{G} - \mathcal{G}(\boldsymbol{\sigma}, \boldsymbol{\tau})$ while the genera $g(\hat{\eta}^c)$ sum up to $\mathcal{G} - \mathcal{G}(\boldsymbol{\sigma}, \boldsymbol{\tau}) - L$ (the other conditions in Prop. 5.9 must also be satisfied). In the example of Fig. 6, the constellations are planar and $(l - \ell)/2 = 2 = L$. For instance, the following gives a prescription for generating all the spaces that contribute at leading order:

Proposition 5.10. *The spaces* $\mathcal{S}(\boldsymbol{\sigma}, \boldsymbol{\tau}, \hat{\boldsymbol{\eta}})$ *that correspond to the leading term in* N *for the expansion* (4.3) *of the cumulant Weingarten functions* $W_C^{(N)}[\boldsymbol{\sigma}, \boldsymbol{\tau}]$ *are those of minimal arithmetic genus* $\mathcal{G}(\boldsymbol{\sigma}, \boldsymbol{\tau}, \hat{\boldsymbol{\eta}}) = \mathcal{G}(\boldsymbol{\sigma}, \boldsymbol{\tau})$, *that is, such that the* $\hat{\eta}^c$ *are all planar and* $G[\Pi(\boldsymbol{\sigma}, \boldsymbol{\tau}), \{\pi_c\}_c; \{\Pi(\nu_c)\}_c]$ *is a tree.*

From this picture the aim is to keep $\mathcal{S}(\boldsymbol{\sigma}, \boldsymbol{\tau})$ fixed and group the contributions of the different constellations $\hat{\boldsymbol{\eta}}$ that lead to the same values of L and the same genera for the connected components of the $\hat{\eta}^c$. This is achieved in the last subsection.

5.3.2. A simpler kind of nodal surfaces. For both m_C and p_C, one fixes $\boldsymbol{\sigma}, \boldsymbol{\tau}$ but sums over the proper $\hat{\boldsymbol{\eta}}$ (denoted by $\hat{\boldsymbol{\rho}}$ or $\hat{\boldsymbol{\mu}}$) satisfying a number of assumptions. In order to understand this geometrically, one may therefore, from $\mathcal{S}(\boldsymbol{\sigma}, \boldsymbol{\tau}; \hat{\boldsymbol{\rho}})$ introduced in the previous subsection, contract to points the connected components of the constellations $\hat{\eta}^c$

but keep the nodal topological constellation $\mathcal{S}(\boldsymbol{\sigma},\boldsymbol{\tau})$ (that is, go only half-way to build the graph $G[\Pi(\boldsymbol{\sigma},\boldsymbol{\tau}),\{\pi_c\}_c;\{\Pi(\nu_c)\}_c]$). The result is a new kind of object $\mathcal{S}(\boldsymbol{\sigma},\boldsymbol{\tau};\{\pi_c\}_c)$ (since only the information on the connected components of the constellations is retained, they have been replaced in the argument by the corresponding partitions of $\{1,\dots,n\}$). It can be obtained directly from $\mathcal{S}(\boldsymbol{\sigma},\boldsymbol{\tau})$ by adding nodal points between the faces of $\mathcal{S}(\boldsymbol{\sigma},\boldsymbol{\tau})$ corresponding to the blocs of $\{\pi_c\}_c$. Let us introduce this object more formally.

We fix $\boldsymbol{\sigma},\boldsymbol{\tau}\in\boldsymbol{S}_n$ as well as $\mathcal{G}\geq 0$, and for $c\in\{1,\dots,D\}$, we let $\pi_c\geq\Pi(\nu_c)$ be a partition of the disjoint cycles of $\nu_c=\sigma_c\tau_c^{-1}$ subject to the conditions

(C′1) $\Pi(\boldsymbol{\sigma},\boldsymbol{\tau})\vee\pi_1\vee\cdots\vee\pi_D=1_n$,

(C′2) $L[\Pi(\boldsymbol{\sigma},\boldsymbol{\tau}),\{\pi_c\}_c;\{\Pi(\nu_c)\}_c]=\mathcal{G}-\mathcal{G}(\boldsymbol{\sigma},\boldsymbol{\tau})$.

For every nodal embedded graph $\Gamma^\bullet\subset X^\bullet$ in the isomorphism class $\mathcal{S}(\boldsymbol{\sigma},\boldsymbol{\tau})$, $\{\pi_c\}_c$ provides a partition of the faces $F_1,\dots,F_{nk-l}$ of $\Gamma^\bullet\subset X^\bullet$ (the connected components of the complement of the graph $\Gamma^\bullet$ in the nodal surface $X^\bullet$). We may see the blocks in this partition as a new kind of node: we choose for each $j\in\{1,\dots,nk-l\}$ a point v_j in the interior of F_j, and we see each block B_c of π_c for each c as a node $P(B_c)=\{v_j\mid F_j\in B_c\}$. We then consider the nodal surface $X^\bullet/\bigsqcup_c\{P(B_c)\}_{B_c\in\pi_c}$ (together with the nodal graph $\Gamma^\bullet$ embedded in $X^\bullet$).

We then denote by $\mathcal{S}(\boldsymbol{\sigma},\boldsymbol{\tau};\{\pi_c\}_c)$ the isomorphism class of such objects, where by isomorphism we mean homeomorphism of nodal surfaces that induces an isomorphism of the nodal embedded graph on $\Gamma^\bullet\subset X^\bullet$ and that preserves the incidence between the nodes $P(B_c)$ and the faces F_j, in the sense that if a node $P(B_c)$ belongs to the interiors of the faces $F_j\in B_c$, then the image of the node also belongs to the interior of the images of the faces. We let

$$\mathfrak{S}_{\mathcal{G}}(\boldsymbol{\sigma},\boldsymbol{\tau})=\big\{\mathcal{S}(\boldsymbol{\sigma},\boldsymbol{\tau};\{\pi_c\}_c)\mid\{\pi_c\}_c\text{ satisfy (C}'1)\text{ and (C}'2)\big\},\tag{5.15}$$

and more generally, we call the $\mathcal{S}(\boldsymbol{\sigma},\boldsymbol{\tau};\{\pi_c\}_c)$ the *foldings of* $\mathcal{S}(\boldsymbol{\sigma},\boldsymbol{\tau})$.

An example is shown in Fig. 7, where now the nodal points in the interiors of the faces are represented by dotted edges that meet at star-vertices labeled by the blocks of the partitions π_c.

As for Lemma 5.7, the following follows directly from the fact that the graph $G[\Pi(\boldsymbol{\sigma},\boldsymbol{\tau}),\{\pi_c\}_c;\{\Pi(\nu_c)\}_c]$ is obtained by contracting the connected components of $\mathcal{S}(\boldsymbol{\sigma},\boldsymbol{\tau})$ to points:

Lemma 5.11. *The number of connected components of* $\mathcal{S}(\boldsymbol{\sigma},\boldsymbol{\tau};\{\pi_c\}_c)$ *is*

$$|\Pi(\boldsymbol{\sigma},\boldsymbol{\tau})\vee\pi_1\vee\cdots\vee\pi_D|.$$

However now, in comparison to Lemma 5.8, the information on the genera of the connected components of the $\hat{\eta}^c$ has been lost:

Lemma 5.12. *The arithmetic genus of a connected folding* $\mathcal{S}(\boldsymbol{\sigma},\boldsymbol{\tau};\{\pi_c\}_c)$ *can be expressed as*

$$\mathcal{G}(\boldsymbol{\sigma},\boldsymbol{\tau};\{\pi_c\}_c)=\mathcal{G}(\boldsymbol{\sigma},\boldsymbol{\tau})+L[\Pi(\boldsymbol{\sigma},\boldsymbol{\tau}),\{\pi_c\}_c;\{\Pi(\nu_c)\}_c].\tag{5.16}$$

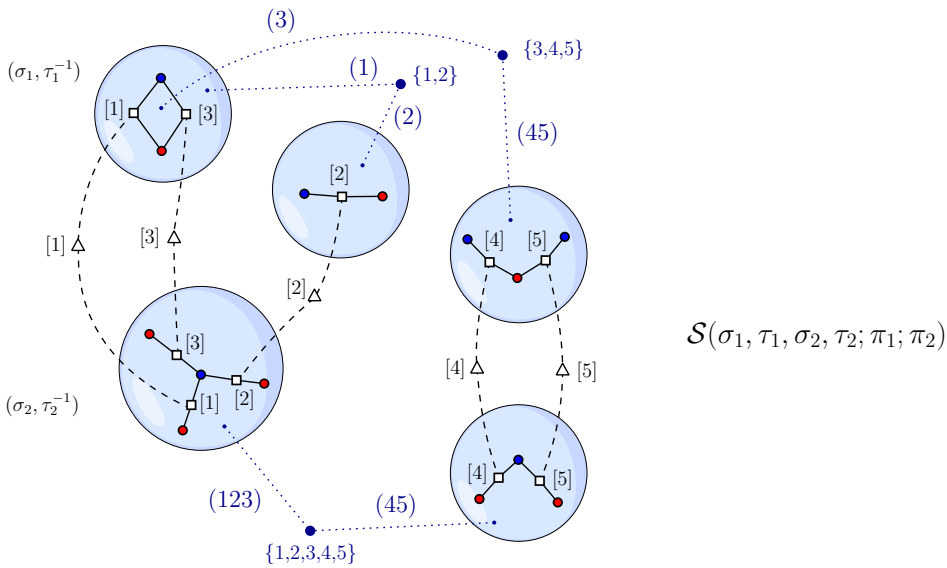

Fig. 7. The nodal surface $\mathcal{S}(\boldsymbol{\sigma}, \boldsymbol{\tau}; \pi_1, \pi_2)$ for the example of Fig. 6, where $\mathcal{S}(\boldsymbol{\sigma}, \boldsymbol{\tau})$ is as in Fig. 4 and with $\pi_1 = \{1, 2, 3\}\{4, 5\}$ and $\pi_2 = \{1, 2, 3, 4, 5\}$.

Proof. From the definition (5.1),

$$\mathcal{G}(\boldsymbol{\sigma}, \boldsymbol{\tau}, \{\pi_c\}_c) = \sum_{c=1}^{D} g(\sigma_c, \tau_c^{-1}) + nD + \sum_{c=1}^{D} \#(\sigma_c \tau_c^{-1}) - n - \sum_{c=1}^{D} (|\Pi(\sigma_c, \tau_c^{-1})| + \pi_c) + 1.$$

The result follows using (5.2) for $\mathcal{S}(\boldsymbol{\sigma}, \boldsymbol{\tau})$ as well as (5.14). ∎

We can therefore express conditions (C′1) and (C′2) geometrically:

Lemma 5.13. *$\mathfrak{S}_{\mathcal{G}}(\boldsymbol{\sigma}, \boldsymbol{\tau})$ is the set of connected foldings of $\mathcal{S}(\boldsymbol{\sigma}, \boldsymbol{\tau})$ whose arithmetic genus is $\mathcal{G}$.*

We may express both m_C and p_C in terms of connected foldings of $\mathcal{S}(\boldsymbol{\sigma}, \boldsymbol{\tau})$ of bounded arithmetic genus. While in Prop. 5.9, the $\mathcal{S}(\boldsymbol{\sigma}, \boldsymbol{\tau}; \hat{\boldsymbol{\rho}})$ were counted with weight 1, now the $\mathcal{S}(\boldsymbol{\sigma}, \boldsymbol{\tau}; \{\pi_c\}_c)$ must be counted with a non-trivial weight that takes into account all the different choices of $\hat{\boldsymbol{\eta}}$ satisfying the conditions in Prop. 5.9 and that lead to the same $\mathcal{S}(\boldsymbol{\sigma}, \boldsymbol{\tau}; \{\pi_c\}_c)$. For $m_C(\boldsymbol{\sigma}, \boldsymbol{\tau}; l, k)$, from (4.8), for each color c this weight is the number $M(\pi_c, \nu_c; l_c, k_c)$ of proper k_c-constellations with respect to the partition π_c, that is, $\Pi(\hat{\eta}^c) = \pi_c$. However, as already mentioned in Sec. 4.2, this quantity does not factor over the connected components of $\hat{\eta}^c$, since a permutation η_i^c may still be different from the identity but reduce to the identity on a subset of $\{1, \ldots, n\}$ corresponding to a connected component of $\hat{\eta}^c$ (graphically, the vertices of flavor i may all be the identity on a connected component of $\hat{\eta}^c$ without it being the case for all connected components of $\hat{\eta}^c$). This means that $M(\pi_c, \nu_c; l_c, k_c)$ cannot be expressed as a product of weights associated to some of the vertices of $\mathcal{S}(\boldsymbol{\sigma}, \boldsymbol{\tau}; \{\pi_c\}_c)$.

On the other hand, for $p_C[\boldsymbol{\sigma},\boldsymbol{\tau};l]$, one has factorization over the blocks of π_c, leading to the following geometrical interpretation of the formula (4.6) of Thm. 4.1 in terms of the nodal surfaces $\mathcal{S}(\boldsymbol{\sigma},\boldsymbol{\tau};\{\pi_c\}_c)$, with non-trivial combinatorial weights. We let $V_\bullet^c(\mathcal{S})$ be the set of nodes of a folding $\mathcal{S}=\mathcal{S}(\boldsymbol{\sigma},\boldsymbol{\tau};\{\pi_c\}_c)$ of $\mathcal{S}(\boldsymbol{\sigma},\boldsymbol{\tau})$ that correspond to the blocks of π_c, called *nodes of color* c, and let $V_\bullet(\mathcal{S})=\bigsqcup_c V_\bullet^c(\mathcal{S})$ be the set of *colored nodes*. For $v\in V_\bullet^c(\mathcal{S})$, we also let $\nu_{c|v}$ be the restriction of $\nu_c=\sigma_c\tau_c^{-1}$ to the block $B_c\in\pi_c$ corresponding to v, and $c(\nu_{c|v})$ its cycle type, which is a partition of $|B_c|$. We recall that for $\alpha=c(\nu_{c|v})$, $|C_\alpha|=\frac{|B_c|!}{\prod_{p\geq 1}p^{d_p(\alpha)}d_p(\alpha)!}$, where $d_p(\alpha)$ is the number of parts of α of size p.

Proposition 5.14. *With these notations, $p_C[\boldsymbol{\sigma},\boldsymbol{\tau};l]$ can be expressed as a sum of connected foldings of $\mathcal{S}(\boldsymbol{\sigma},\boldsymbol{\tau})$ of bounded arithmetic genus, whose colored nodes are weighted by monotone single Hurwitz numbers:*

$$p_C[\boldsymbol{\sigma},\boldsymbol{\tau};l] = \sum_{\mathcal{G}=\mathcal{G}(\boldsymbol{\sigma},\boldsymbol{\tau})}^{\frac{1}{2}[l-\ell(\boldsymbol{\sigma},\boldsymbol{\tau})]} \sum_{\mathcal{S}\in\mathfrak{S}_{\mathcal{G}}(\boldsymbol{\sigma},\boldsymbol{\tau})} \sum_{\substack{\{g_v\}_{v\in V_\bullet(\mathcal{S})}\geq 0\\ \sum_{v\in V_\bullet(\mathcal{S})} g_v=\frac{1}{2}[l-\ell(\boldsymbol{\sigma},\boldsymbol{\tau})]+\mathcal{G}(\boldsymbol{\sigma},\boldsymbol{\tau})-\mathcal{G}}} \prod_{v\in V_\bullet^c(\mathcal{S})} \frac{\vec{H}_{g_v}(c(\nu_{c|v}))}{|C_{c(\nu_{c|v})}|}. \tag{5.17}$$

The topological expansion of the cumulant Weingarten functions in Prop. 5.9 can therefore also be re-expressed as a topological expansion over connected foldings of $\mathcal{S}(\boldsymbol{\sigma},\boldsymbol{\tau})$ of fixed arithmetic genus.

With this interpretation, generating all the contributions to $p_C[\boldsymbol{\sigma},\boldsymbol{\tau};l]$ is quite simple: fixing $\boldsymbol{\sigma},\boldsymbol{\tau}$ and l, one sums over the excess L between 0 and $(l-\ell)/2$, and over all possible ways to add nodes of color c (represented by star-vertices of color c) for every color $c\in\{1,\dots,D\}$ between all the faces of $\mathcal{S}(\boldsymbol{\sigma},\boldsymbol{\tau})$, so that the resulting (class of) nodal surface is connected and the graph obtained when contracting the connected components of $\mathcal{S}(\boldsymbol{\sigma},\boldsymbol{\tau})$ to points has L excess edges. This generates all the foldings of $\mathcal{S}(\boldsymbol{\sigma},\boldsymbol{\tau})$ of arithmetic genus $\mathcal{G}(\boldsymbol{\sigma},\boldsymbol{\tau})+L$. For each such folding $\mathcal{S}=\mathcal{S}(\boldsymbol{\sigma},\boldsymbol{\tau};\{\pi_c\}_c)$, each node of color c corresponds to a block B_c of π_c. One then distributes the total genus $(l-\ell)/2-L$ (see (5.13)) among all the nodes of color c, and each such node is endowed with a weight $\gamma_{l(g_{B_c})}(\nu_{c|B_c})$, which precisely takes into account the contributions of all the connected constellations of genus g_{B_c} corresponding to the connected component of $\hat{\eta}^c$ for every $\mathcal{S}(\boldsymbol{\sigma},\boldsymbol{\tau},\hat{\boldsymbol{\eta}})$ that contracts to $\mathcal{S}$. From (4.21) this factor $\gamma_{l(g_{B_c})}(\nu_{c|B_c})$ is proportional to a genus-g_{B_c} single monotone Hurwitz number (the signs combine into the overall $(-1)^l$ factor in (4.6)).

The simplified version of this geometrical picture corresponding to $D=1$ and minimal $l=\ell$ has been introduced in [46].

Acknowledgments. L.L. thanks JSPS and Kyoto University, where the discussions at the origin of this project took place. L.L. thanks G. Chapuy for useful references on constellations. The authors would like to thank Jonathan Novak for insightful comments on a preliminary version.

Funding. B.C. was supported by JSPS KAKENHI 17K18734 and 17H04823. R.G. is supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 818066) and by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC-2181/1 - 390900948 (the Heidelberg STRUCTURES Cluster of Excellence). L.L. acknowledges support of the START-UP 2018 programme with project number 740.018.017, which is financed by the Dutch Research Council (NWO). This project has also received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. ERC-2016-STG 716083, "CombiTop").

References

[1] Alexandrov, A., Chapuy, G., Eynard, B., Harnad, J.: Fermionic approach to weighted Hurwitz numbers and topological recursion. Comm. Math. Phys. **360**, 777–826 (2018) Zbl 1436.14060 MR 3800798

[2] Alexandrov, A., Chapuy, G., Eynard, B., Harnad, J.: Weighted Hurwitz numbers and topological recursion. Comm. Math. Phys. **375**, 237–305 (2020) Zbl 1472.37078 MR 4082183

[3] Ben Geloun, J., Rivasseau, V.: A renormalizable 4-dimensional tensor field theory. Comm. Math. Phys. **318**, 69–109 (2013) Zbl 1261.83016 MR 3017064

[4] Bleher, P. M., Kuijlaars, A. B. J.: Random matrices with external source and multiple orthogonal polynomials. Int. Math. Res. Notices **2004** 109–129 Zbl 1082.15035 MR 2038771

[5] Bousquet-Mélou, M., Schaeffer, G.: Enumeration of planar constellations. Adv. Appl. Math. **24**, 337–368 (2000) Zbl 0955.05004 MR 1761777

[6] Brézin, E., Hikami, S.: Correlations of nearby levels induced by a random potential. Nuclear Phys. B **479**, 697–706 (1996) Zbl 0925.82117 MR 1418841

[7] Brézin, E., Hikami, S.: Level spacing of random matrices in an external source. Phys. Rev. E (3) **58**, 7176–7185 (1998) MR 1662382

[8] Brézin, E., Hikami, S.: Extension of level-spacing universality. Phys. Rev. E **56**, 264–269 (1997)

[9] Bychkov, B., Dunin-Barkowski, P., Shadrin, S.: Combinatorics of Bousquet-Mélou-Schaeffer numbers in the light of topological recursion. Eur. J. Combin. **90**, art. 103184, 35 pp. (2020) Zbl 1458.05262 MR 4125529

[10] Carrozza, S., Oriti, D., Rivasseau, V.: Renormalization of a $SU(2)$ tensorial group field theory in three dimensions. Comm. Math. Phys. **330**, 581–637 (2014) Zbl 1300.83023 MR 3223482

[11] Cavalieri, R., Miles, E.: Riemann Surfaces and Algebraic Curves. London Math. Soc. Student Texts 87, Cambridge Univ. Press, Cambridge (2016) Zbl 1354.14001 MR 3585685

[12] Chapuy, G.: Asymptotic enumeration of constellations and related families of maps on orientable surfaces. Combin. Probab. Comput. **18**, 477–516 (2009) Zbl 1221.05204 MR 2507734

[13] Collins, B.: Moments and cumulants of polynomial random variables on unitary groups, the Itzykson–Zuber integral, and free probability. Int. Math. Res. Notices **2003**, 953–982 Zbl 1049.60091 MR 1959915

[14] Collins, B., Matsumoto, S.: Weingarten calculus via orthogonality relations: new applications. ALEA Latin Amer. J. Probab. Math. Statist. **14**, 631–656 (2017) Zbl 1375.60018 MR 3680193

[15] Collins, B., Śniady, P.: Integration with respect to the Haar measure on unitary, orthogonal and symplectic group. Comm. Math. Phys. **264**, 773–795 (2006) Zbl 1108.60004 MR 2217291

[16] Di Francesco, P., Ginsparg, P., Zinn-Justin, J.: 2D gravity and random matrices. Phys. Rep. **254**, 133 pp. (1995) MR 1320471

[17] Disertori, M., Gurau, R., Magnen, J., Rivasseau, V.: Vanishing of beta function of non-commutative Φ_4^4 theory to all orders. Phys. Lett. B **649**, 95–102 (2007) Zbl 1248.81253 MR 2308801

[18] Do, N., Dyer, A., Mathews, D. V.: Topological recursion and a quantum curve for monotone Hurwitz numbers. J. Geom. Phys. **120**, 19–36 (2017) Zbl 1373.14051 MR 3712146

[19] Eichhorn, A., Koslowski, T.: Continuum limit in matrix models for quantum gravity from the Functional Renormalization Group. Phys. Rev. D **88**, art. 084016, 15 pp. (2013)

[20] Eichhorn, A., Koslowski, T., Pereira, A. D.: Status of background-independent coarse-graining in tensor models for quantum gravity. Universe **5**, no. 2, 53 pp. (2019)

[21] Eynard, B.: Eigenvalue distribution of large random matrices, from one matrix to several coupled matrices. Nuclear Phys. B **506**, 633–664 (1997) Zbl 0925.82121 MR 1488592

[22] Eynard, B.: Counting Surfaces. Progr. Math. Phys. 70, Birkhäuser/Springer, (2016) Zbl 1338.81005 MR 3468847

[23] Goulden, I. P., Guay-Paquet, M., Novak, J.: Monotone Hurwitz numbers in genus zero. Canad. J. Math. **65**, 1020–1042 (2013) Zbl 1280.05008 MR 3095005

[24] Goulden, I. P., Guay-Paquet, M., Novak, J.: Polynomiality of monotone Hurwitz numbers in higher genera. Adv. Math. **238**, 1–23 (2013) Zbl 1285.05008 MR 3033628

[25] Goulden, I. P., Guay-Paquet, M., Novak, J.: Monotone Hurwitz numbers and the HCIZ integral. Ann. Math. Blaise Pascal **21**, 71–89 (2014) Zbl 1296.05202 MR 3248222

[26] Goulden, I. P., Guay-Paquet, M., Novak, J.: Monotone Hurwitz numbers and the HCIZ integral II. arXiv:1107.1001 (2011)

[27] Grosse, H., Wulkenhaar, R.: Renormalisation of ϕ^4-theory on noncommutative $\mathbb{R}^2$ in the matrix base. J. High Energy Phys. **2003**, no. 12, art. 019, 26 pp. Zbl 1075.82005 MR 2041157

[28] Guay-Paquet, M., Harnad, J.: Generating functions for weighted Hurwitz numbers. J. Math. Phys. **58**, art. 083503, 28 pp. (2017) Zbl 1369.05013 MR 3683833

[29] Guionnet, A.: Large deviations and stochastic calculus for large random matrices. Probab. Surv. **1**, 72–172 (2004) Zbl 1189.60059 MR 2095566

[30] Harish-Chandra: Differential operators on a semisimple Lie algebra. Amer. J. Math. **79**, 87–120 (1957) Zbl 0072.01901 MR 84104

[31] Harnad, J.: Weighted Hurwitz numbers and hypergeometric τ-functions: an overview. In: String-Math 2014, Proc. Sympos. Pure Math. 93, Amer. Math. Soc., Providence, RI, 289–333 (2016) Zbl 1358.33004 MR 3525997

[32] Itzykson, C., Zuber, J. B.: The planar approximation. II. J. Math. Phys. **21**, 411–421 (1980) Zbl 0997.81549 MR 562985

[33] Kontsevich, M.: Intersection theory on the moduli space of curves and the matrix Airy function. Comm. Math. Phys. **147**, 1–23 (1992) Zbl 0756.35081 MR 1171758

[34] Lando, S. K., Zvonkin, A. K.: Graphs on Surfaces and Their Applications. Encyclopaedia Math. Sci. 141, Springer, Berlin (2004) Zbl 1040.05001 MR 2036721

[35] Lando, S. K., Zvonkine, D.: Counting ramified converings and intersection theory on spaces of rational functions. I. Cohomology of Hurwitz spaces. Moscow Math. J. **7**, 85–107, 167 (2007) Zbl 1131.14034 MR 2324558

[36] Louf, B.: Simple formulas for constellations and bipartite maps with prescribed degrees. Canad. J. Math. **73**, 160–176 (2021) Zbl 1436.05134 MR 4201537

[37] Matsumoto, S., Novak, J.: Jucys–Murphy elements and unitary matrix integrals. Int. Math. Res. Notices **2013**, 362–397 (2013) Zbl 1312.05143 MR 3010693

[38] Mehta, M. L.: A method of integration over matrix variables. Comm. Math. Phys. **79**, 327–340 (1981) Zbl 0471.28007 MR 627056

[39] Nica, A., Speicher, R.: Lectures on the Combinatorics of Free Probability. London Math. Soc. Lecture Note Ser. 335, Cambridge Univ. Press, Cambridge (2006) Zbl 1133.60003 MR 2266879

[40] Novak, J.: On the complex asymptotics of the HCIZ and BGW integrals. arXiv:2006.04304 (2020)

[41] Rota, G.-C.: On the foundations of combinatorial theory. I. Theory of Möbius functions. Z. Wahrsch. Verw. Gebiete **2**, 340–368 (1964) Zbl 0121.02406 MR 174487

[42] Seiberg, N., Witten, E.: String theory and noncommutative geometry. J. High Energy Phys. **1999**, art. 32, 93 pp. Zbl 0957.81085 MR 1720697

[43] Weingarten, D.: Asymptotic behavior of group integrals in the limit of infinite rank. J. Math. Phys. **19**, 999–1001 (1978) Zbl 0388.28013 MR 471696

[44] Zinn-Justin, P.: Random Hermitian matrices in an external field. Nuclear Phys. B **497**, 725–732 (1997) Zbl 0933.82022 MR 1463645

[45] Zinn-Justin, P.: Universality of correlation functions of Hermitian random matrices in an external field. Comm. Math. Phys. **194**, 631–650 (1998) Zbl 0912.15028 MR 1631489

[46] Zinn-Justin, P., Zuber, J.-B.: On some integrals over the $\mathrm{U}(N)$ unitary group and their large N limit. J. Phys. A **36**, 3173–3193 (2003) Zbl 1074.82013 MR 1986413

[47] Zvonkine, D.: An introduction to moduli spaces of curves and their intersection theory. In: Handbook of Teichmüller Theory. Vol. III, IRMA Lect. Math. Theor. Phys. 17, Eur. Math. Soc., Zürich, 667–716 (2012) Zbl 1264.14071 MR 2952773