Ann. Inst. H. Poincaré D
Comb. Phys. Interact. (Online first)
DOI 10.4171/AIHPD/201

Published by EMS Press

Topological expansion of unitary integrals and maps

Thomas Buc-d'Alché

Abstract. We study integrals on the unitary group with respect to the Haar measure. We give a combinatorial interpretation in terms of maps of the asymptotic topological expansion, established previously by Guionnet and Novak. The maps we introduce—the maps of unitary type—satisfy Tutte-like equations. It allows us to show that in the perturbative regime, they describe the different orders of the asymptotic topological expansion. Furthermore, they generalize the monotone Hurwitz numbers.

1. Introduction

In the breakthrough article [5], Brézin et al. used random matrix theory to address the problem of enumeration of maps, graphs embedded in surfaces up to homeomorphisms. The topological properties of Feynman diagrams had previously been shown to be critical in the work of 't Hooft [39], thus relating the combinatorics of maps to field theory (see also the review article [1]). For instance, the planar diagrams give the leading order in the expansion of physically significant quantities.

The random matrix approach to the enumeration of maps pioneered by Brézin et al. subsequently found many applications. Harer and Zagier used the same approach to study the topological properties of the moduli space of curves [24]. In the celebrated article [28], Kontsevitch used matrix integrals to solve Witten's conjecture. See also [12] for a review of the application of random matrix theory to combinatorial problems appearing in 2D gravity. More generally, random matrices provide a powerful tool to address hard combinatorial problems such as the problem of the enumeration of Riemann surfaces, see the work of Eynard [17]. For another approach on the enumeration of maps, see, for instance, [3].

In all the problems above, the matrix models used are related to the Gaussian Unitary Ensemble (GUE). Let $\mathrm{d}M = \prod_i \mathrm{d}M_{ii} \prod_{i<j} \mathrm{d}\,\mathrm{Re}(M_{ij})\mathrm{d}\,\mathrm{Im}(M_{ij})$ be the Lebesgue measure on the space of Hermitian matrices $\mathcal{H}_N$ and V be a polynomial

---

Mathematics Subject Classification 2020: 15B52 (primary); 60B20 (secondary).
Keywords: random matrix theory, unitary group, Haar measure, Weingarten calculus.

called the potential. We consider the measure

$$\mu^N_{\mathrm{GUE},V} = \frac{1}{Z^N_{\mathrm{GUE},V}} \mathrm{e}^{-N \operatorname{Tr} V(M) - \frac{N}{2} \operatorname{Tr} M^2} \mathrm{d}M,$$

where the normalization constant is the partition function

$$Z^N_{\mathrm{GUE},V} = \int_{\mathcal{H}_N} \mathrm{e}^{-N \operatorname{Tr} V(M) - \frac{N}{2} \operatorname{Tr} M^2} \mathrm{d}M.$$

Many relevant quantities, such as the partition function can be expressed as a formal series of maps using Wick's formula (see [44] for an introduction). For instance, for $V(M) = tq(M)$, with $q(M) = M^4$ and $t \in \mathbb{R}$, we get the formal expansion in the dimension N of the random matrix M,

$$\ln Z^N_{\mathrm{GUE},V} = \sum_{g \geq 0} N^{2-2g} \sum_{n \geq 0} \frac{(-t)^n}{n!} \mathcal{M}^{\mathrm{GUE},(g)}_{q,n},$$

where $\mathcal{M}^{\mathrm{GUE},(g)}_{q,n}$ is the number of connected maps of genus g with n vertices, all of them of degree 4. Notice that the term of order N^{2-2g} in this expansion is a generating series of maps of genus g. We call such an expansion a (formal) *topological expansion*.

In general, the above equality holds in the sense of formal power series, see [16] for instance. By the above equality, we mean that the derivatives with respect to t at $t = 0$ of the left and right sides of the equation above coincide. In fact, the series of maps on the right side may not converge in general.

We can replace this divergent series with an asymptotic expansion as $N \to \infty$, where the equality holds up to an error of order N^{-p}, for some integer p. Ercolani and McLaughlin obtained such an expansion in a one-matrix model for a potential whose coefficients are close to zero [15]. The case with several random matrices was studied by Guionnet and Maurel-Segala [21, 22], and Maurel-Segala [31]. More complicated models involving not only a matrix from the GUE but also deterministic matrices, sometimes called models with external sources, have been studied, see [4].

The multi-matrix models display much more variety. As for the one-matrix models, they were first studied by physicists, see the reviews [12, 20]. From an analytical point of view, they are harder to solve than one-matrix models; see, for instance, the works of Mehta [32], and from a combinatorial point of view, they allow to address a wealth of combinatorial problems as they are related to the enumeration of colored maps; see, for instance, [21].

In this article, we establish a similar link between integrals of unitary matrices and the combinatorics of some maps. More precisely, we introduce new maps, the maps of unitary type (Definition 3.13), that describe the topological expansion. These maps allow us to relate the Weingarten calculus and the Dyson–Schwinger equation—two important ways to study unitary integrals. In a particular case, the maps of unitary type are related to the Hurwitz numbers. In this way, we generalize part of the results obtained in [19], which relate a particular integral, the HCIZ integral, to Hurwitz numbers.

The Haar unitary matrices share the same unitary invariance as matrices of the Ginibre ensemble. An expansion in terms of non-crossing permutations for expectation of traces of words of Ginibre matrices G_i has been obtained in [33]. This point of view in terms of non-crossing permutations is similar to the interpretation in terms of maps. For instance, the annuli considered in multi-annulus permutations correspond to the vertices in a map. A genus expansion in terms of maps has been obtained in [13]. They consider only some pairings, *admissible pairings*, which correspond to the fact that edges are oriented in the maps of unitary type. In particular, the unitary invariance of the Ginibre ensemble implies that to have a non-zero expectation, the words in Ginibre matrices considered must be *balanced*, i.e., contain as many G_i as G_i^*. A similar condition appears for Haar unitary matrices.

We introduce some notation. We consider matrices of dimension $N \in \mathbb{N}^*$, where $\mathbb{N}^* = \{1, 2, 3, \ldots\}$. We denote by $\operatorname{Tr} A = \sum_{i=1}^N A_{ii}$ the trace of a matrix A, and by $\operatorname{tr} = \operatorname{Tr}/N$ the normalized trace. Notice that both Tr and tr depend on the dimension N. The conjugate transpose of a matrix M is denoted by M^*. Let $p \in \mathbb{N}^*$. For all $N \geq 1$, we fix p deterministic matrices $A_1^N, \ldots, A_p^N$ of size $N \times N$. The matrix U^N will be a unitary matrix of size $N \times N$, i.e., an element of the unitary group $\mathbb{U}(N)$, and $(U^N)^* = (U^N)^{-1}$ will be its conjugate transpose.

Let $\mathrm{d}U^N$ be the Haar measure on the unitary group $\mathbb{U}(N)$, and V be a polynomial in several non-commutative variables, which does not depend on N. The measure μ_V^N is given by

$$\mathrm{d}\mu_V^N(U^N) = \frac{1}{Z_V^N} \exp\bigl(N \operatorname{Tr} V\bigl(U^N, (U^N)^*, A_1^N, (A_1^N)^*, \ldots, A_p^N, (A_p^N)^*\bigr)\bigr)\mathrm{d}U^N, \tag{1.1}$$

where the partition function Z_V^N is

$$Z_V^N = \int_{\mathbb{U}(N)} \exp\bigl(N \operatorname{Tr} V\bigl(U^N, (U^N)^*, A_1^N, (A_1^N)^*, \ldots, A_p^N, (A_p^N)^*\bigr)\bigr)\mathrm{d}U^N. \tag{1.2}$$

We will evaluate all non-commutative polynomials at the matrices

$$U^N, (U^N)^*, A_1^N, (A_1^N)^*, \ldots, A_p^N, (A_p^N)^*$$

and will omit writing this explicitly in the sequel, e.g., writing $\operatorname{Tr}(V)$ to mean

$$\operatorname{Tr}(V(U^N, (U^N)^*, A_1^N, \ldots, A_p^N)).$$

In Section 5, we will consider measures of the form

$$\frac{1}{Z_V^N} \exp(N \operatorname{Tr} V) \mathrm{d}U_1^N \cdots \mathrm{d}U_n^N,$$

where V is a noncommutative polynomial that depends on $U_1^N, \ldots, U_n^N$, all independent and Haar-distributed.

We will assume the two following hypotheses.

Hypothesis 1.1. For all $N \geq 1$ and for all $U_1, \ldots, U_n \in \mathbb{U}(N)^n$, $\operatorname{Tr} V$ is real.

Hypothesis 1.2. Assume that

$$\sup_{N \geq 1} \sup_{1 \leq i \leq p} \|A_i^N\| < \infty,$$

where $\|\cdot\|$ is the operator norm.

In most of the article, we will not assume Hypothesis 1.2, but rather Hypothesis 1.3.

Hypothesis 1.3. For all $N \geq 1$ and for all $1 \leq i \leq p$, $\|A_i^N\| \leq 1$, where $\|\cdot\|$ is the operator norm.

This will prove convenient and will not change the main result, Theorem 1.4 stated below. Stating Theorem 1.4 with Hypothesis 1.2 instead of Hypothesis 1.3 corresponds to rescaling the coefficients of the polynomial V.

Hypothesis 1.1 implies that the measure μ_V^N is a probability measure, and in particular that $Z_V^N \in (0, +\infty)$. We write the potential V as a sum of monomials q_i with complex coefficients z_i, $V = \sum_i z_i q_i$. Thus, we will sometimes consider the partition functions, cumulants, etc., as functions of $\boldsymbol{z} = (z_1, z_2, \ldots)$. With this notation, the reality conditions is

$$\sum_i z_i \operatorname{Tr}(q_i) = \sum_i \overline{z_i} \operatorname{Tr}(q_i^*).$$

Notice that for generic q_i's, $\operatorname{Tr} V$ might be real for only specific values of $\boldsymbol{z}$.

When considering the partition function with potential $V = tAU^N B(U^N)^*$, where $t \in \mathbb{C}$ and A, B are self-adjoint matrices, we recover the Harish–Chandra–Itzykson–Zuber (HCIZ) integral

$$Z_V^N = \int_{\mathbb{U}(N)} \exp\big(tN \operatorname{Tr}(AU^N B(U^N)^*)\big) \mathrm{d}U^N,$$

which was first studied by Harish–Chandra [25] and Itzykson and Zuber [26], and whose asymptotics have been since investigated, see [19, 23, 35, 43].

We will compute joint moments and cumulants (see Definition 2.1) of the random variables $\mathrm{Tr}(P_1), \dots, \mathrm{Tr}(P_l)$ under μ_V^N (for V small), where the P_i are non-commutative polynomials. In [8], the first-order asymptotics of partition functions was studied. In [23], it has been shown that the joint cumulants admit an asymptotic expansion as $N \to \infty$, when the coefficients of the potential V are small enough.

The goal of this article is to give a combinatorial interpretation of the coefficients of this expansion. We show that unitary matrix integrals enumerate a particular family of maps, which we call maps of unitary type. They are introduced in Section 3.2, Definition 3.13. This interpretation links the Dyson–Schwinger equation, which is satisfied by sums of maps of unitary type, and the Weingarten calculus studied first by Weingarten [42], and then by [37], whose results were rediscovered and expanded upon by Collins [7] and Collins and Śniady [11]. See [10] for a review.

Expansions in terms of combinatorial objects have already been introduced for unitary matrices. For instance, in the case of the HCIZ integral, expansions for the free energy using double Hurwitz numbers are computed in [19]. In [8], the leading order of the expansion of unitary integrals is expressed in terms of maps with "dotted edges". However, to our knowledge, no interpretation of these expansions using maps has been obtained at all orders for the unitary integrals we consider. As an interesting particular case, when considering alternated polynomials (see Definition 3.39), the combinatorics of maps of unitary type is related to triple Hurwitz numbers.

In the case of the GUE, integrals of random matrices and enumeration of maps are related by Wick's formula. In the case of unitary matrices, Wick's formula is replaced by Weingarten's formula. In Section 2, we express joint moments of random variables $\mathrm{Tr}(P_i)$, for non-commutative polynomials P_i, using Weingarten's formula. In the case where the potential $V = 0$, we can express such moments as weighted sums of permutations. In Section 3, we recall a few notions on maps and introduce the maps of unitary type, which are our main combinatorial tools. This allows us to deduce a topological expansion for the joint cumulants in the case of no potential (i.e., $V = 0$). To address the general case $V \neq 0$, we introduce generating series of maps of unitary type of the form

$$\mathcal{M}_{V,l}^{(g),N}(P_1, \dots, P_l) = \sum_{\boldsymbol{n} \in \mathbb{N}^k} \frac{\boldsymbol{z}^{\boldsymbol{n}}}{\boldsymbol{n}!} \times \sum w_N(\mathcal{C}, \boldsymbol{n}, V, P_1, \dots, P_l),$$

where the second sum is on a set of connected maps $\mathcal{C}$ of unitary type (see Definition 3.13) of genus g which depends on $V, P_1, \dots, P_l, \boldsymbol{n}$. The term $w_N(\mathcal{C}, \boldsymbol{n}, V, P_1, \dots, P_l)$ is a weight which depends on the size N, $\mathcal{C}$, $\boldsymbol{n}$ and on $V, P_1, \dots, P_l$. See Definition 3.38.

In Section 4, we describe a decomposition of maps of unitary type, which can be interpreted as a cutting procedure. It allows us to deduce induction relations—similar to the topological recursion of Chekhov, Eynard, and Orantin, see [6, 18]—on weighted sums $\mathcal{M}_{V,l}^{(g),N}$ of maps of unitary type of a given genus g. This decomposition is reminiscent of a procedure introduced by Tutte [40]. In Section 5, we extend the results obtained so far to the case of integrals over several independent random unitary matrices $U_1^N, \dots, U_n^N$.

It turns out that the induction relations obtained in Section 4 are related to the Dyson–Schwinger lattice. The Dyson–Schwinger lattice (see [23]) is a family of equations relating cumulants together, which generalize the Dyson–Schwinger equation (see equation (6.1)). This equation admits under some hypotheses a unique solution [8]. Furthermore, in [23], the Dyson–Schwinger lattice has been used to establish the existence of an asymptotic expansion of the cumulants, when $N \to \infty$. Let us assume Hypotheses 1.1 and 1.3, and that the joint law of the matrices A_i^N, tr admits an asymptotic expansion as $N \to \infty$. For all h, we have an asympotic expansion for the renormalized joint cumulants $N^{l-2}\mathcal{W}_{V,l}^N(P_1, \dots, P_l)$ (introduced in Definition 2.3) when the coefficients of the potential V are small enough

$$N^{l-2}\mathcal{W}_{V,l}^N(P_1, \dots, P_l) = \sum_{g=0}^{h} \frac{\tau_{l,g}^V(P_1, \dots, P_l)}{N^{2g}} + o(N^{-2h}), \tag{1.3}$$

where the coefficients $\tau_{l,g}^V(P_1, \dots, P_l)$ are uniquely defined by some induction relations.

In Section 6, we use the same techniques to express the terms of this expansion in terms of maps of unitary type. We obtain a topological expansion: the coefficient of $\frac{1}{N^{2g}}$ in the expansion is a generating series of weighted unitary type maps of genus g.

We thus improve on the result of [23, Theorem 25] by relaxing the hypotheses, showing that the convergence is uniform in g and l, and by giving a combinatorial interpretation to the coefficients $\tau_{l,g}^V(P_1, \dots, P_l)$.

Theorem 1.4 (Main theorem). *Assume that for all $N \geq 1$, $\mathrm{Tr}(V)$ is real for all unitary matrices $U_1, \dots, U_n \in \mathbb{U}(N)^n$ and that $\sum_{N\geq 1} \sup_{1\leq i\leq p} \|A_i^N\| < \infty$.*

There exists $\varepsilon > 0$ such that if

$$\|z\|_\infty < \varepsilon,$$

then for all $l \geq 1$, $g \geq 0$, and $\boldsymbol{P} = (P_1, \dots, P_l)$, we have the asymptotic expansion as $N \to \infty$

$$N^{l-2}\mathcal{W}_{V,l}^N(P_1, \dots, P_l) = \sum_{h=0}^{g} \frac{1}{N^{2h}} \mathcal{M}_{V,l}^{(h),N}(P_1, \dots, P_l) + \mathcal{O}(N^{-2g-2}).$$

Notice that we do not require the trace Tr to have an asymptotic expansion as in [23, Theorem 25].

An interesting particular case described in Section 3.6 is when all the polynomials involved are *alternated*, see Definition 3.39, that is, if they can be written as

$$P = B_1^N U^N C_1^N (U^N)^* \cdots B_m^N U^N C_m^N (U^N)^*,$$

where B_i^N and C_i^N for $i = 1, \dots, m$ are square $N \times N$ matrices. This is the case of the HCIZ integral in particular. In that case, our sums of maps are related to the triple monotone Hurwitz numbers, which count some ramified coverings of the sphere with at most three nonsimple ramification points. We thus generalize the link between the (double) monotone Hurwitz numbers and the HCIZ integral, which had already been studied in [19]. See also [9] for a study of the HCIZ integral in the tensor setting.

In Section 2, we give definitions and recall important consequences of Weingarten calculus. In Section 3, we introduce the maps of unitary types and show that they describe the topological expansion of cumulants with respect to the Haar measure. When the polynomials are alternated, these maps are related to the triple Hurwitz numbers. In Section 4, we give a decomposition of maps of unitary type and deduce induction relations on sums of maps of a given genus and with prescribed vertices, in the spirit of the work of Tutte [40]. In Section 6, we study the Dyson–Schwinger equation and give the proof of the main result.

2. Weingarten calculus

In this section, we first give a few definitions and introduce notation pertaining to moments and cumulants of traces of random matrices. Then, we give a short review of the Weingarten calculus. This allows us to give expression for the expectation of a product of traces of monomials in the matrices U^N, $(U^N)^*$, A_i^N, $(A^*)^N$.

2.1. Moments and cumulants

Let us consider $l \geq 1$ non-commutative polynomials $P_1, P_2, \dots, P_l$ in the variables u, u^{-1}, and a_i, a_i^* for $1 \leq i \leq p$, with $p \in \mathbb{N}$. We define the involution $*$ such that $u^* = u^{-1}$, for $1 \leq i \leq p$, $(a_i)^* = a_i^*$, and for any letters $X_1, \dots, X_k$ in $\{u, u^*, a_i, a_i^* \colon 1 \leq i \leq p\}$ and $z \in \mathbb{C}$, we have $(zX_1 \cdots X_k)^* = z^* X_k^* \cdots X_1^*$. We denote the unital ast-algebra generated by such polynomials by

$$\mathcal{A} = \mathbb{C}\langle u, u^{-1}, a_i, a_i^*; 1 \leq i \leq p\rangle.$$

The unital ast-algebra generated by the non-commutative polynomials in the formal variables $a_1, a_1^*, \dots, a_p, a_p^*$ only is denoted by $\mathcal{B}$. We will evaluate all polynomials P_i at the matrices U^N, $(U^N)^*$, A_1^N, $(A_1^N)^*, \dots, A_p^N$, $(A_p^N)^*$ and will omit writing this explicitly in the sequel, e.g., writing $\mathrm{Tr}(P)$ to mean $\mathrm{Tr}(P(U^N, (U^N)^*, A_1^N, \dots, A_p^N))$. Notice that there is no relation between the formal variables u and u^{-1}, or a_i and a_i^* for $i \in \mathbb{N}^*$ (except for those involving ast). We will denote by $\mathrm{tr} = \frac{1}{N}\mathrm{Tr}$ the normalized trace.

In this article, we study the random variables $\mathrm{Tr}(P_1), \dots, \mathrm{Tr}(P_l)$, seen as functions of U^N, under the measure μ_V^N (see (1.1)). We will be interested in computing the joint moments and cumulants of these random variables. To state the definition of the cumulants, we introduce some notation about partitions. We denote by $\mathcal{P}(I)$ the set of partitions of a finite set I. In particular, for $n \in \mathbb{N}^*$, we denote the set $\{1, 2, \dots, n\}$ by $[n]$. We denote the cardinality of a finite set I by $|I|$. Given a partition $\pi \in \mathcal{P}(I)$, $|\pi|$ is the number of blocks of π.

Definition 2.1. Let $k \in \mathbb{N}^*$. The *joint moment* of the complex random variables $X_1, \dots, X_k$ is

$$m_k(X_1, \dots, X_k) = \mathbb{E}[X_1 X_2 \cdots X_k].$$

The *joint cumulant* of the complex random variables $X_1, \dots, X_k$ is $c_k(X_1, \dots, X_k)$, defined recursively by

$$c_k(X_1, \dots, X_k) = m_k(X_1, \dots, X_k) - \sum_{\substack{\pi \in \mathcal{P}([k]) \\ |\pi| \geq 2}} \prod_{B \in \pi} c_{|B|}(X_i : i \in B).$$

Notice that both the joint moments and cumulants are symmetric, multilinear functions. It can be proved inductively. The symmetry makes $c_{|B|}(X_i : i \in B)$ above unambiguous.

Remark 2.2. The cumulants can also be defined, see [38], as the coefficients of the series of the logarithm of the exponential generating series of the moments

$$\sum_{n \geq 0} \frac{z^n}{n!} c_n(X_1, \dots, X_1) = \ln \sum_{n \geq 0} \frac{z^n}{n!} m_n(X_1, \dots, X_1).$$

Definition 2.3. For $(P_1, \dots, P_l) \in \mathcal{A}^l$, we write the joint moments of the traces of P_i's under μ_V^N as

$$\alpha_{V,l}^N(P_1, \dots, P_l) = m_l(\mathrm{Tr}(P_1), \dots, \mathrm{Tr}(P_l)) = \int_{\mathbb{U}(N)} \mathrm{Tr}(P_1) \cdots \mathrm{Tr}(P_l) \mathrm{d}\mu_V^N.$$

We write the joint cumulants under μ_N^V as

$$\mathcal{W}_{V,l}^N(P_1, \dots, P_l) = c_l(\mathrm{Tr}(P_1), \dots, \mathrm{Tr}(P_l)),$$

and introduce the renormalized cumulants

$$\widetilde{\mathcal{W}}_{V,l}^N(P_1, \dots, P_l) = N^{l-2} c_l(\mathrm{Tr}(P_1), \dots, \mathrm{Tr}(P_l)).$$

In Section 6, we will discuss an asymptotic expansion (as $N \to \infty$) for the joint cumulants. For now, we study the moments for N fixed. When $V = 0$, we can compute directly the moments using Weingarten's formula, see Section 2.2. When $V \neq 0$, we can compute the cumulants using the free energy F_V^N defined in terms of the partition function Z_V^N. Recall that

$$V = \sum_{i=1}^{k} z_i q_i$$

is the potential, a sum of k polynomials $q_1, \dots, q_k \in \mathcal{A}$ with complex coefficients $z_1, \dots, z_k$. Note that V does not depend on N. We have

$$Z_V^N = \int_{\mathbb{U}(N)} \exp(N \operatorname{Tr}(V)) \mathrm{d}U^N,$$

and we define the free energy as

$$F_V^N = \frac{1}{N^2} \ln Z_V^N. \tag{2.1}$$

The free energy is always well defined when $\operatorname{Tr} V$ is real.

In the expression of the partition function, we can develop the exponential as a series and exchange the sum and the integral

$$\begin{aligned} Z_V^N &= \int_{\mathbb{U}(N)} \sum_{n_1,\dots,n_k \geq 0} \prod_{i=1}^{k} \frac{(N z_i \operatorname{Tr}(q_i))^{n_i}}{n_i!} \mathrm{d}U^N \\ &= \sum_{n_1,\dots,n_k \geq 0} \prod_{i=1}^{k} \frac{(N z_i)^{n_i}}{n_i!} \int_{\mathbb{U}(N)} \operatorname{Tr}(q_1)^{n_1} \cdots \operatorname{Tr}(q_k)^{n_k} \mathrm{d}U^N. \end{aligned}$$

In the second line, we used Hypothesis 1.3, which implies that $|\operatorname{Tr}(q_i)| \leq N$, and the fact that we are integrating with respect to the Haar measure on the compact group $\mathbb{U}(N)$ to exchange the sum and the integral. Notice that this expression is valid for all $\boldsymbol{z}$, even if $\operatorname{Tr} V$ is not real.

We introduce the notation $\boldsymbol{z} = (z_1, \dots, z_k)$, and for $\boldsymbol{n} = (n_1, \dots, n_k) \in \mathbb{N}^k$,

$$\boldsymbol{z}^{\boldsymbol{n}} = \prod_{i=1}^{k} z_i^{n_i} \quad \text{and} \quad \boldsymbol{n}! = \prod_{i=1}^{k} n_i!.$$

Then,

$$Z_V^N = \sum_{n \geq 0} N^n \sum_{\substack{\boldsymbol{n} \in \mathbb{N}^k \\ n_1 + \dots + n_k = n}} \frac{\boldsymbol{z}^{\boldsymbol{n}}}{\boldsymbol{n}!} \alpha_{0,n}^N (\underbrace{q_1, \dots, q_1}_{n_1 \text{ times}}, \dots, \underbrace{q_k, \dots, q_k}_{n_k \text{ times}}),$$

and therefore, the partition function is a generating series of the moments with respect to the Haar measure (i.e., with $V = 0$).

Similarly, the free energy is a generating series of the renormalized cumulants for $V = 0$ (see [2, Theorem 1.3.3, 4.])

$$\begin{aligned} F_V^N &= \sum_{n \geq 1} \sum_{\substack{\boldsymbol{n} \in \mathbb{N}^k \\ n_1 + \cdots + n_k = n}} \frac{(N\boldsymbol{z})^{\boldsymbol{n}}}{\boldsymbol{n}!} \frac{1}{N^2} \mathcal{W}_{0,n}^N(\underbrace{q_1, \ldots, q_1}_{n_1 \text{ times}}, \ldots, \underbrace{q_k, \ldots, q_k}_{n_k \text{ times}}) \\ &= \sum_{n \geq 1} \sum_{\substack{\boldsymbol{n} \in \mathbb{N}^k \\ n_1 + \cdots + n_k = n}} \frac{\boldsymbol{z}^{\boldsymbol{n}}}{\boldsymbol{n}!} \widetilde{\mathcal{W}}_{0,n}^N(\underbrace{q_1, \ldots, q_1}_{n_1 \text{ times}}, \ldots, \underbrace{q_k, \ldots, q_k}_{n_k \text{ times}}). \end{aligned}$$

Notice that the free energy a priori exists only for $\boldsymbol{z}$ sufficiently small. Indeed, Z_V^N is defined for all $\boldsymbol{z}$ but is nonzero on an open neighborhood of 0 which depends on N. In particular, the radius of convergence of F_V^N a priori depends on N.

Notice that by modifying the potential V and differentiating, we have

$$\begin{aligned} \left.\frac{\partial}{\partial t}\right|_{t=0} F_{V+tP}^N &= \frac{1}{N} \int_{\mathbb{U}(N)} \operatorname{Tr}(P) \mathrm{d}\mu_V^N(U^N) \\ &= \frac{1}{N} \alpha_{V,1}^N(P) = \widetilde{\mathcal{W}}_{V,1}^N(P). \end{aligned}$$

In general, we can prove by induction the following lemma, which is a consequence of the definition of cumulants given in Remark 2.2.

Lemma 2.4. *The renormalized joint cumulants are given by*

$$\widetilde{\mathcal{W}}_{V,l}^N(P_1, \ldots, P_l) = \left.\frac{\partial^l}{\partial t_1 \partial t_2 \cdots \partial t_l}\right|_{t_1 = \cdots = t_l = 0} F_{V + \sum_i t_i P_i}^N.$$

Lemma 2.4 implies that for a fixed N, there exists a neighborhood $U_0 \in \mathbb{C}^k$ of 0 such that for $\boldsymbol{z} \in U_0$,

$$\begin{aligned} &\widetilde{\mathcal{W}}_{V,l}^N(P_1, \ldots, P_l) \\ &\quad = \sum_{n \geq 0} \sum_{\substack{\boldsymbol{n} \in \mathbb{N}^k \\ n_1 + \cdots + n_k = n}} \frac{\boldsymbol{z}^{\boldsymbol{n}}}{\boldsymbol{n}!} \widetilde{\mathcal{W}}_{0,n}^N(\underbrace{q_1, \ldots, q_1}_{n_1 \text{ times}}, \ldots, \underbrace{q_k, \ldots, q_k}_{n_k \text{ times}}, P_1, \ldots, P_l). \end{aligned} \tag{2.2}$$

In the next subsections, we compute the moments with respect to the Haar measure. From these moments and Definition 2.1, we can compute the cumulants with respect to the Haar measure. The expression (2.2) motivates the introduction in Section 3.5 of a formal sum. The first terms of this sum are shown to give the asymptotic expansion of the cumulants in Theorem 1.4.

2.2. The Weingarten formula

To compute the moments with respect to Haar measure, the key tool is Weingarten's formula, first obtained in [42], which expresses the average of coefficients of a unitary matrix in terms of the Weingarten function defined below (Definition 2.5). See [10] for a review on the Weingarten calculus.

The Weingarten formula involves a sum over permutations. Let us fix some notation pertaining to permutations. For I a finite set, we denote by $\mathfrak{S}(I)$ the set of permutations on this set. In particular, $\mathfrak{S}_n = \mathfrak{S}([n])$ is the set of permutations on $[n] = \{1, 2, \dots, n\}$. A permutation σ admits a decomposition in disjoint cycles. The set of cycles of σ is denoted by $\mathrm{Cycles}(\sigma)$ and its number of cycles is denoted by $c(\sigma)$. A cycle $c \in \mathrm{Cycles}(\sigma)$ is written as $(u_1\, u_2\, \dots\, u_k)$ with distinct $u_1, \dots, u_k$. It is the permutation whose support is $\{u_1, \dots, u_k\}$ and such that $c(u_i) = u_{i+1}$ with the convention $u_{k+1} = u_1$.

We also introduce the modified traces $\mathrm{Tr}_\sigma(\boldsymbol{M})$ and $\mathrm{tr}_\sigma(\boldsymbol{M})$ for $\sigma \in \mathfrak{S}(I)$ and $\boldsymbol{M} = (M_i, i \in I)$ a tuple of matrices, defined by

$$\begin{aligned} \mathrm{Tr}_\sigma(\boldsymbol{M}) &= \prod_{c \in \mathrm{Cycles}(\sigma)} \mathrm{Tr}\Big(\overrightarrow{\prod}_{i \in c} M_i\Big), \\ \mathrm{tr}_\sigma(\boldsymbol{M}) &= \prod_{c \in \mathrm{Cycles}(\sigma)} \mathrm{tr}\Big(\overrightarrow{\prod}_{i \in c} M_i\Big) = N^{-c(\sigma)}\,\mathrm{Tr}_\sigma(\boldsymbol{M}), \end{aligned} \tag{2.3}$$

where, if $c = (i_1\, \dots\, i_k)$ is a cycle of the permutation σ, the notation $\overrightarrow{\prod}_{i \in c} M_i$ stands for the non-commutative product $M_{i_1} M_{i_2} \cdots M_{i_k}$. Notice that such a non-commutative product is defined up to circular permutation. The trace property ensures that the quantity $\mathrm{Tr}_\sigma(\boldsymbol{M})$ is well defined.

Definition 2.5. Let $q \le N$ be an integer. The *Weingarten function* $\mathrm{Wg}_N : \mathfrak{S}_q \to \mathbb{C}$ is defined for all $\pi \in \mathfrak{S}_q$ by

$$\mathrm{Wg}_N(\pi) = \int_{\mathbb{U}(N)} (U^N)_{11} \cdots (U^N)_{qq} \overline{(U^N)_{1\pi(1)} \cdots (U^N)_{q\pi(q)}} \mathrm{d}U^N.$$

This function can also be defined for all $q \in \mathbb{N}^*$ using characters of the symmetric group (see [11]). The invariance of the Haar measure by multiplication by permutation matrices implies that the Weingarten function is invariant by conjugation, i.e., for all $\sigma, \pi \in \mathfrak{S}_q$, we have

$$\mathrm{Wg}_N(\sigma\pi\sigma^{-1}) = \mathrm{Wg}_N(\pi).$$

With our definition of the Weingarten function, Weigarten's formula is valid in the case $q \le N$. It actually holds for all $q \ge 1$ with an appropriate definition of the Weingarten function.

Theorem 2.6 (Weingarten's formula, see [7] and [11]). *Let U^N be a Haar-distributed unitary matrix of size $N \times N$ and $\boldsymbol{i} = (i_1, i_2, \dots, i_q)$, $\boldsymbol{j} = (j_1, j_2, \dots, j_q)$, $\boldsymbol{i'} = (i'_1, i'_2, \dots, i'_{q'})$ and $\boldsymbol{j} = (j'_1, j'_2, \dots, j'_{q'})$ be elements of $[N]^q$ or $[N]^{q'}$ for $q, q' \geq 1$.*

$$\begin{aligned} &\int_{\mathbb{U}(N)} (U^N)_{i_1 j_1} \cdots (U^N)_{i_q j_q} \overline{(U^N)_{i'_1 j'_1} \cdots (U^N)_{i'_{q'} j'_{q'}}} \mathrm{d}U^N \\ &\quad = \delta_{q,q'} \sum_{\rho,\sigma \in \mathfrak{S}_q} \prod_{k=1}^{q} \delta_{i_k, i'_{\sigma(k)}} \delta_{j_k, j'_{\rho(k)}} \mathrm{Wg}_N(\sigma\rho^{-1}). \end{aligned} \tag{2.4}$$

Before giving the expression for the moments with respect to the Haar measure $\alpha^N_{0,l}(P_1, \dots, P_l)$, let us make a simplifying assumption on our polynomials P_i.

We introduce the set $\mathcal{Y}$ of words in the letters $a_1, a_1^*, \dots, a_p, a_p^*$. We assume that for all i, P_i can be written uniquely as

$$M_{i,1} u^{\varepsilon_{i,1}} M_{i,2} u^{\varepsilon_{i,2}} \cdots M_{i,d_i} u^{\varepsilon_{i,d_i}}, \tag{2.5}$$

where $d_i \geq 1$, $\boldsymbol{\varepsilon_i} = (\varepsilon_{i,1}, \dots, \varepsilon_{i,d_i}) \in \{\pm 1\}^{d_i}$, and $M_{i,j}$ is either the empty word or an element of $\mathcal{Y}$,. We write $\mathcal{X}$ the set of such monomials. We have $\mathcal{Y} \subset \mathcal{X}$. Notice that $\mathcal{A}$ is generated by the elements of $\mathcal{X}$ up to cyclic permutation of the factors in a monomial.

The integer d_i, that we will sometime write $\deg P_i$, is the *degree* of the monomial P_i. Notice that there is no relation between the formal variables, in particular between u and u^{-1} (except for those involving ast). Therefore, the degree of (2.5) is defined by counting the total number of letters u or u^* in a word. In particular, $\deg(uu^{-1}) = 2$.

Definition 2.7. With $(P_1, \dots, P_l) \in \mathcal{X}^l$, and with the notation (2.5), we set

- $\boldsymbol{P} = (P_1, \dots, P_l)$,
- $\boldsymbol{M_P} = (M_i)_{i \in [\sum_i \deg P_i]} = (M_{1,1}, \dots, M_{1,d_1}, \dots, M_{l,1}, \dots, M_{l,d_l})$,
- $\boldsymbol{\varepsilon_P} = (\varepsilon(i))_{i \in [\sum_i \deg P_i]} = (\varepsilon_{1,1}, \dots, \varepsilon_{1,d_1}, \dots, \varepsilon_{l,1}, \dots, \varepsilon_{l,d_l})$.

Notice that we change the indices of the monomials $M_{i,j}$ and of $\boldsymbol{\varepsilon}_{i,j}$, by setting for all $1 \leq i \leq k$, $1 \leq j \leq d_i$, $M_{d_1 + \dots + d_{i-1} + j} = M_{i,j}$ and $\varepsilon(d_1 + \cdots d_{i-1} + j) = \varepsilon_{i,j}$.

We set $\deg \boldsymbol{P} = \sum_i \deg P_i$.

Furthermore, we define the permutation

$$\gamma_{\boldsymbol{P}} = (1 \dots d_1)(d_1 + 1 \dots d_2) \cdots (d_{l-1} + 1 \dots d_l). \tag{2.6}$$

In the sequel, we will consider $\boldsymbol{\varepsilon_P}$ as a function, but sometimes using vector notation for convenience. In particular, we consider the sets

$$\boldsymbol{\varepsilon}_{\boldsymbol{P}}^{-1}(+1) = \big\{ i \in [\deg \boldsymbol{P}] \colon \boldsymbol{\varepsilon_P}(i) = +1 \big\}, \quad \boldsymbol{\varepsilon}_{\boldsymbol{P}}^{-1}(-1) = \big\{ i \in [\deg \boldsymbol{P}] \colon \boldsymbol{\varepsilon_P}(i) = -1 \big\}.$$

Remark 2.8. The permutation $\gamma_{\boldsymbol{P}}$ defined by (2.6) gives a choice of labeling for the letters u and u^* in the monomials we consider. Notice that this choice is arbitrary. The cycle notation is convenient here as we are interested in traces of such monomials. The ordering of the letters in the words needs only to be specified up to cyclic permutation.

For any permutation $\sigma \in \mathfrak{S}_{\deg \boldsymbol{P}}$, we can replace $\gamma_{\boldsymbol{P}}, \boldsymbol{M}_{\boldsymbol{P}} = (M_i), \boldsymbol{\varepsilon}_{\boldsymbol{P}} = (\varepsilon(i))$ by $\gamma' = \sigma^{-1}\gamma_{\boldsymbol{P}}\sigma$, $\boldsymbol{M}' = (M_i') = (M_{\sigma(i)}), \boldsymbol{\varepsilon}' = (\varepsilon'(i)) = (\varepsilon(\sigma(i)))$. This new data describes the same polynomials. By this, we mean that if we write $\gamma' = c_1' \cdots c_l'$ the decomposition in disjoint cycles of γ', we have

$$\prod_{i=1}^{l} \mathrm{Tr}(P_i) = \prod_{i=1}^{l} \mathrm{Tr}\Big(\overrightarrow{\prod}_{j \in c_i} M_j u^{\boldsymbol{\varepsilon}'(j)}\Big).$$

Notice that the non-commutative product is only defined up to the cyclic permutation of the factors. The cyclic property of the trace ensures that the quantity on the right-hand side is well defined.

We can assume all the polynomials are of the form (2.5) without loss of generality as $\alpha_{V,l}^N$ is multilinear and satisfies the trace property

$$\alpha_{V,l}^N(P_1,\dots,P_{i-1},P_iQ,P_{i+1},\dots,P_l) = \alpha_{V,l}^N(P_1,\dots,P_{i-1},QP_i,P_{i+1},\dots,P_l),$$

as $\mathrm{Tr}(P_iQ) = \mathrm{Tr}(QP_i)$. Furthermore, if there exists i such that P_i contains no letter u nor u^{-1}, we can factor the term $\mathrm{Tr}(P_i)$ out of the moment.

The formula for the moments with respect to the Haar measure involves permutations belonging to the set $\mathfrak{S}^{(\varepsilon)}(I) \subset \mathfrak{S}(I)$ of permutations (introduced in [34]), for $\varepsilon \in \{\pm 1\}^I$.

Definition 2.9. Let $\varepsilon \in \{\pm 1\}^I$. The set $\mathfrak{S}^{(\varepsilon)}(I) \subset \mathfrak{S}(I)$ is the set of permutations $\pi \in \mathfrak{S}(I)$ such that

$$\pi(\varepsilon^{-1}(+1)) = \varepsilon^{-1}(-1).$$

Furthermore, we define $\pi^{(\varepsilon)} = \pi^2|_{\varepsilon^{-1}(+1)} \in \mathfrak{S}(\varepsilon^{-1}(+1))$.

Notice that the set $\mathfrak{S}^{(\varepsilon)}(I)$ is empty if $|\varepsilon^{-1}(+1)| \neq |\varepsilon^{-1}(-1)|$.

Example 2.10. For instance, if $\varepsilon = (+1,+1,-1,+1,-1,-1)$, then $\pi = (1\,3\,4\,6)(2\,5) \in \mathfrak{S}_6^{(\varepsilon)}$, and $\pi^{(\varepsilon)} = (1\,4)(2)$.

The notation of Definitions 2.7 and 2.9 allows us to express the moments in a compact way.

Proposition 2.11 ([34, Proposition 3.4]). *Let $\boldsymbol{P} = (P_1,\dots,P_l) \in \mathcal{X}^l$. We have*

$$\alpha_{0,l}^N(\boldsymbol{P}) = \alpha_{0,l}^N(P_1,\dots,P_l) = \sum_{\pi \in \mathfrak{S}_{\deg \boldsymbol{P}}^{(\varepsilon_{\boldsymbol{P}})}} \mathrm{Tr}_{\gamma_{\boldsymbol{P}}\pi^{-1}}(\boldsymbol{M}_{\boldsymbol{P}}) \mathrm{Wg}_N(\pi^{(\varepsilon_{\boldsymbol{P}})}). \tag{2.7}$$

2.3. Expansion of the Weingarten function

We wish to express the moments and cumulants uniquely in terms of combinatorial objects and traces. To this end, we now present a result of Novak [36] expressing the Weingarten function in terms of walks on the Cayley graph of $\mathfrak{S}_n$ generated by the transpositions.

Definition 2.12. The *value* of a transposition $\tau = (i\ j) \in \mathfrak{S}(I)$, where I is a finite subset of $\mathbb{N}^*$, is $\mathrm{val}(\tau) = \max\{i, j\}$.

Definition 2.13. Let ρ and σ be in $\mathfrak{S}(I)$, with I a finite subset of $\mathbb{N}^*$.

A *(weakly) monotone walk with r steps on* $\mathfrak{S}(I)$ from ρ to σ is a tuple $(\tau_1, \dots, \tau_r)$ of transpositions of $\mathfrak{S}(I)$ such that

- $\tau_r \cdots \tau_1 \rho = \sigma$,
- $\mathrm{val}(\tau_1) \leq \cdots \leq \mathrm{val}(\tau_r)$.

We denote the set of such walks by $\overrightarrow{W}^r(\rho, \sigma)$, and we define $\vec{w}^r(\rho, \sigma)$ as the cardinality of the set $\overrightarrow{W}^r(\rho, \sigma)$.

In this definition, we use the arrow notation $\vec{w}$ and $\overrightarrow{W}$ to emphasize the monotonicity property, as in [19].

Proposition 2.14 ([36, Theorem 3.1]). *Let $\pi \in \mathfrak{S}_q$ with $N \geq q$. We have*

$$\mathrm{Wg}_N(\pi) = \sum_{r \geq 0} \frac{(-1)^r}{N^{r+q}} \vec{w}^r(\mathrm{Id}, \pi),$$

and the series is absolutely convergent.

Propositions 2.14 and 2.11 imply the following result (recall notation from Definition 2.7).

Corollary 2.15. *Let $N \geq 1$ be an integer, $\boldsymbol{P} = (P_1, \dots, P_l) \in \mathcal{X}^l$ with $m = \deg \boldsymbol{P}/2 \leq N$. The moments admits the expansion*

$$\alpha^N_{0,l}(P_1, \dots, P_l) = \sum_{r \geq 0} \frac{(-1)^r}{N^{r+m}} \sum_{\pi \in \mathfrak{S}^{(\varepsilon_P)}_{2m}} \mathrm{Tr}_{\gamma_{\boldsymbol{P}} \pi^{-1}}(\boldsymbol{M}_{\boldsymbol{P}}) \vec{w}^r(\mathrm{Id}, \pi^{(\varepsilon_P)}).$$

Moreover, the series is absolutely convergent, uniformly on $\boldsymbol{M}_{\boldsymbol{P}}$.

Notice that if $\deg P$ is odd, there are a different number of occurrences of u and of u^*, and such moments are 0.

Proof. Starting from the expression for the moment of Proposition 2.11, we use the expansion for the Weingarten function of Proposition 2.14. Notice that this second result can only be used if $m = \deg \boldsymbol{P}/2 \leq N$, where m is the total number of letter u in the monomials $P_1, \ldots, P_l$. One of the sums is finite, we can exchange the sums and get the wanted expression.

Finally, as the matrices (A_i) have their operator norm bounded by 1, we can crudely bound the trace by

$$\left|\mathrm{Tr}_{\gamma_{\boldsymbol{P}}\pi^{-1}}(\boldsymbol{M}_{\boldsymbol{P}})\right| \leq N^{c(\gamma_{\boldsymbol{P}}\pi^{-1})} \leq N^{2m}.$$

This implies that the convergence is uniform in $\boldsymbol{M}_{\boldsymbol{P}}$. ∎

3. Maps and maps of unitary type

In this section, we introduce combinatorial objects, the so-called maps of unitary type, that will be convenient to express the moments $\alpha^N_{0,l}$, and then the cumulants $\mathcal{W}^N_{0,l}$. These maps are particular cases of the maps appearing in the Gaussian case.

3.1. Maps

First, we give a few definitions regarding maps. See [30] for more details on maps.

Definition 3.1. An *embedded graph* is a pair (Γ, S), where S is a compact topological surface and Γ is a graph (with possibly loops and multiple edges) embedded in S, so that we write $\Gamma \subset S$, such that

- the vertices of Γ are distinct points on the surface S,
- the edges of Γ are simple paths on S that can intersect only at their endpoints,
- the complement $S \setminus \Gamma$ of the graph is a disjoint union of simply connected open sets. Each of these connected components is called a *face*.

The genus of an embedded graph is the genus of the surface S.

An embedded graph will be said to be oriented if Γ is an oriented graph.

We will sometimes refer to Γ and S as the underlying graph and surface of an embedded graph. Notice that the genus of the surface is well defined: it follows from the fact that the faces are homeomorphic to disks, see [30, Section 1.3.2.2.].

Remark 3.2. In this article, the embedded graphs are in general disconnected. We will specify it when the maps we consider are connected.

Definition 3.3. Two (oriented or unoriented) embedded graphs (Γ_1, S_1) and (Γ_2, S_2) are said to be isomorphic if there is an orientation-preserving homeomorphism $h: S_1 \to S_2$ such that $h|_\Gamma$ is an isomorphism of (oriented or unoriented) graphs.

Definition 3.4. A *map* (or oriented map) is an equivalence class of embedded graphs (or oriented embedded graphs) up to isomorphism.

Figure 1. The left and right sides of a half-edge and of an ingoing half-edge.

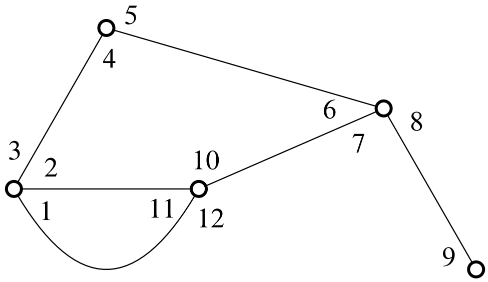

Figure 2. A map with labeled half-edges.

As the genus of a surface is a topological invariant, the genus of a map is the genus of any of its representatives.

Definition 3.5. A face will be said to be *incident* to a vertex or an edge if the vertex or the edge belongs to the boundary of the face.

It will be convenient to regard each edge of a map as being made of two half-edges. As a part of an embedded graph they can be described as follows. Each edge $e = \{v, v'\}$ (with possibly $v = v'$) can be parametrized by $\gamma_e \colon [0,1] \to S$, with $\gamma_e(0) = v$ and $\gamma_e(1) = v'$. The two half-edges that compose e are $h = \gamma_e([0, 1/2])$ and $h' = \gamma_e([1/2, 1])$. We will say that h (respectively, h') is connected to v (resp., v'). As we will be concerned only with combinatorial data, the choice of parametrization γ_e will be unimportant. When going from the vertex of the half-edge to the other end of the half-edge (connected to another half-edge), we can distinguish a left side and a right side (see Figure 1). Notice that the left and right sides are defined relative to the position of the incident vertex, and does not depend on the eventual orientation.

Defining a left side of a half-edge is crucial for the labeling procedure. Indeed, we regard each half-edge as being incident to the face on its left. A face will thus be described by the labels of the half-edges incident to it.

We label the half-edges of a map $\mathcal{C}$ from 1 to $2m$, where m is the number of edges of $\mathcal{C}$. By convention, we write each label at the left of its half-edge. See Figure 2. In an oriented map with labeled half-edges, the edges can be represented as an ordered pair of two half-edges. The first half-edge is connected to the first vertex of the edge and is said to be *outgoing*. The second half-edge is connected to the second vertex of the edge and is said to be *ingoing*.

These labels allow us to define three permutations that encode the labeled map, see [30, Section 1.3.3.].

Definition 3.6. Let $\mathcal{C}$ be a map with $2m \geq 2$ labeled half-edges. We define the three permutations $\sigma_{\mathcal{C}}, \alpha_{\mathcal{C}}, \varphi_{\mathcal{C}} \in \mathfrak{S}_{2m}$ as follows.

- Let $i \in [2m]$. The half-edge labeled by i is attached to a vertex v_i. Starting from the half-edge i and turning in the clockwise direction around v_i, the next half-edge we encounter is labeled j (possibly $i = j$). We set $\sigma_{\mathcal{C}}(i) = j$.
- Let $i \in [2m]$. The half-edge labeled by i is attached to another half-edge labeled j. We set $\alpha_{\mathcal{C}}(i) = j$.
- Let $i \in [2m]$. The half-edge labeled i has a face f_i to its left. Starting from the half-edge i, we turn in the counterclockwise direction around the face f_i. The next half-edge we encounter with f_i to its left is labeled j. We set $\varphi_{\mathcal{C}}(i) = j$.

The three permutations $\sigma_{\mathcal{C}}$, $\alpha_{\mathcal{C}}$, $\varphi_{\mathcal{C}}$ constitute the permutational model of $\mathcal{C}$. If the map has no edges then we do not define any permutational data.

The permutation $\sigma_{\mathcal{C}}$ describes how the half-edges are arranged around a vertex (we call this data "the local structure of the map"), and $\alpha_{\mathcal{C}}$ describes how to attach them. The permutation $\alpha_{\mathcal{C}}$ only depends on the underlying graph of the map. Notice that $\alpha_{\mathcal{C}}$ belongs to the set of involutions without fixed points

$$\mathcal{I}_{2m} = \left\{\alpha \in \mathfrak{S}_{2m} \colon \alpha^2 = \mathrm{Id}\ \forall i \in [2m], \alpha(i) \neq i\right\}.$$

Notice that we chose different conventions than in [30], resulting in our permutation σ being the inverse of theirs.

Example 3.7. The map $\mathcal{C}$ of Figure 2 is described by

$$\begin{aligned}
\sigma_{\mathcal{C}} &= (1\,3\,2)(4\,5)(6\,8\,7)(9)(10\,12\,11),\\
\alpha_{\mathcal{C}} &= (1\,12)(2\,11)(3\,4)(5\,6)(7\,10)(8\,9),\\
\varphi_{\mathcal{C}} &= (1\,11)(2\,10\,6\,4)(3\,5\,8\,9\,7\,12).
\end{aligned}$$

For an oriented map, we must also describe the orientation of each half-edge.

Definition 3.8. Let $\mathcal{C}$ be an oriented map with $2m$ labeled half-edges. We define the function $\boldsymbol{\varepsilon}_{\mathcal{C}} \in \{\pm 1\}^{[2m]}$ as follows. For all $i \in [2m]$, we set $\varepsilon(i) = +1$ if the half-edge labeled i is outgoing and $\varepsilon(i) = -1$ if the half-edge labeled i is ingoing.

Such an $\boldsymbol{\varepsilon}$ belongs to the set $\mathcal{E}_{2m} = \{\boldsymbol{\varepsilon} \in \{\pm 1\}^{2m} \colon \sum_{i=1}^{2m} \varepsilon(i) = 0\}$.

In the case of an oriented map, α is in the set $\mathcal{I}_{2m}^{(\varepsilon)}$ of the permutations of $\mathcal{I}_{2m}$ such that for all $i \in [2m]$, $\varepsilon(\alpha(i)) = -\varepsilon(i)$.

Lemma 3.9. [30, Proposition 1.3.16] *Let $\mathcal{C}$ be a map with labeled half-edges. We have*

$$\varphi\alpha = \sigma.$$

Conversely, we can reconstruct a map from two permutations $\sigma \in \mathfrak{S}_{2m}, \alpha \in \mathcal{I}_{2m}$. The following theorem is essentially a restatement of a result obtained in [14].

Theorem 3.10. *Let $m \geq 1$, $\sigma \in \mathfrak{S}_{2m}$ and $\mathfrak{C}(m, \sigma)$ be the set of maps with labeled half-edges $\mathcal{C}$ such that $\sigma_{\mathcal{C}} = \sigma$. Then, the mapping*

$$\begin{aligned} \mathfrak{C}(m, \sigma) &\to \mathcal{I}_{2m}, \\ \mathcal{C} &\mapsto \alpha_{\mathcal{C}} \end{aligned}$$

is a bijection.

This theorem shows that once the local structure of the map (and a labeling of the half-edges) is fixed, the map only depends on the underlying graph. We have the corresponding result for oriented maps.

Theorem 3.11. *Let $m \geq 1$, $\sigma \in \mathfrak{S}_{2m}$, $\boldsymbol{\varepsilon} \in \mathcal{E}_{2m}$ and $\mathfrak{C}(m, \boldsymbol{\varepsilon}, \sigma)$ be the set of oriented maps with $2m$ labeled half-edge $\mathcal{C}$ such that $\sigma_{\mathcal{C}} = \sigma$ and $\boldsymbol{\varepsilon}_{\mathcal{C}} = \boldsymbol{\varepsilon}$. Then,*

$$\begin{aligned} \mathfrak{C}(m, \varepsilon, \sigma) &\to \mathcal{I}_{2m}^{(\boldsymbol{\varepsilon})}, \\ \mathcal{C} &\mapsto \alpha_{\mathcal{C}} \end{aligned}$$

is a bijection.

3.2. Maps of unitary type

We have just seen how to describe a map with permutations. We now define a particular type of map, which we call map of unitary type, whose edge structure is described by a permutation $\pi \in \mathfrak{S}_{2m}^{(\varepsilon)}$ for some ε and $m \geq 1$ and a monotone walk $(\tau_1, \ldots, \tau_r) \in \overrightarrow{W}^r(\mathrm{Id}, \pi^{(\varepsilon)})$.

Definition 3.12. A vertex in an oriented map will be said to be *alternated* if when going around this vertex the half-edges connected to it are alternatively ingoing and outgoing.

Definition 3.13. Let I be a finite subset of $\mathbb{N}^*$ and $r \in \mathbb{N}$. A *map of unitary type* with labels in I with r black vertices is an oriented map with vertices colored in white or black such that

(1) there are r black vertices, which are alternated of degree 4 and numbered from 1 to r;

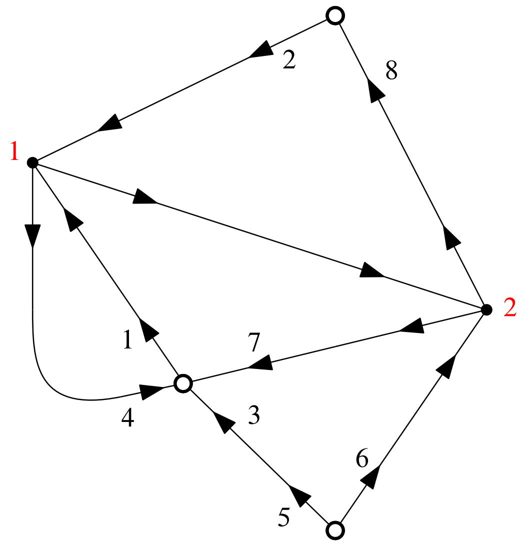

Figure 3. A unitary type map. The numbers in red (1 near the black vertex on the left, 2 near the black vertex on the right) are the numbers of the black vertices, the labels in black are the labels of the white half-edges.

(2) there are $|I|$ half-edges that are connected to white vertices. We call these half-edges *white half-edges*. Each element of I labels exactly one white half-edge;

(3) if an oriented edge connects the black vertex numbered k to the black vertex numbered l, with the orientation from k to l, then $k < l$.

See Figure 3 for an example.

Remark 3.14. There is a correspondence between a tuple $\boldsymbol{P} = (P_1, \ldots, P_l)$ of monomials and a family of maps of unitary type. The number of white vertices is l, each of them corresponds to a monomial. The white outgoing half-edges correspond to occurrences of u, the white ingoing half-edges correspond to occurrences of u^*. The black vertices correspond to steps in a walk as defined in Definition 2.13. Note however that the monotonicity condition of the walk corresponds to the increasing condition defined in Definition 3.20. This link will be described in more details in Section 3.4.

Remark 3.15. The map has oriented edges so there are as many ingoing as outgoing half-edges, of any color. The black vertices are alternated and of edgree 4, so there are as many ingoing black half-edges as black outgoing half-edges. Thus, there are as many white ingoing half-edges as white outgoing half-edges.

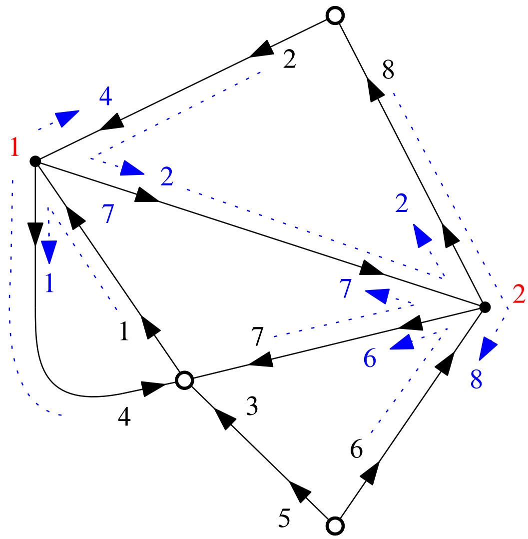

Figure 4. Procedure to assign labels to half-edges. The newly labeled half-edges are in blue (and follow the dotted arrows).

Remark 3.16. Notice that condition (3) in Definition 3.13 implies that each face is incident to at least one white vertex. Indeed, if it were not the case, there would be a face incident to only black vertices, numbered $n_1 < n_2 < \cdots < n_k$, with $n_k < n_1$, a contradiction.

Remark 3.17. The maps of unitary type are very similar to the maps introduced in [8] to describe the leading term in the asymptotics of the cumulants when $N \to \infty$. In fact, the two kinds of maps in genus 0 are related by a surgery that transforms black vertices of unitary maps into "dotted edges" of the maps from [8]. Here, we consider the non-planar cases as well.

We denote by $w_k(\mathcal{C})$ the white vertex in the unitary type map $\mathcal{C}$ connected to the half-edge labeled k. We will omit the notation $\mathcal{C}$ if there is no ambiguity.

Notice that in a map of unitary type, the half-edges connected to black vertices are not unlabeled. We now explain how to label them. Consider, in a map of unitary type, an unlabeled half-edge which we denote by h. This half-edge has a face f to its left (see Figure 1). Starting from h, we turn around the face in the clockwise direction until we encounter a labeled half-edge connected to a white vertex, which is labeled by i. We assign to h the label i. See Figures 3 and 4.

Notice that by Remark 3.16, all faces are incident to at least one white vertex, so all unlabeled half-edges can be labeled by this procedure, in a unique way.

The following lemma will be used to prove Lemma 3.24.

Lemma 3.18. *Let h be a half-edge labeled by i. There exists a unique white half-edge h' labeled by i. If h is ingoing then h' is ingoing. If h is outgoing then h' is outgoing.*

Proof. Consider an ingoing half-edge h. The existence and uniqueness of h' is a consequence of the definition. If h is a white half-edge, the statement is obvious. If not, then consider the face f to its left. Starting from h we turn around f in the clockwise direction until we reach a white vertex w. All the vertices we encounter before w are black. The black vertices are alternated so all the half-edges such that f is at their left are ingoing as well, and so is the white half-edge h' that we reach, whose label is the same as the label of h. We proceed similarly for outgoing half-edges. ■

The labels for the edges allow us to define the notion of value of a black vertex.

Definition 3.19. Consider a black vertex b. Let i and j be the labels of the two outgoing half-edges at b. The *value* of the black vertex b is $\mathrm{val}(b) = \max(i, j)$.

Definition 3.20. A map of unitary type with r black vertices $b_1, \dots, b_r$ numbered, respectively, $1, \dots, r$ is *nondecreasing* if

$$\mathrm{val}(b_1) \le \mathrm{val}(b_2) \le \cdots \le \mathrm{val}(b_r).$$

Example 3.21. Figure 4 displays an example. The labels of the black vertices are in red. The values of the black vertices s_1 and s_2 are $\mathrm{val}(s_1) = 2, \mathrm{val}(s_2) = 6$.

3.3. Permutational model

Similarly as in Section 3.1, we define a permutational model for the maps of unitary type.

Definition 3.22. Let $I \subset \mathbb{N}^*$ be finite and $r \in \mathbb{N}$. Let $\mathcal{C}$ be a map of unitary type with labels in $I \neq \emptyset$ and r black vertices.

We define $\boldsymbol{\varepsilon}_{\mathcal{C}} = (\varepsilon(i), i \in I)$ as follows. If the white half-edge labeled $i \in I$ is outgoing, we set $\varepsilon(i) = +1$, else we set $\varepsilon(i) = -1$.

We define $\gamma_{\mathcal{C}}, \pi_{\mathcal{C}}, \phi_{\mathcal{C}} \in \mathfrak{S}(I)$ and $\tau_{\mathcal{C}} = (\tau_1, \dots, \tau_r) \in \mathfrak{S}(\varepsilon_{\mathcal{C}}^{-1}(+1))^r$ as follows.

- Let $i \in I$. The white half-edge h_i, labeled i, is connected to a white vertex w_i. Starting from h_i, we turn in the clockwise direction around w_i. Let j be the label of the next half-edge connected to w_i. We set $\gamma_{\mathcal{C}}(i) = j$.
- Let $i \in I$. The white half-edge h_i labeled i is connected to another half-edge h_j, which is labeled by j. We set $\pi_{\mathcal{C}}(i) = j$.
- Let $i \in I$. The white half-edge labeled i has a face f_i to its left. Starting from the half-edge i, we turn in the counterclockwise direction around the face f_i. The next white half-edge with f_i on its left we encounter is labeled j. We set $\phi_{\mathcal{C}}(i) = j$.

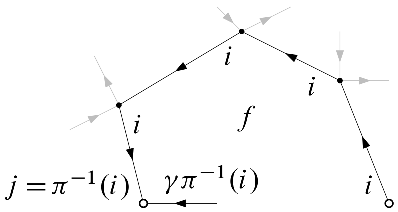

Figure 5. Proof of Lemma 3.25.

- Let b_l be the black vertex numbered l. The outgoing half-edges that are connected to it are labeled by i and j. We set $\tau_l = (i\ j)$.

The permutations $\gamma_{\mathcal{C}}, \pi_{\mathcal{C}}, \phi_{\mathcal{C}}$ are the counterparts for maps of unitary type of the permutations $\sigma_{\mathcal{C}}, \alpha_{\mathcal{C}}, \varphi_{\mathcal{C}}$ defined in Definition 3.6.

Example 3.23. For the map in Figure 4, we have $r = 2$ and

$$\begin{aligned}
\gamma_{\mathcal{C}} &= (1\,7\,3\,4)(5\,6)(2\,8),\\
\varepsilon_{\mathcal{C}} &= (+1, +1, -1, -1, +1, +1, -1, -1),\\
\tau_1 &= (1\,2), \quad \tau_2 = (2\,6),\\
\pi_{\mathcal{C}} &= (1\,7\,6\,8\,2\,4)(3\,5),\\
\phi_{\mathcal{C}} &= (1)(2)(3\,6)(4\,8\,5)(7).
\end{aligned}$$

Lemma 3.24. *The permutation $\pi_{\mathcal{C}}$ belongs to $\mathfrak{S}^{(\varepsilon_{\mathcal{C}})}(I)$, defined in Definition 2.9.*

Proof. An edge consists of an outgoing half-edge h attached to an ingoing half-edge h'. Assume that h is white. Let i be the label of h and j be the label of h'. We have $\pi(i) = j$. By Lemma 3.18, j is the label of a white ingoing half-edge. Thus, $\varepsilon(i) = -1$ and $\varepsilon(j) = +1$. We proceed similarly if h' is white. ■

We have the following counterpart of Lemma 3.9.

Lemma 3.25. *For a unitary type map $\mathcal{C}$, we have $\gamma_{\mathcal{C}}\pi_{\mathcal{C}}^{-1} = \phi_{\mathcal{C}}$.*

Proof. Let $i \in I$ be the label of a white outgoing half-edge, and f the face at the left of the half-edge. Starting from the half-edge labeled i, we follow the boundary of the face until we encounter a white vertex. The last half-edge we traversed, which was ingoing, was labeled j. This half-edge is connected to a outgoing half-edge labeled i. By definition, we thus have $\pi_{\mathcal{C}}(j) = i$. The next labeled half-edge when going around f in the counterclockwise order is the half-edge following the half-edge j when turning in the clockwise direction around the white vertex. This next half-edge is thus labeled $\gamma_{\mathcal{C}}(j) = \gamma_{\mathcal{C}}\pi_{\mathcal{C}}^{-1}(i)$, see Figure 5.

The proof is identical if i is the label of an ingoing half-edge. ■

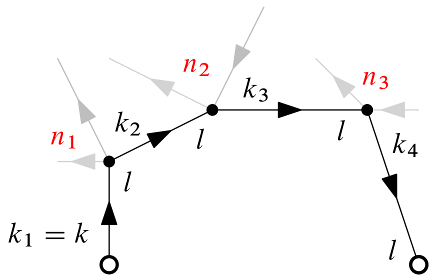

Figure 6. Chain of edges around the face f.

Proposition 3.26. *Let I be a finite subset of $\mathbb{N}^*$, $r \in \mathbb{N}^*$, $\gamma \in \mathfrak{S}(I)$, and $\varepsilon \in \{\pm 1\}^I$. Let $\mathcal{C}$ be a unitary type map with set of labels I and with r black vertices such that $\gamma_{\mathcal{C}} = \gamma$ and $\varepsilon_{\mathcal{C}} = \varepsilon$, and let $\tau_{\mathcal{C}} = (\tau_1, \dots, \tau_r)$. Then, $\tau_r \cdots \tau_1 = \pi_{\mathcal{C}}^{(\varepsilon)}$.*

Proof. Let $k \in I$ be the label of a white outgoing half-edge connected to a vertex $w_k = u_0$. Let f be the face at its right. We construct a path starting from the half-edge labeled k as follows, see also Figure 6. Consider the edge $e_1 = (u_0, u_1)$ of which the half-edge labeled k is part. If u_1 is white then for all $1 \le j \le r$, $\tau_j(k) = k = \pi_{\mathcal{C}}^{(\varepsilon)}(k)$.

If u_1 is black, we can find vertices $u_2, u_3, \dots, u_{p+1}$ such that $u_2, \dots, u_p$ are black and u_{p+1} is white, and (u_j, u_{j+1}) follows (u_{j-1}, u_j) when going around the vertex u_j in the counterclockwise order. Notice that these edges are all part of the boundary of f.

Let $n_1, n_2, \dots, n_p$ be the labels of the black vertices $u_1, \dots, u_p$, and k_j, $1 \le j \le p+1$ be the labels of the outgoing half-edges edges (connected to u_{j-1}) in (u_{j-1}, u_j). Notice that $1 \le n_1, \dots, n_p \le r$ as black vertices have labels in $[r]$. By construction, we have $\tau_{n_j}(k_{j-1}) = k_j$.

We have $\tau_{n_p}\tau_{n_{p-1}} \cdots \tau_{n_1}(k) = k_p$, so the labels of the black ingoing vertices incident to f are all equal to $l = \pi_{\mathcal{C}}(k)$, by construction of $\pi_{\mathcal{C}}$. We have $\pi_{\mathcal{C}}(k) = l = \pi_{\mathcal{C}}^{-1}(k_p)$. Thus, $\tau_{n_p}\tau_{n_{p-1}} \cdots \tau_{n_1}(k) = \pi^{(\varepsilon)}(k)$.

Assume now that $\tau_r \cdots \tau_1(k) \ne \tau_{n_p}\tau_{n_{p-1}} \cdots \tau_{n_1}(k)$. Let j be the minimal index such that there exists p' satisfying $n_{p'} \le j < n_{p'+1}$ (with the convention $n_{p+1} = r+1$) and $\tau_j \cdots \tau_1(k) \ne \tau_{n_{p'}} \cdots \tau_{n_1}(k)$. The index j is minimal so $j > n_{p'}$ (else we would have a contradiction as $\tau_{j-1} \cdots \tau_1(k) = \tau_{n_{p'-1}} \cdots \tau_{n_1}(k)$). We have $k_{p'} = \tau_{j-1} \cdots \tau_1(k) = \tau_{n_{p'}} \cdots \tau_{n_1}(k)$. By construction, all the half-edges labeled $k_{p'}$ are on the boundary of a same face f', and they follow each other. We have just seen that there is such an half-edge in the edge between $u_{p'}$ and $u_{p'+1}$. The fact that $\tau_j(k_{p'}) \ne k_{p'}$ implies that there is an half-edge labeled $k_{p'}$ that is connected to the jth black vertex. However, this edge must be before (when going around the face f') or after the edge $(u_{p'}, u_{p'+1})$ in the boundary of f'. This contradicts the fact that if there is an edge going from a black vertex i to a black vertex labeled j we have $i < j$, as $n_{p'} < j < n_{p'+1}$. ■

Definition 3.27. We denote by $\mathfrak{C}^r(I, \varepsilon, \gamma)$ the set of nondecreasing unitary type maps $\mathcal{C}$ with set of labels I and with r black vertices such that $\gamma_{\mathcal{C}} = \gamma$ and $\varepsilon_{\mathcal{C}} = \varepsilon$.

Similarly, we denote by $\mathfrak{C}(g, I, \varepsilon, \gamma)$ the set of nondecreasing unitary type maps $\mathcal{C}$ with set of labels I and with genus g such that $\gamma_{\mathcal{C}} = \gamma$ and $\varepsilon_{\mathcal{C}} = \varepsilon$.

Theorem 3.28. *Let I be a finite subset of the positive integers, $r \in \mathbb{N}^*$, $\varepsilon \in \{\pm 1\}^I$ and $\gamma \in \mathfrak{S}(I)$.*

The mapping

$$\begin{aligned} \mathfrak{C}^r(I, \varepsilon, \gamma) &\to \bigcup_{\pi \in \mathfrak{S}^{(\varepsilon)}(I)} \{\pi\} \times \overrightarrow{\mathcal{W}}^r(\mathrm{Id}, \pi^{(\varepsilon)}), \\ \mathcal{C} &\mapsto (\pi_{\mathcal{C}}, \tau_{\mathcal{C}}) \end{aligned}$$

is a bijection.

Proof. Lemma 3.24 and Proposition 3.26 show that this map has values in

$$\bigcup_{\pi \in \mathfrak{S}^{(\varepsilon)}(I)} \{\pi\} \times \overrightarrow{\mathcal{W}}^r(\mathrm{Id}, \pi^{(\varepsilon)}).$$

We now construct an inverse mapping. To do so, we explicitly construct a map corresponding to permutations π and $\boldsymbol{\tau} = (\tau_1, \dots, \tau_r)$. By Theorem 3.10, it suffices to construct from π and $\boldsymbol{\tau}$ the incidence relation of the underlying graph.

To this end, we introduce the set whose elements represent the half-edges $\widetilde{I} = \{h_i : i \in I\} \cup \bigcup_{j=1}^r \{h_{j,1}, h_{j,2}, h_{j,3}, h_{j,4}\}$. We can split this set into the set of ingoing and outgoing edges $\widetilde{I} = \widetilde{I}_{\text{in}} \cup \widetilde{I}_{\text{out}}$. We have $\widetilde{I}_{\text{out}} = \{h_i : i \in I, \varepsilon(i) = +1\} \cup \bigcup_{j=1}^r \{h_{j,2}, h_{j,4}\}$. The elements $h_{j,k}$ represent the half-edges of the black vertices of the map we are going to construct, and the elements h_i represent the half-edges of the white vertices. We are going to define a labeling function $L : \widetilde{I} \to I$. We set for all $i \in I$, $L(h_i) = i$. The function L is constructed by induction. At the initial step, it is only defined for the white half-edges. We then define it for the black half-edges of the black vertex i at step i.

To construct a map, we use Theorem 3.10. We define two permutations $\sigma, \alpha \in \mathfrak{S}(\widetilde{I})$ as follows. We define $\widetilde{\gamma} \in \mathfrak{S}(\widetilde{I})$ by $\widetilde{\gamma}(h_i) = h_{\gamma(i)}$ and the identity otherwise. We set

$$\sigma = \widetilde{\gamma}(h_{1,1}\, h_{1,2}\, h_{1,3}\, h_{1,4}) \cdots (h_{r,1}\, h_{r,2}\, h_{r,3}\, h_{r,4}).$$

The permutation α is given by the following algorithm. Let $\pi \in \mathfrak{S}^{(\varepsilon)}(I)$, and $\tau = (\tau_1, \dots, \tau_r) \in \overrightarrow{\mathcal{W}}^r(\mathrm{Id}, \pi^{(\varepsilon)})$. We consider first the permutation $\tau_1 = (i_1, j_1)$, with

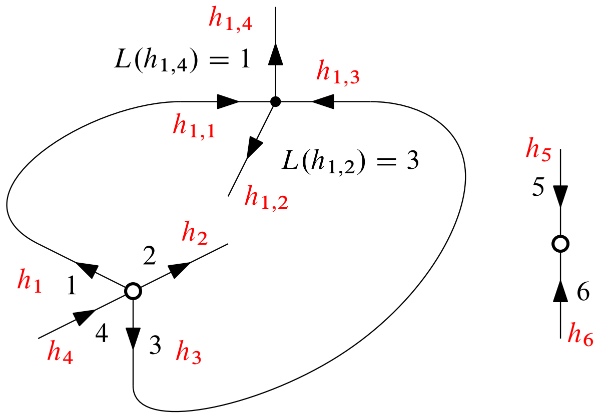

Figure 7. First step of the construction of the permutation α from one transposition $\tau_1 = (1\,3)$, represented as a map. The name of the half-edges are in red ($h_1, \ldots, h_6$ near the white vertices and $h_{1,1}, \ldots, h_{1,4}$ near the black vertices), and the labels are in black. Here, $\alpha_1 = (h_1\, h_{1,1})(h_3\, h_{1,3})$.

$i_1 < j_1$. We set $\alpha_1 = (h_{i_1}\, h_{1,1})(h_{j_1}, h_{1,3})$. We set $L(h_{1,2}) = j_1$ and $L(h_{1,4}) = i_1$. In terms of maps, this procedure corresponds to connecting two edges to a same black vertex, see Figure 7.

We proceed similarly to construct the black vertices labeled $2, 3, \ldots, r$ from the transpositions $\tau_2, \ldots, \tau_r$. At the kth step, we consider the transposition $\tau_k = (i_k\, j_k)$, with $i_k < j_k$. There is only one half-edge h (respectively, h') in $\widetilde{I}_{\text{out}}$ such that $L(h) = i_k$ and $\alpha_{k-1}(h) = h$ (respectively, $L(h') = j_k$ and $\alpha_{k-1}(h') = h'$). We set $\alpha_k = \alpha_{k-1}(h\, h_{k,1})(h'\, h_{k,3})$.

Finally, we connect each remaining outgoing half-edge labeled i to the ingoing half-edge $\pi^{-1}(i)$. For all $i \in I$, there is a unique h such that $\alpha_r(h) = h$ and $L(h) = i$. We set $\alpha_{r+1,i} = (h\, h_{\pi^{-1}(i)})$ and define $\alpha = \alpha_r \prod_{i \in I} \alpha_{r+1,i}$. We define $\widetilde{\varepsilon} \in \{\pm 1\}^{\widetilde{I}}$ by $\widetilde{\varepsilon}(i) = +1$ if $i \in \widetilde{I}_{\text{out}}$ and $\widetilde{\varepsilon}(i) = -1$ otherwise.

Theorem 3.10 implies that given σ, α, and $\widetilde{\varepsilon}$, we construct a unique map $\widetilde{\mathcal{C}}$. By construction, the resulting map is of unitary type: the vertices attached to the half-edges h_i are the white vertices and the other are the black vertices. The black vertex attached to the half-edges $h_{j,k}$ is numbered j. The map is constructed such that $\pi_{\mathcal{C}} = \pi_{\widetilde{\mathcal{C}}}$ and $\tau_{\mathcal{C}} = \tau_{\widetilde{\mathcal{C}}}$.

Furthermore, the map is nondecreasing (recall Definition 3.20) as the tuple $\boldsymbol{\tau}$ is a monotone walk.

We have constructed a right inverse, so the map $\mathcal{C} \mapsto (\pi_{\mathcal{C}}, \tau_{\mathcal{C}})$ is surjective. We now show that this map is injective. We show that the incidence relation of a map of unitary type $\mathcal{C}$ is determined by the permutations. Indeed, consider a map of unitary type described by π and $\tau = (\tau_1, \dots, \tau_r)$, and an outgoing half-edge h_i labeled i. There are four cases.

- If h_i is a white half-edge such that for all j we have $\tau_j(i) = i$, then h_i is necessarily attached to the white half-edge labeled $\pi^{-1}(i)$.
- If h_i is a white half-edge and there exists k such that $\tau_k(i) \neq i$, then h_i is necessarily connected to the k'th black vertex, where $k' = \min\{k\colon \tau_k(i) \neq i\}$.
- If h_i is a half-edge connected to the kth black vertex and for all $l > k$ $\tau_l(i) = i$, then h_i is necessarily attached to a white half-edge labeled $\pi^{-1}(i)$
- If h_i is a half-edge connected to the kth black vertex and l is the smallest integer such that $l > k$ and $\tau_l(i) \neq i$, then h_i is necessarily attached to the l-black vertex.

Thus, two maps of unitary type in $\mathfrak{C}^r(I, \varepsilon, \gamma)$ described by the same permutations π and τ have necessarily the same edges, i.e., are identical. ∎

For a tuple of permutations, $(\sigma_1, \dots, \sigma_k) \in \mathfrak{S}(I)^k$, we denote the subgroup of $\mathfrak{S}(I)$ they generate by

$$\langle \sigma_1, \dots, \sigma_k \rangle.$$

We can associate to the triplet $(\gamma_{\mathcal{C}}, \pi_{\mathcal{C}}, \tau_{\mathcal{C}})$ the group

$$G(\mathcal{C}) = \langle \gamma_{\mathcal{C}}, \pi_{\mathcal{C}}, \tau_1, \dots, \tau_r \rangle, \tag{3.1}$$

where $\tau_{\mathcal{C}} = (\tau_1, \dots, \tau_r)$.

Proposition 3.29. *A unitary type map $\mathcal{C}$ with set of labels I is connected if and only if the group $G(\mathcal{C})$ defined by (3.1) acts transitively on I.*

Proof. First, assume that $\mathcal{C}$ is connected. Let $i, j \in I$. There is a path ρ (made up by vertices and edges) connecting the white vertices w_i and w_j. First, let us assume that ρ contains only black vertices, except for its boundary which is made up of w_i and w_j. The path encounters the black vertices $n_1, \dots, n_p$, the labels on the left of the edges that constitute ρ are $k_1, \dots, k_{p+1}$. The first and last edges are connected to w_i and w_j so $k_1 = \gamma_{\mathcal{C}}^{m_1}(i)$ and $k_{p+1} = \gamma_{\mathcal{C}}^{m_2}\pi_{\mathcal{C}}^{m_3}(j)$ for some integers m_1, m_2, m_3.

Let $1 \leq i \leq p$. If $k_i = k_{i+1}$, we set $\sigma_i = \mathrm{Id}$, and if $\tau_{n_i}(k_i) = k_{i+1}$, we set $\sigma_i = \tau_{n_i}$, see Figure 8. Those are the only two possibilities as the half-edges connected to a black vertex labeled k, with $\tau_k = (u\ v)$ can only be labeled by u or v.

Thus, we have proved that there is $\sigma_\rho = \pi_{\mathcal{C}}^{-m_3}\gamma_{\mathcal{C}}^{-m_2}\sigma_p \cdots \sigma_1\gamma_{\mathcal{C}}^{-m_1} \in G(\mathcal{C})$ such that $\sigma_\rho(i) = j$.

In general, any path connecting w_i and w_j can be written as the concatenation of paths with only black vertices in their interiors, we can thus construct by composition a permutation in $G(\mathcal{C})$ that sends i to j. Thus, $G(\mathcal{C})$ is transitive.

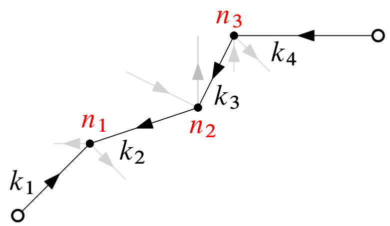

Figure 8. Three situations for σ_i. We set $\sigma_1 = \tau_{n_1}$, $\sigma_2 = \mathrm{Id}$, and $\sigma_3 = \tau_{n_3}$. The black vertices are labeled n_1, n_2, n_3 in red. In grey are the two half-edges that do not play a role for each black vertex.

Conversely, if $G(\mathscr{C})$ is transitive, for any $k, l \in I$, there exists $\sigma \in G(\mathscr{C})$ such that $\sigma(k) = l$. We can write $\sigma = \sigma_p \cdots \sigma_1$, with for all i, σ_i is one of $\gamma_{\mathscr{C}}, \pi_{\mathscr{C}}^{-1}, \tau_1, \dots, \tau_r$. We use this to construct a path connecting v_k to v_l. For all i, we attach to σ_i a path ρ_i starting from a half-edge labeled k_i. We set $k_1 = k$, and we will show that $k_{p+1} = l$.

- If $\sigma_i = \gamma_{\mathscr{C}}$, ρ_i is the empty path, and $k_{i+1} = \gamma_{\mathscr{C}}(k_i)$.
- If $\sigma_i = \pi_{\mathscr{C}}^{-1}$, ρ_i is the path connecting the half-edge k_i to the half-edge $\pi^{-1}(k_i)$. Such a path exists by the propagation of labels procedure. We set $k_{i+1} = \pi^{-1}(k_i)$.
- If $\sigma_i = \tau_{n_i}$, for some n_i, and $\tau_{n_i}(k_i) = k_i$, then ρ_i is the empty path and $k_{i+1} = k_i$.
- If $\sigma_i = \tau_{n_i}$, for some n_i, and $\tau_{n_i}(k_i) \neq k_i$, then we set $k_{i+1} = \tau_{n_i}(k_i)$. Both k_i and k_{i+1} are labels of outgoing half-edges. We set ρ_i to be the path that starts from the half-edge k_i, follows the half-edges labeled k_i until it reaches the black vertex n_i, and then follows the half-edges labeled k_{i+1} until the half-edge k_{i+1}, and the vertex $w_{k_{i+1}}$.

We have constructed a path going from the half-edge i to the half-edge $k_{p+1} = \sigma(i) = j$, as wanted. ■

3.4. Expression of the moments in terms of maps of unitary type

Theorem 3.28 allows us to rewrite the expression for the moments given in Corollary 2.15 (see Definition 2.7 for relevant notation).

Corollary 3.30. *Let $N \geq 1$ be an integer, $\boldsymbol{P} = (P_1, \dots, P_l) \in \mathcal{X}^l$ be monomials with $m = \frac{1}{2} \deg \boldsymbol{P} \leq N$. The moments under the Haar measure μ_0^N (see Definition 2.3) admit the following expansion:*

$$\alpha_{0,l}^N(P_1, \dots, P_l) = \sum_{r \geq 0} \frac{(-1)^r}{N^{r+m}} \sum_{\mathscr{C} \in \mathfrak{C}^r([2m], \boldsymbol{\varepsilon_P}, \gamma_{\boldsymbol{P}})} \operatorname{Tr}_{\phi_{\mathscr{C}}}(\boldsymbol{M_P}).$$

Furthermore, the series is absolutely convergent.

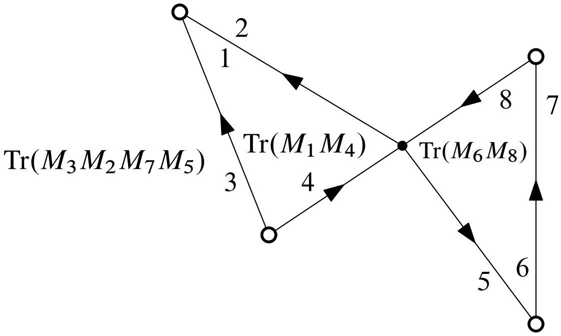

Figure 9. A map with its weights. This weighted map gives (up to a sign) a contribution from the sum $\alpha_{0,4}^{(0),N}(M_1u^{-1}M_2u^{-1}, M_3uM_4u, M_5u^{-1}M_6u, M_7u^{-1}M_8u)$.

The weights $\mathrm{Tr}_{\phi_{\mathcal{C}}}(\boldsymbol{M_P})$ can be interpreted as product of weights given by the faces of the map $\mathcal{C}$; see Figure 9.

Proof of Corollary 3.30. Recall the expression of Corollary 2.15:

$$\alpha_{0,l}^N(P_1,\ldots,P_l) = \sum_{r\geq 0}\frac{(-1)^r}{N^{r+m}}\sum_{\pi\in\mathfrak{S}_{2m}^{(\varepsilon_{\boldsymbol{P}})}}\mathrm{Tr}_{\gamma_{\boldsymbol{P}}\pi^{-1}}(\boldsymbol{M_P})\vec{w}^r(\mathrm{Id},\pi^{(\varepsilon_{\boldsymbol{P}})}).$$

By definition of $\vec{w}^r(\mathrm{Id},\pi^{(\varepsilon_{\boldsymbol{P}})})$, we can rewrite this as

$$\begin{aligned}\alpha_{0,l}^N(P_1,\ldots,P_l) &= \sum_{r\geq 0}\frac{(-1)^r}{N^{r+m}}\sum_{\substack{\pi\in\mathfrak{S}_{2m}^{(\varepsilon_{\boldsymbol{P}})}\\(\tau_1,\ldots,\tau_r)\in\overrightarrow{\mathcal{W}}^r(\mathrm{Id},\pi^{(\varepsilon_{\boldsymbol{P}})})}}\mathrm{Tr}_{\gamma_{\boldsymbol{P}}\pi^{-1}}(\boldsymbol{M_P})\\ &= \sum_{r\geq 0}\frac{(-1)^r}{N^{r+m}}\sum_{\mathcal{C}\in\mathfrak{C}^r([2m],\boldsymbol{\varepsilon_P},\gamma_{\boldsymbol{P}})}\mathrm{Tr}_{\gamma_{\boldsymbol{P}}\pi_{\mathcal{C}}^{-1}}(\boldsymbol{M_P}),\end{aligned}$$

where we used Theorem 3.28 in the last line.

We get the result by using Lemma 3.25, which gives $\gamma_{\boldsymbol{P}}\pi_{\mathcal{C}}^{-1} = \phi_{\mathcal{C}}$. ∎

Definition 2.1 and Corollary 3.30 allow us to express the cumulants in terms of maps of unitary types. We deduce the following lemma.

Lemma 3.31. *Let $N\geq 1$ be an integer $\boldsymbol{P} = (P_1,\ldots,P_l)\in\mathcal{X}^l$ be monomials with $m = \frac{1}{2}\deg\boldsymbol{P}$. The cumulants admit the expansion*

$$\mathcal{W}_{0,l}^N(P_1,\ldots,P_l) = \sum_{r\geq 0}\frac{(-1)^r}{N^{r+m}}\sum_{\substack{\mathcal{C}\in\mathfrak{C}^r([2m],\boldsymbol{\varepsilon_P},\gamma_{\boldsymbol{P}})\\ \mathcal{C}\text{ is connected}}}\mathrm{Tr}_{\phi_{\mathcal{C}}}(\boldsymbol{M_P}).$$

Furthermore, the series is absolutely convergent.

Proof. We show the formula by induction using Corollary 3.30. Notice first that when $l = 1$, $\alpha^N_{0,1}(P_1) = \mathcal{W}^N_{0,1}(P_1)$ and the maps in $\mathfrak{C}^r([2m], \boldsymbol{\varepsilon}_{\boldsymbol{P}}, \gamma_{\boldsymbol{P}})$ are connected.

Then, we notice that a map can be decomposed into its connected components. This decomposition gives a partition of the set of labels of half-edges. Each block contains the labels appearing in one connected component. Using Definitions 2.1 and 2.3, we obtain that

$$\begin{aligned}
\mathcal{W}^N_{0,l}(P_1,\dots,P_l) &= \alpha^N_{0,l}(P_1,\dots,P_l) - \sum_{\substack{\Pi\in\mathcal{P}([l])\\ |\Pi|\geq 2}} \prod_{B\in\Pi} \mathcal{W}^N_{0,|B|}(P_i, i\in B)\\
&= \sum_{r\geq 0}\frac{(-1)^r}{N^{r+m}} \sum_{\mathcal{C}\in\mathfrak{C}^r([2m],\boldsymbol{\varepsilon}_{\boldsymbol{P}},\gamma_{\boldsymbol{P}})} \mathrm{Tr}_{\phi_{\mathcal{C}}}(\boldsymbol{M}_{\boldsymbol{P}})\\
&\quad - \sum_{r\geq 0}\frac{(-1)^r}{N^{r+m}} \sum_{\substack{\mathcal{C}\in\mathfrak{C}^r([2m],\boldsymbol{\varepsilon}_{\boldsymbol{P}},\gamma_{\boldsymbol{P}})\\ \mathcal{C}\text{ has at least 2 connected components}}} \mathrm{Tr}_{\phi_{\mathcal{C}}}(\boldsymbol{M}_{\boldsymbol{P}}).
\end{aligned}$$

Hence, we get the result. ∎

Remark 3.32. The formulae imply that we can express moments and cumulants with respect to the Haar measure as a weighted sum over maps. The maps are the non-decreasing maps of unitary type whose local structure (i.e., how the half-edges are attached to the vertices, but not how the half-edges are attached together) is determined by $\gamma_{\boldsymbol{P}}^{-1}$ and $\boldsymbol{\varepsilon}_{\boldsymbol{P}}$. To each face is associated a weight, which is the trace of a certain word in the matrices of $\boldsymbol{M}_{\boldsymbol{P}}$, times a sign.

A topological expansion for the Haar measure. We now rewrite Lemma 3.31 as a sum over the genus g of the maps rather than on the number of black vertices r. We will see that this gives us an expansion in powers of $\frac{1}{N^2}$. We first recall Euler's formula

$$2 - 2g(\mathcal{C}) = V(\mathcal{C}) - E(\mathcal{C}) + F(\mathcal{C}), \tag{3.2}$$

where $V(\mathcal{C})$, $E(\mathcal{C})$ and $F(\mathcal{C})$ are the number of vertices, edges and faces of a map $\mathcal{C}$, and $g(\mathcal{C})$ is its genus. In the case of a map of unitary type labeled by a set of $2m$ integers, and with r black vertices, we have

- $c(\gamma_{\mathcal{C}})$ white and r black vertices,
- $2m$ white half-edges and $4r$ half-edges out of black vertices, for a total of $m + 2r$ edges,
- $c(\phi_{\mathcal{C}})$ faces (see Definition 3.1).

Thus, we get

$$2 - 2g(\mathcal{C}) = (c(\gamma_{\mathcal{C}}) + r) - (m + 2r) + c(\phi_{\mathcal{C}}) = c(\gamma_{\mathcal{C}}) + c(\phi_{\mathcal{C}}) - m - r. \tag{3.3}$$

A change of variable in the sum of Lemma 3.31, and the identities $l = c(\gamma_{\boldsymbol{P}})$ and $\mathrm{Tr}_{\phi_{\mathcal{C}}}(\boldsymbol{M_P}) = N^{c(\phi_{\mathcal{C}})}\,\mathrm{tr}_{\phi_{\mathcal{C}}}(\boldsymbol{M_P})$ give the following proposition.

Proposition 3.33. *Let $N \geq 1$ be an integer, $\boldsymbol{P} = (P_1, \dots, P_l) \in \mathcal{X}^l$ be monomials with $m = \frac{1}{2}\deg \boldsymbol{P}$. The cumulants admit the expansion*

$$\mathcal{W}_{0,l}^{N}(P_1, \dots, P_l) = N^{2-l}(-1)^{m+l} \sum_{g \geq 0} \frac{1}{N^{2g}} \sum_{\substack{\mathcal{C} \in \mathfrak{C}(g,[2m],\boldsymbol{\varepsilon_P},\gamma_{\boldsymbol{P}}) \\ \mathcal{C} \text{ is connected}}} (-1)^{c(\phi_{\mathcal{C}})}\,\mathrm{tr}_{\phi_{\mathcal{C}}}(\boldsymbol{M_P}).$$

Furthermore, the series is absolutely convergent.

Notice that this expansion is in terms of the normalized trace $\mathrm{tr} = \frac{1}{N}\mathrm{Tr}$. The factors with the trace are bounded by 1 if we assume $\|A_i^N\| \leq 1$ for all $1 \leq i \leq N$ and $N \geq 1$.

Remark 3.34. The sum in Proposition 3.33 is in general not finite. Indeed, even for $l = 1$ and $P_1 = AUBU^*$, the sum contains terms of arbitrary genus. They appear for instance because of the factorization of the identity $\mathrm{Id} = (1\,2)^{2k}$ for all $k \geq 0$.

Definition 3.35. Let $N \geq 1$ be an integer, $\boldsymbol{P} = (P_1, \dots, P_l) \in \mathcal{X}^l$ be monomials with $m = \frac{1}{2}\deg \boldsymbol{P}$. The term of order $2g$ in the expansion of the cumulant is denoted by

$$\mathcal{M}_{0,l}^{(g),N}(P_1, \dots, P_l) = (-1)^{m+l} \sum_{\substack{\mathcal{C} \in \mathfrak{C}(g,[2m],\boldsymbol{\varepsilon_P},\gamma_{\boldsymbol{P}}) \\ \mathcal{C} \text{ is connected}}} (-1)^{c(\phi_{\mathcal{C}})}\,\mathrm{tr}_{\phi_{\mathcal{C}}}(\boldsymbol{M_P}).$$

We extend this definition to all monomials in $\mathcal{A}$ by setting for $P_1, \dots, P_l \in \mathcal{X}$ and M a word in $a_1, a_1^*, \dots, a_p, a_p^*$,

$$\begin{aligned} &\mathcal{M}_{0,l}^{(g),N}(P_1, \dots, P_{i-1}, P_i M, P_{i+1}, \dots, P_l) \\ &\quad = \mathcal{M}_{0,l}^{(g),N}(P_1, \dots, P_{i-1}, M P_i, P_{i+1}, \dots, P_l) \end{aligned}$$

for all $1 \leq i \leq l$.

The last property is enforced so that $\mathcal{M}_{0,l}^{(g),N}$ has a property of cyclicity, as does the trace.

Stationary distribution of the $(A_i^N)_{1 \leq i \leq p}$. Let us consider a particular choice for the sequence of matrices $(A_i^N)_{1 \leq i \leq p, N \geq 1}$. Fix a family of p matrices of fixed size $M \times M$, $(A_i^M)_{1 \leq i \leq p}$, and consider the sequence of matrices $(A_i^{qM})_{1 \leq i \leq p, q \geq 1}$, where A_i^{qM} is the block-diagonal matrix with q blocks, whose blocks are A_i^M. When considering the sums of maps for $N = qM$, the traces $\mathrm{tr}_\phi(\boldsymbol{M})$ no longer depend on q or $N = qM$.

In the case of zero potential ($V = 0$), by Proposition 3.33, the renormalized cumulant $\widetilde{W}^N_{0,1}$ converges with limit

$$\lim_{q\to\infty} \widetilde{W}^{qM}_{0,1}(P) = \mathcal{M}^{(0),M}_{0,1}(P)$$

for $P \in \mathcal{A}$. This fact allows us to prove the following lemma.

Lemma 3.36. *Fix $N \in \mathbb{N}^*$. Assume that $\|A_i^N\| \le 1$ for all $1 \le i \le p$. Let $P \in \mathcal{X}$ be a monomial. We have for all choices of $(A_i^N)_{1\le i\le p}$ that*

$$\left|\mathcal{M}^{(0),N}_{0,1}(P)\right| \le 1.$$

Proof. By the previous remark, we have for the choice of with the $(A_i^N)_{1\le i\le p}$ block diagonal as above,

$$\left|\mathcal{M}^{(0),M}_{0,1}(P)\right| = \lim_{N\to\infty}\left|\widetilde{W}^N_{0,1}(P)\right| = \lim_{N\to\infty}\mathbb{E}[\operatorname{tr}(P)] \le 1,$$

as $\|P\| \le 1$. ∎

More generally, with the $(A_i^N)_{1\le i\le p}$ block diagonal as above, we have

$$\widetilde{W}^{qM}_{0,l}(\boldsymbol{P}) = (-1)^{m+l}\sum_{g\ge 0}\frac{1}{(qM)^{2g}}\sum_{\substack{\mathcal{C}\in\mathfrak{C}(g,[2m],\boldsymbol{\varepsilon_P},\boldsymbol{\gamma_P})\\ \mathcal{C}\text{ is connected}}}(-1)^{c(\phi_{\mathcal{C}})}\operatorname{tr}_{\phi_{\mathcal{C}}}(\boldsymbol{M_P}),$$

where $\operatorname{tr}_{\phi_{\mathcal{C}}}(\boldsymbol{M_P})$ does not depend on q. This implies the following lemma.

Lemma 3.37. *For all $N \ge 1$, $g \ge 0$, $l \ge 1$, and $\boldsymbol{P} \in \mathcal{X}^l$, we have the following properties.*

(i) *(Traciality) For all $Q \in \mathcal{X}$,*

$$\mathcal{M}^{(g),N}_{0,l}(P_1,\dots,P_{l-1},P_lQ) = \mathcal{M}^{(g),N}_{0,l}(P_1,\dots,P_{l-1},QP_l).$$

(ii) *(Symmetry) For all permutations $\sigma \in \mathfrak{S}_l$,*

$$\mathcal{M}^{(g),N}_{0,l}(P_1,\dots,P_l) = \mathcal{M}^{(g),N}_{0,l}(P_{\sigma(1)},\dots,P_{\sigma(l)}).$$

(iii) *(Simplification) We have*

$$\mathcal{M}^{(g),N}_{0,l}(P_1,\dots,P_{l-1},u^*P_lu) = \mathcal{M}^{(g),N}_{0,l}(P_1,\dots,P_l).$$

(iv) *(Conjugation) We have*

$$\mathcal{M}^{(g),N}_{0,l}(P_1^*,\dots,P_l^*) = \overline{\mathcal{M}^{(g),N}_{0,l}(P_1,\dots,P_l)}.$$

Proof. Consider the series

$$G(\hbar) = (-1)^{m+l} \sum_{g\geq 0} \hbar^{2g} \sum_{\substack{\mathcal{C}\in\mathfrak{C}(g,[2m],\boldsymbol{\varepsilon}_{\boldsymbol{P}},\gamma_{\boldsymbol{P}})\\ \mathcal{C}\text{ is connected}}} (-1)^{c(\phi_{\mathcal{C}})} \operatorname{tr}_{\phi_{\mathcal{C}}}(\boldsymbol{M}_{\boldsymbol{P}}),$$

where the polynomials in the tuple $\boldsymbol{M}_{\boldsymbol{P}}$ are evaluated at the matrices $A_1^M, (A_1^M)^*, \dots, A_{2m}^M, (A_{2m}^M)^*$. Proposition 3.33 implies that $G(1/qM) = \widetilde{\mathcal{W}}_{0,l}^N(\boldsymbol{P})$.

Thus, as the renormalized cumulant under the Haar measure $\widetilde{\mathcal{W}}_{0,l}^N = N^{l-2}\mathcal{W}_{0,l}^N$ satisfies all four properties, and the set $\{1/qM\}_{q\geq 1}$ has an accumulation point, we get the result. ∎

3.5. Formal topological expansion

When the potential V is not zero, we expect to have an expansion of the free energy as in Proposition 3.33. Let us now consider a potential of the form $V = \sum_{i=1}^k z_i q_i$, with $\boldsymbol{z} = (z_1, \dots, z_k) \in \mathbb{C}^k$ and $\boldsymbol{q} = (q_1, \dots, q_k) \in \mathcal{X}^k$.

Proposition 3.33 motivates the introduction of the formal series

$$\begin{aligned} F_V^{N,f} &= \frac{1}{N^2} \sum_{\boldsymbol{n}\in\mathbb{N}^k} \frac{(-\boldsymbol{z})^{\boldsymbol{n}}}{\boldsymbol{n}!} \widetilde{\mathcal{W}}^N_{0,\sum_i n_i}(\boldsymbol{q}_{\boldsymbol{n}}) \\ &= \sum_{g\geq 0} \frac{1}{N^{2g}} \sum_{\boldsymbol{n}\in\mathbb{N}^k} \frac{\boldsymbol{z}^{\boldsymbol{n}}}{\boldsymbol{n}!} (-1)^{\deg \boldsymbol{q}_{\boldsymbol{n}}} \sum_{\substack{\mathcal{C}\in\mathfrak{C}(g,[2\deg \boldsymbol{q}_{\boldsymbol{n}}],\boldsymbol{\varepsilon}_{\boldsymbol{q}_{\boldsymbol{n}}},\gamma(\boldsymbol{q}_{\boldsymbol{n}}))\\ \mathcal{C}\text{ is connected}}} (-1)^{c(\phi_{\mathcal{C}})} \operatorname{tr}_{\phi_{\mathcal{C}}}(\boldsymbol{M}_{\boldsymbol{q}_{\boldsymbol{n}}}), \end{aligned} \tag{3.4}$$

where we use the notation $\boldsymbol{q}_{\boldsymbol{n}} = (\underbrace{q_1, \dots, q_1}_{n_1\text{ times}}, \dots, \underbrace{q_k, \dots, q_k}_{n_k\text{ times}})$ for $\boldsymbol{n} = (n_1, \dots, n_k)$, as well as $\boldsymbol{z}^{\boldsymbol{n}} = \prod_{i=1}^k z_i^{n_i}$ and $\boldsymbol{n}! = \prod_{i=1}^k n_i!$.

Similarly, we introduce formal series corresponding to the cumulants.

Definition 3.38. Let $N \in \mathbb{N}^*$, $\boldsymbol{P} = (P_1, \dots, P_l) \in \mathcal{X}^l$ be monomials with $m = \frac{1}{2}\deg \boldsymbol{P}$. The *formal cumulant* of $\boldsymbol{P}$ is the formal series

$$\mathcal{M}_{V,l}^N(P_1, \dots, P_l) = \sum_{g\geq 0} \frac{1}{N^{2g}} \mathcal{M}_{V,l}^{(g),N}(P_1, \dots, P_l),$$

where the gth term is

$$\begin{aligned} \mathcal{M}_{V,l}^{(g),N}(P_1, \dots, P_l) = &\sum_{\boldsymbol{n}\in\mathbb{N}^k} \frac{\boldsymbol{z}^{\boldsymbol{n}}}{\boldsymbol{n}!} (-1)^{\deg \boldsymbol{q}_{\boldsymbol{n}} + \deg \boldsymbol{P}} \\ &\times \sum_{\substack{\mathcal{C}\in\mathfrak{C}(g,[\deg \boldsymbol{q}_{\boldsymbol{n}}+\deg \boldsymbol{P}],\boldsymbol{\varepsilon}_{\boldsymbol{q}_{\boldsymbol{n}}\boldsymbol{P}},\gamma(\boldsymbol{q}_{\boldsymbol{n}}\boldsymbol{P}))\\ \mathcal{C}\text{ is connected}}} (-1)^{c(\phi_{\mathcal{C}})} \operatorname{tr}_{\phi_{\mathcal{C}}}(\boldsymbol{M}_{\boldsymbol{q}_{\boldsymbol{n}}\boldsymbol{P}}), \end{aligned}$$

where $\boldsymbol{q_n P}$ is the concatenation of the two tuples $\boldsymbol{q_n}$ and $\boldsymbol{P}$. In particular, $\gamma_{\boldsymbol{q_n P}}$ is the permutation defined in (2.6) associated to the tuple $\boldsymbol{q_n P}$.

Notice that the total numbers of u and u^* (or white half-edges) are $\sum_i n_i \deg q_i + 2m$, and the number of white vertices is $\sum_i n_i + l$. The case of the formal free energy corresponds to $m = 0$, $l = 0$.

At this point, it is not clear whether the series $\mathcal{M}_{V,l}^{(g),N}(P_1, \ldots, P_l)$ converge. It will be shown in Section 4.3. In Section 6, we will show that in the asymptotic regime, the cumulant $\mathcal{W}_{V,l}^N(P_1, \ldots, P_l)$ coincides with the formal cumulant up to an arbitrary order, for $\boldsymbol{z}$ small enough.

3.6. Alternated polynomials and Hurwitz numbers

In this section, we consider a particular case; that is, we assume that all polynomials are alternated monomials (see Definition 3.39). In particular, this covers the case of a potential of the form $V = zAU^N B(U^N)^*$ encountered in the HCIZ integral. In [19], the HCIZ integral had been expressed in terms of monotone double Hurwitz numbers. In the multimatrix case, results relating the more general tensor HCIZ integral to the Hurwitz numbers have been obtained in [9].

Here, we consider only the case where we have a single unitary matrix.

Definition 3.39. A monomial $P \in \mathcal{A}$ is said to be *alternated* if it can be written as

$$P = B_1 u C_1 u^{-1} \cdots B_m u C_m u^{-1},$$

with $B_i, C_i, 1 \le i \le m$ words in $a_1, a_1^*, \ldots, a_p, a_p^*$.

In this section, we assume that all the polynomials involved, $P_1, \ldots, P_l, q_1, \ldots, q_k$, are alternated monomials. We write as before $\boldsymbol{P} = (P_1, \ldots, P_l)$. We now explain how we can give different expressions for the renormalized cumulants in this case. To do so, we reason only using the permutational model of the maps of unitary type.

In this case, $\boldsymbol{\varepsilon} = (+1, -1, +1, -1, \ldots)$, and we have $\gamma_{\boldsymbol{P}}(\boldsymbol{\varepsilon}^{-1}(+1)) = \boldsymbol{\varepsilon}^{-1}(-1)$ and $\gamma_{\boldsymbol{P}}(\boldsymbol{\varepsilon}^{-1}(-1)) = \boldsymbol{\varepsilon}^{-1}(+1)$. Thus, $\gamma_{\boldsymbol{P}} \in \mathfrak{S}_{2m}^{(\boldsymbol{\varepsilon})}$. We define $\tilde{\gamma} = \gamma_{\boldsymbol{P}}^2|_{\boldsymbol{\varepsilon}^{-1}(+1)}$.

In particular, this implies for all $\mathcal{C} \in \mathfrak{C}(g, [2m], \boldsymbol{\varepsilon}, \gamma_{\boldsymbol{P}})$, we have $\phi_{\mathcal{C}}(\boldsymbol{\varepsilon}^{-1}(+1)) = \boldsymbol{\varepsilon}^{-1}(+1)$ and $\phi_{\mathcal{C}}(\boldsymbol{\varepsilon}^{-1}(-1)) = \boldsymbol{\varepsilon}^{-1}(-1)$. That is, we can write $\phi_{\mathcal{C}}$ as a product of two permutations, one, $\rho_{\mathcal{C}}$, having its support in $\boldsymbol{\varepsilon}^{-1}(+1)$, and the other, $\sigma_{\mathcal{C}}$, having its support in $\boldsymbol{\varepsilon}^{-1}(-1)$.

Write $\tau_{\mathcal{C}} = (\tau_1, \ldots, \tau_r)$. We notice that the group generated by $\gamma_{\boldsymbol{P}}, \phi_{\mathcal{C}}, \tau_1, \ldots, \tau_r$ is transitive if and only if the group generated by $\gamma_{\boldsymbol{P}}, \rho_{\mathcal{C}}, \sigma_{\mathcal{C}}, \tau_1, \ldots, \tau_r$ is transitive. As conjugating the elements of a group by one of the elements does not change the group, we have

$$\langle \gamma_{\boldsymbol{P}}, \sigma_{\mathcal{C}}, \rho_{\mathcal{C}}, \tau_1, \ldots, \tau_r \rangle = \langle \gamma_{\boldsymbol{P}}, \gamma_{\boldsymbol{P}}^{-1} \sigma_{\mathcal{C}} \gamma_{\boldsymbol{P}}, \rho_{\mathcal{C}}, \tau_1, \ldots, \tau_r \rangle.$$

Now, we remark that the subgroup of $\mathfrak{S}_{2m}$ on the right-hand side is transitive if and only if the subgroup $\langle \widetilde{\gamma}, \gamma_{\boldsymbol{P}}^{-1}\sigma\gamma_{\boldsymbol{P}}, \rho, \tau_1, \ldots, \tau_r\rangle$ of $\mathfrak{S}(\varepsilon^{-1}(+1))$ is transitive.

This remark allows us to rewrite the sum of Definition 3.35,

$$\begin{aligned}\mathcal{W}_{0,l}^{(g),N}(P_1,\ldots,P_l) &= (-1)^m \sum_{\substack{\mathcal{C}\in\mathfrak{C}(g,[2m],\boldsymbol{\varepsilon}_{\boldsymbol{P}}\gamma_{\boldsymbol{P}})\\ \mathcal{C}\text{ connected}}} (-1)^{c(\phi_{\mathcal{C}})}\operatorname{tr}_{\phi_{\mathcal{C}}}(\boldsymbol{M_P})\\ &= (-1)^m \sum_{\substack{\sigma,\rho\in\mathfrak{S}_m\\ \tau\in\overrightarrow{W}_g(\mathrm{Id},\rho^{-1}\widetilde{\gamma}\sigma)\\ \langle\widetilde{\gamma},\sigma,\rho,\tau_1,\ldots,\tau_r\rangle\text{ transitive}}} (-1)^{c(\sigma)+c(\rho)}\operatorname{tr}_\rho(\boldsymbol{B_P})\operatorname{tr}_\sigma(\boldsymbol{C_P}),\end{aligned}$$

where we introduced the notation

$$\begin{aligned}\boldsymbol{B_P} &= (M_i; i\in\boldsymbol{\varepsilon}^{-1}(+1)),\\ \boldsymbol{C_P} &= (M_i; i\in\boldsymbol{\varepsilon}^{-1}(-1)),\end{aligned}$$

if $\boldsymbol{M_P} = (M_i)_{1\le i\le \deg \boldsymbol{P}}$.

Definition 3.40. Let $\rho,\gamma,\sigma\in\mathfrak{S}_m$. The *$r$th monotone triple Hurwitz number* associated to ρ,γ,σ, denoted by $\vec{h}^r(\rho,\gamma,\sigma)$, is the number of r-tuples of transpositions $(\tau_1,\ldots,\tau_r)\in\mathfrak{S}_m^r$ such that

- $\tau_r\cdots\tau_1 = \rho\gamma\sigma$;
- $\mathrm{val}(\tau_1)\le\mathrm{val}(\tau_2)\le\cdots\le\mathrm{val}(\tau_r)$;
- the group $\langle\gamma,\rho,\sigma,\tau_1,\ldots,\tau_r\rangle\subset\mathfrak{S}_m$ is transitive.

When g satisfies the Euler equation

$$2-2g = c(\gamma)+c(\rho)+c(\sigma)-r-m,$$

we set $\vec{h}_g(\gamma,\sigma,\rho) = \vec{h}^r(\gamma,\sigma,\rho)$.

This gives us the following proposition.

Proposition 3.41. *Let $\boldsymbol{P} = (P_1,\ldots,P_l)$ be alternated monomials. We have*

$$\begin{aligned}&\mathcal{W}_{0,l}^{(g),N}(P_1,\ldots,P_l)\\ &\quad= (-1)^m\sum_{\sigma,\rho\in\mathfrak{S}_m}(-1)^{c(\sigma)+c(\rho)}\operatorname{tr}_\rho(\boldsymbol{B_P})\operatorname{tr}_\sigma(\boldsymbol{C_P})\vec{h}_g(\rho^{-1},\widetilde{\gamma},\sigma).\end{aligned}\tag{3.5}$$

Remark 3.42. In the case of the HCIZ integral, we have $\widetilde{\gamma} = \mathrm{Id}$; thus, the monotone triple Hurwitz numbers reduce to the monotone double Hurwitz numbers.

Remark 3.43. Notice that when all the polynomials are alternated, all the white vertices in the unitary type maps involved are alternated vertices (see Definition 3.12).

Remark 3.44. In this particular case, the maps of unitary type bear similarity with the ribbon graphs introduced in [27].

4. Tutte-like equations

We will now state induction relations that apply to the sums of maps $\mathcal{M}_{0,l}^{(g),N}$ defined in Definition 3.35. They are obtained by a procedure very similar to the one used by Tutte in [40]. These induction relations are the analog of the topological recursion for matrices of the GUE [18]. Similar induction relations have been obtained for maps related to the Gaussian case in [22] and [31], and for maps with "dotted edges" in the unitary case for $g = 0$ in [8]. More precisely, we will prove the following theorem.

Theorem 4.1. *Let $N \in \mathbb{N}^*$, $\boldsymbol{P} = (P_1, \dots, P_l u) \in \mathcal{X}^l$ be monomials. Then, for $g \geq 0$ and $m = \frac{1}{2} \deg \boldsymbol{P} \geq 2$, we have the induction relation*

$$\begin{aligned}
&\mathcal{M}_{0,l}^{(g),N}(P_1, \dots, P_l u)\\
&\quad = - \sum_{P_l = QuR} \mathcal{M}_{0,l+1}^{(g-1),N}(P_1, \dots, P_{l-1}, Qu, Ru)\\
&\qquad - \sum_{P_l = QuR} \sum_{\substack{g_1+g_2=g\\ I \subset [l-1]}} \mathcal{M}_{0,|I|+1}^{(g_1),N}(\boldsymbol{P}|_I, Qu) \mathcal{M}_{0,l-|I|}^{(g_2),N}(\boldsymbol{P}|_{I^c}, Ru)\\
&\qquad - \sum_{j=1}^{l-1} \sum_{P_j = QuR} \mathcal{M}_{0,l-1}^{(g),N}(P_1, \dots, P_{j-1}, P_{j+1}, \dots, P_{l-1}, RQuP_l u)\\
&\qquad + \sum_{P_l = Qu^* R} \mathcal{M}_{0,l+1}^{(g-1),N}(P_1, \dots, P_{l-1}, Q, R)\\
&\qquad + \sum_{P_l = Qu^* R} \sum_{\substack{g_1+g_2=g\\ I \subset [l-1]}} \mathcal{M}_{0,|I|+1}^{(g_1),N}(\boldsymbol{P}|_I, Q) \mathcal{M}_{0,l-|I|}^{(g_2),N}(\boldsymbol{P}|_{I^c}, R)\\
&\qquad + \sum_{j=1}^{l-1} \sum_{P_j = Qu^* R} \mathcal{M}_{0,l-1}^{(g),N}(P_1, \dots, P_{j-1}, P_{j+1}, \dots, P_{l-1}, RQP_l),
\end{aligned}$$

where we use the notation $\boldsymbol{P}|_I = (P_i)_{i \in I}$ and we set by convention, $\mathcal{M}_{0,l}^{(-1),N} = 0$ and $\mathcal{M}_{0,0}^{(g),N} = 0$.

With a similar proof, we can state a similar theorem for $\boldsymbol{P} = (P_1, \dots, P_l u^*)$. Equivalently, this is a consequence of the invariance by conjugation of the sums of maps, see Lemma 3.37.

Theorem 4.1 describes how a map of unitary type can be decomposed into one or several maps. In Section 4.1, we will describe the precise procedure used to cut maps of unitary type into one or more maps of unitary types, or equivalently to decompose the permutations describing a particular map. The procedure can be understood as an elaborate way of contracting an edge in a map. This allows us to give an interpretation to the terms appearing in the recurrence.

Consider first the terms in the last two lines (the other terms have a similar interpretation). When contracting an edge, three situations can occur.

- The edge is a non-contractible loop; informally, it goes around a handle of the surface, in that case when contracting it, we remove the handle and create a new vertex, this corresponds to the term

$$\sum_{P_l = Qu^* R} \mathcal{M}_{0,l+1}^{(g-1),N}(P_1, \dots, P_{l-1}, Q, R).$$

- The edge is a contractible loop. When contracting the edge, we cut the map into two disconnected components. This corresponds to the term

$$\sum_{P_l = Qu^* R} \sum_{\substack{g_1+g_2=g \\ I \subset [l-1]}} \mathcal{M}_{0,|I|+1}^{(g_1),N}(\boldsymbol{P}|_I, Q)\mathcal{M}_{0,l-|I|}^{(g_2),N}(\boldsymbol{P}|_{I^c}, R).$$

- The edge is not a loop. When contracting the edge, we just merge two vertices. This corresponds to the term

$$\sum_{j=1}^{l-1} \sum_{P_j = Qu^* R} \mathcal{M}_{0,l-1}^{(g),N}(P_1, \dots, P_{j-1}, P_{j+1}, \dots, P_{l-1}, RQP_l).$$

The other terms have a similar form. They do not correspond to the contraction of an edge, but rather to the erasure of a black vertex. This theorem shows that a sum of maps can be expressed in terms of the sums of maps with either lower genus, lower overall degree, or lower number of vertices.

Before describing precisely the procedure used to cuts the maps of unitary type, and giving the proof of Theorem 4.1, we show how to rewrite these equations in a compact way. The function $\mathcal{M}_{0,l}^{(g),N}$ was defined only on some monomials, i.e., on the set $\mathcal{X}^l$. Recall that $\mathcal{A}$ is the algebra of noncommutative polynomials in the variables $u, u^*, a_1, a_1^*, \dots$. We now extend $\mathcal{M}_{0,l}^{(g),N}$ by linearity to a linear form on the tensor space $\mathcal{A}^{\otimes l} = \mathcal{A} \otimes_{\mathbb{C}} \cdots \otimes_{\mathbb{C}} \mathcal{A}$. We will use the notation $\mathcal{M}_{0,l}^{(g),N}(P_1, \dots, P_l)$ ($\mathcal{M}_{0,l}^{(g),N}$ as a multilinear function) and $\mathcal{M}_{0,l}^{(g),N}(P_1 \otimes \cdots \otimes P_l)$ ($\mathcal{M}_{0,l}^{(g),N}$ as a linear function on $\mathcal{A}^{\otimes l}$) interchangeably. We can then rewrite Theorem 4.1 using the notion of noncommutative derivative.

Let $\boldsymbol{P} = (P_1, \dots, P_l)$ be a k-tuple of polynomials, and $I = \{i_1 < i_2 < \cdots < i_p\}$ be a non-empty subset of $[l]$; then, we define

$$\boldsymbol{P}_I = P_{i_1} \otimes P_{i_2} \otimes \cdots \otimes P_{i_p}.$$

We define the operations $\times$ and $\sharp$ as follows. Let $\boldsymbol{P} \in \mathcal{A}^l$, $I \subset [l-1]$, $Q = Q_1 \otimes Q_2 \in \mathcal{A}^{\otimes 2}$ and $S = S_1 \otimes S_2 \in \mathcal{A}^{\otimes 2}$; then, we set

$$Q \times S = (Q_1 \otimes Q_2) \times (S_1 \otimes S_2) = (Q_1 S_1) \otimes (Q_2 S_2) \tag{4.1}$$

and

$$\boldsymbol{P}_I \otimes \boldsymbol{P}_{I^c} \sharp Q = \boldsymbol{P}_I \otimes Q_1 \otimes \boldsymbol{P}_{I^c} \otimes Q_2. \tag{4.2}$$

Definition 4.2. The *logarithmic non-commutative derivative*

$$\partial \colon \mathcal{A} \to \mathcal{A}^{\otimes 2}$$

with respect to u of a monomial $P \in \mathcal{A}$ is defined by

$$\partial P = \sum_{P = QuR} Qu \otimes R - \sum_{P = Qu^{-1}R} Q \otimes u^{-1} R \in \mathcal{A}^{\otimes 2}.$$

The definition extends by linearity to any polynomial in $\mathcal{A}$.

Remark 4.3. This derivative was introduced by Voiculescu in [41, Section 8.1]. However, as remarked by an anonymous reviewer, this corresponds to the derivation on the unitary group invariant under multiplication from the right. Near the identity, this derivation corresponds to the ordinary gradient ∇_H, where $\exp^H = U$, hence the name logarithmic derivative.

Remark 4.4. Consider a monomial $P \in \mathcal{A}$ evaluated in $U, U^*, A_1, A_1^*, \ldots$, and denote by $\partial_{m,j}$ the derivative with respect to the coefficient (m, j) of U. We have

$$\sum_m U_{m,k} (\partial_{m,j} P)_{i,l} = \sum_m \sum_{P = QUR} Q_{i,m} U_{m,k} R_{j,l} = \sum_{P = QUR} (QU)_{i,k} R_{j,l},$$

and similarly, if $\overline{\partial}_{m,j}$ is the derivative with respect to the coefficient (m, j) of U^*,

$$\sum_m U^*_{j,m} (\overline{\partial}_{k,m} P)_{i,l} = \sum_m \sum_{P = QU^*R} Q_{i,k} U^*_{j,m} R_{m,l} = \sum_{P = QUR} Q_{i,k} (U^* R)_{j,l}.$$

In coordinates, the non-commutative derivative thus corresponds to

$$(\partial P)_{\substack{i,k \\ j,l}} = \sum_m \big(U_{m,k} \partial_{m,j} P - U^*_{j,m} \overline{\partial}_{k,m} P \big)_{i,l}.$$

Definition 4.5. The *cyclic derivative* $\mathcal{D} \colon \mathcal{A} \to \mathcal{A}$ with respect to u of a monomial $P \in \mathcal{A}$ is defined by

$$\mathcal{D} P = \sum_{P = QuR} RQu - \sum_{P = Qu^{-1}R} u^{-1} RQ.$$

The definition extends by linearity to any polynomial in $\mathcal{A}$.

Remark 4.6. Let P be a monomial. In coordinates, the cyclic derivative is

$$(\mathcal{D}P)_{i,j} = \sum_{k,m} \big(U_{m,j}\partial_{m,i}P - U^*_{m,j}\overline{\partial}_{j,m}P\big)_{k,k} = \sum_k (\partial P)_{\substack{j,k \\ i,k}}.$$

To take advantage of the notation just introduced, we will write

$$\mathcal{M}^{(g),N}_{0,l} \otimes \mathcal{M}^{(g'),N}_{0,l'}(P_1 \otimes \cdots \otimes P_{l+l'})$$

with $P_i \in \mathcal{A}$ to mean

$$\begin{aligned}&\mathcal{M}^{(g),N}_{0,l} \otimes \mathcal{M}^{(g'),N}_{0,l'}(P_1 \otimes \cdots \otimes P_{l+l'}) \\ &\quad = \mathcal{M}^{(g),N}_{0,l}(P_1,\ldots,P_l)\mathcal{M}^{(g'),N}_{0,l'}(P_{l+1},\ldots,P_{l+l'}).\end{aligned}$$

Theorem 4.1 can then be rewritten as follows.

Corollary 4.7. *For $m \geq 2$, $g \geq 1$, and $P_1,\ldots,P_l \in \mathcal{A}$, we have the following equation:*

$$\begin{aligned}&\sum_{I\subset[l-1]}\sum_{g=g_1+g_2} \mathcal{M}^{(g_1),N}_{0,|I|+1} \otimes \mathcal{M}^{(g_2),N}_{0,|I^c|+1}(\boldsymbol{P}_I \otimes \boldsymbol{P}_{I^c}\sharp\partial P_l) \\ &\quad = -\mathcal{M}^{(g-1),N}_{0,l+1}(P_1 \otimes \cdots \otimes P_{l-1} \otimes \partial P_l) \\ &\qquad -\sum_{j=1}^{l-1} \mathcal{M}^{(g),N}_{0,l-1}(P_1 \otimes \cdots \otimes \check{P}_j \otimes \cdots \otimes (\mathcal{D}P_j)P_l),\end{aligned} \tag{4.3}$$

where $\check{P}_j$ means that the factor P_j is omitted.

Proof. We group the terms two by two. We have

$$\begin{aligned}&-\sum_{P_l=QuR} \mathcal{M}^{(g-1),N}_{0,l+1}(P_1,\ldots,P_{l-1},Qu,Ru) \\ &\qquad + \sum_{P_l=Qu^*R} \mathcal{M}^{(g-1),N}_{0,l+1}(P_1,\ldots,P_{l-1},Q,R) \\ &\quad = -\mathcal{M}^{(g-1),N}_{0,l+1}(P_1,\ldots,P_{l-1},\partial(P_l u)) \\ &\qquad + \mathcal{M}^{(g-1),N}_{0,l+1}(P_1,\ldots,P_{l-1},P_l,1) \\ &\quad = -\mathcal{M}^{(g-1),N}_{0,l+1}(P_1,\ldots,P_{l-1},\partial(P_l u)),\end{aligned}$$

where we used the traciality property of Lemma 3.37 to replace u^*Ru by R in the last argument of $\mathcal{M}^{(g-1),N}_{0,l+1}$ in the second line. In the third line, we used the fact that there are no connected map with $l+1$ vertices, $l \geq 1$, and one vertex of degree 0.

We proceed similarly for the two other terms. We have

$$\begin{aligned}
&-\sum_{P_l=QuR}\sum_{\substack{g_1+g_2=g\\ I\subset[l-1]}} \mathcal{M}^{(g_1),N}_{0,|I|+1}(\boldsymbol{P}|_I, Qu)\mathcal{M}^{(g_2),N}_{0,l-|I|}(\boldsymbol{P}|_{I^c}, Ru)\\
&+\sum_{P_l=Qu^*R}\sum_{\substack{g_1+g_2=g\\ I\subset[l-1]}} \mathcal{M}^{(g_1),N}_{0,|I|+1}(\boldsymbol{P}|_I, Q)\mathcal{M}^{(g_2),N}_{0,l-|I|}(\boldsymbol{P}|_{I^c}, R)\\
&\quad=-\sum_{\substack{g_1+g_2=g\\ I\subset[l-1]}} \mathcal{M}^{(g_1),N}_{0,|I|+1}\otimes\mathcal{M}^{(g_2),N}_{0,l-|I|}(\boldsymbol{P}|_I\otimes\boldsymbol{P}|_{I^c}\sharp(\partial P u)) + \mathcal{M}^{(g),N}_{0,l}(\boldsymbol{P}),
\end{aligned}$$

where we used that there are no maps with one vertex of degree 0 if $g\geq 1$ or $l\geq 2$.

The third term is

$$\begin{aligned}
&-\sum_{j=1}^{l-1}\sum_{P_j=QuR}\mathcal{M}^{(g),N}_{0,l-1}(P_1,\dots,P_{j-1},P_{j+1},\dots,P_{l-1},RQuP_lu)\\
&+\sum_{j=1}^{l-1}\sum_{P_j=Qu^*R}\mathcal{M}^{(g),N}_{0,l-1}(P_1,\dots,P_{j-1},P_{j+1},\dots,P_{l-1},RQP_l)\\
&\quad=-\sum_{j=1}^{l-1}\mathcal{M}^{(g),N}_{0,l-1}(P_1,\dots,P_{j-1},P_{j+1},\dots,P_{l-1},(\mathcal{D}P_j)P_lu).
\end{aligned}$$

Putting the three terms together, we get the result when P_l is a monomial that finishes by u. The cyclicity property of Lemma 3.37 implies that it is then the same if P_l contains a u. Finally, if P_l does not contain a u, then it is either a constant and the formula is clear, or it contains a u^*. In the latter case, the conjugation property of Lemma 3.37 allows us to recover the result.

Finally, the result extends by linearity to all polynomials, as it is linear in each of the P_i. ∎

Notice that the formula in Corollary 4.7 is valid not only for monomials P_l finishing by a u but also for all polynomials.

4.1. How to cut maps

In this section, we fix two integers $m\geq 2$ and $g\geq 0$, a permutation $\gamma\in\mathfrak{S}_{2m}$ and $\boldsymbol{\varepsilon}\in\mathcal{E}_{2m}=\{\boldsymbol{\varepsilon}=(\varepsilon(i))_{i\in[2m]}\in\{\pm 1\}^{2m}\colon \sum_{i=1}^{2m}\varepsilon(i)=0\}$. By the cyclicity and the symmetry properties proved in Lemma 3.37, we can assume that the polynomial P_l finishes by the letter u. We thus assume that $\varepsilon(2m)=+1$. We consider a map of unitary type, $\mathcal{C}\in\mathfrak{C}(g,m,\gamma,\boldsymbol{\varepsilon})$, with r black vertices. Let S denote the underlying surface of the map $\mathcal{C}$.

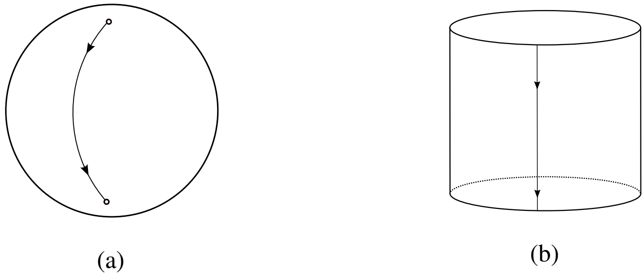

Figure 10. (a) An oriented map of genus 0 with two vertices. (b) The corresponding map where vertices are seen as boundary components. The underlying surface is a cylinder.

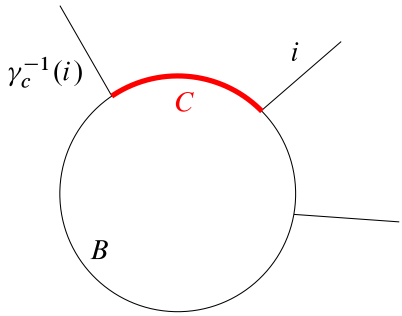

Figure 11. The corner C at the left of i, displayed with a red thick line. Here, i is a white half-edge.

By Theorem 3.28, this map is described by the permutation $\pi := \pi_{\mathcal{C}}$ and the tuple of transpositions $\tau_{\mathcal{C}} = (\tau_1, \ldots, \tau_r) \in \overrightarrow{W}^r(\pi^{(\varepsilon)})$, with r related to g, $\pi_{\mathcal{C}}$, and $\gamma_{\mathcal{C}}$ by Euler's formula, see (3.3). We will consider two ways of cutting this map, depending on whether $\tau_r(2m) = 2m$ or not.

Remark 4.8. We can also see vertices as "holes" in the surface, that is, we take the underlying surface S to be a surface with boundaries. A vertex is then a boundary component (homeomorphic to a circle) of the surface S. An edge is then a path connecting two boundary components. See Figure 10 for an example.

We will see white vertices as boundary components of the underlying surface, as explained in Remark 4.8. To describe the cutting procedure, we introduce the definition of a corner; see Figure 11.

Definition 4.9. Let $\mathcal{C}$ be a map of unitary type. Consider a (white or black) vertex v in $\mathcal{C}$, seen as a boundary component of a surface, as explained in Remark 4.8, and a half-edge h.

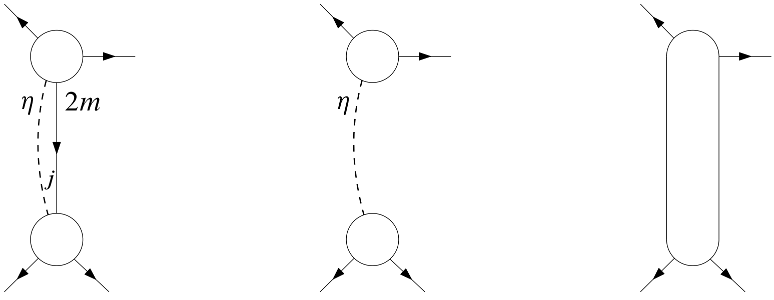

Figure 12. First way to cut the map.

The *corner* at the left (respectively, at the right) of h is the connected, closed set C such that

- C is a subset of the boundary component B corresponding to v;
- C is in the boundary of the face f at the left (respectively, right) of the half-edge h;
- the boundary of C is $\{u, v\}$, where u is the point of intersection of B and the half-edge h, and v is the point of intersection of B and the half-edge h' that follows the half-edge h when going around the face f in the clockwise (resp., counterclockwise) direction starting from the left (resp., the right) of h.

4.1.1. First case: $\boldsymbol{\tau_r(2m) = 2m}$. In this case, $\mathrm{val}(\tau_r) < 2m$, see Definition 2.12. The monotonicity of the walk $\boldsymbol{\tau}$ implies that for all $i \in [r]$, $\mathrm{val}(\tau_i) < 2m$, and in particular $\tau_i(2m) = 2m$. By Definition 3.22, the half-edge $2m$ is not connected to any black half-edge, and is thus connected to a white-half-edge, say the jth one. Note that by our assumption that $\varepsilon(2m) = +1$, $\varepsilon(j) = -1$. Notice that because $\mathcal{C}$ is non-decreasing, $\tau_r(2m) = 2m$ implies that for all i, $\tau_i(2m) = 2m$.

We construct a map $\mathcal{C}'$ of unitary type from $\mathcal{C}$ using the following procedure, depicted in Figure 12.

(1) We choose a path η in the face f at the right of the half-edge $2m$. This path is chosen to start from the white vertex w_{2m}, attached in the corner (see Definition 4.9) at the right of the half-edge $2m$, and end at w_j, attached in the corner at the left of the half-edge j. As faces are homeomorphic to disks, there is only one way to choose η up to homotopy.

(2) We remove the edge containing the half-edges j and $2m$.

(3) We cut the surface along η. Depending on the cases we connect two distinct boundary components of S, or we connect one boundary component to itself.

Remark 4.10. Notice that if w_j and w_{2m} are distinct vertices, this surgery is the usual contraction of an edge.

Remark 4.11. If the path η is a loop, then it may be that the resulting map $\mathcal{C}'$ is disconnected. Furthermore, one of the connected components can be a vertex with no edges. In this case, we remove this component.

Lemma 4.12. *If $\mathcal{C}$ is a map of unitary type with label set I of size $2m$, then $\mathcal{C}'$ is a map of unitary type, with label set $I \setminus \{j, 2m\}$, of size $2m-2$. Furthermore, if $\mathcal{C}$ is non-decreasing, then so is $\mathcal{C}'$.*

Proof. We first check that $\mathcal{C}'$ is a map. We only need to check that each face of $\mathcal{C}'$ is homeomorphic to a disk. We only modify the faces f_{left} and f_{right} at the left and the right of the edge (w_{2m}, w_j). At step 2, when we remove the edge, we connect f_{left} and f_{right}. However, at step 3, we cut along a path homotopic to the edge, thus separating the two faces. All the faces of $\mathcal{C}'$ thus remain disks.

We now check that this map is a map of unitary type. The map $\mathcal{C}'$ has $2m-2$ labeled white half-edges (maybe 0 if $m = 1$), with label set $I \setminus \{j, 2m\}$ (property 2 in Definition 3.13 is satisfied). The black vertices have not been modified when transforming $\mathcal{C}$ into $\mathcal{C}'$, so properties 1 and 3 in Definition 3.13 are satisfied. As the black vertices are not modified, a non-decreasing map remains non-decreasing. ∎

Let us now compute the permutations that represent $\mathcal{C}'$. We will need the notion of the trace of a permutation introduced by Kreweras [29]. This notion has nothing to do with the notion of trace of a matrix.

Definition 4.13. Let A be a finite subset of $\mathbb{N}$. The *trace* of a permutation $\sigma \in \mathfrak{S}(A)$, on $B \subset A$, denoted by $\mathrm{Tr}(\sigma; B)$, is the permutation in $\mathfrak{S}(B)$ defined for each $x \in B$ by

$$\mathrm{Tr}(\sigma; B)(x) = \sigma^{p_x}(x),$$

with $p_x \geq 1$ the smallest integer so that $\sigma^p(x) \in B$.

Computing the trace of a permutation in cycle notation is straightforward: write the cycle decomposition of σ, and erase all elements in the cycles that do not belong to B.

Let $I_j = [2m-1] \setminus \{j\}$.

Lemma 4.14. *Let $\pi' := \mathrm{Tr}(\pi_{\mathcal{C}}; I_j)$. We have $\pi' = \pi_{\mathcal{C}}(j\ 2m) = (j\ 2m)\pi_{\mathcal{C}}$, and $\pi_{\mathcal{C}'} = \pi' = \pi_{\mathcal{C}}|_{I_j}$.*

Proof. We have assumed that the half-edges $2m$ and j are connected to form an edge. This imply $\pi_{\mathcal{C}}(2m) = j$ and $\pi_{\mathcal{C}}(j) = 2m$. Thus, as $\pi_{\mathcal{C}}$ is a permutation, for all $k \in I_j$,

$\pi_{\mathcal{C}}(k) \in I_j$. This means that in the notation of Definition 4.13, $p_k = 1$. We get the first claim, and that $\pi' = \pi_{\mathcal{C}}|_{I_j}$.

When removing the edge at step 2, it is clear that the map we obtain is still described by $\pi_{\mathcal{C}}$, with a cycle removed. When cutting the map at step 3, we do not modify the edges further. ■

Lemma 4.15. *Let* $\gamma' := \operatorname{Tr}(\gamma(j\ 2m); I_j)$. *We have* $\gamma_{\mathcal{C}'} = \gamma'$.

Proof. Assume first that w_j and w_{2m} are two distinct vertices. Let $c = (u_1\ \ldots\ u_p\ j)$ and $c' = (u'_1\ \ldots\ u'_{p'}\ 2m)$ be the cycles that represent them. After cutting the map at step 3, the vertices are replaced by a vertex with structure $(u_1\ \ldots\ u_p\, u'_1\ \ldots\ u'_{p'}) = \operatorname{Tr}(cc'(j\ 2m); I_j)$.

If $w_j = w_{2m}$, this vertex is represented by a cycle $c = (u_1 \ldots u_p\ j\ u'_1 \ldots u'_{p'}\ 2m)$, which we cut using the transposition $(j\ 2m)$. We obtain two vertices represented by the two cycles $\operatorname{Tr}(c(j\ 2m); I_j)$. ■

Lemma 4.16. *We have* $\phi_{\mathcal{C}'} = \operatorname{Tr}(\phi_{\mathcal{C}}; I_j)$.

Proof. By Lemmas 3.25, 4.14, and 4.15,

$$\phi_{\mathcal{C}'} = \gamma'\pi'^{-1} = \operatorname{Tr}(\gamma_{\mathcal{C}}(j\ 2m); I_j)\pi_{\mathcal{C}}|_{I_j}^{-1}.$$

Notice first that for any $k \in I$ and $p \in \mathbb{N}$, $(\gamma_{\mathcal{C}}\pi_{\mathcal{C}}^{-1})^p(k) \in \{j, 2m\}$ if and only if $((j\ 2m)\gamma_{\mathcal{C}}\pi_{\mathcal{C}}^{-1}(j\ 2m))^p(k) \in \{j, 2m\}$. This implies that

$$\operatorname{Tr}(\gamma_{\mathcal{C}}\pi_{\mathcal{C}}^{-1}; I_j) = \operatorname{Tr}((j\ 2m)\gamma_{\mathcal{C}}\pi_{\mathcal{C}}^{-1}(j\ 2m); I_j) = \operatorname{Tr}((j\ 2m)\gamma_{\mathcal{C}}^{-1}\pi'^{-1}; I_j),$$

where we used Lemma 4.14 for the last equality.

Then, as $\pi'(j) = j$ and $\pi'(2m) = 2m$, we have

$$\operatorname{Tr}((j\ 2m)\gamma\pi'^{-1}; I_j) = \operatorname{Tr}((j\ 2m)\gamma; I_j)\pi'^{-1} = \phi_{\mathcal{C}'}. \quad ■$$

Lemma 4.17. *If the map* $\mathcal{C}$ *of unitary type is connected, then the map* $\mathcal{C}'$ *has one or two connected components. Furthermore, if* j *and* $2m$ *do not belong to the same cycle in* γ, *then* $\mathcal{C}'$ *is connected.*

Proof. Assume first that j and $2m$ belong to the same cycle in γ. This means that $w_j = w_{2m}$. If we erase the edge containing the half-edges $2m$ and j, $\mathcal{C}$ stays connected. However, when we cut the map along the path η, we may separate the map into two connected components. More precisely, we separate the map into two connected components if and only if η is homologically trivial, that is, the boundary of a surface embedded in S.

If j and $2m$ belong to different cycles, that is $w_j \neq w_{2m}$, then when removing the edge we may disconnect the two vertices but we then merge them. Consequently, the map $\mathcal{C}'$ is connected. ■

Using the permutations $\gamma' = \gamma_{\mathcal{C}'}$ and $\phi_{\mathcal{C}}$, and Lemma 4.17, we can now compute the genus of $\mathcal{C}'$. We recall Euler's formula (3.3) for a map of genus $g_{\mathcal{C}}$ with $C_{\mathcal{C}}$ connected components

$$2C_{\mathcal{C}} - 2g_{\mathcal{C}} = c(\gamma_{\mathcal{C}}) + c(\phi_{\mathcal{C}}) - m - r.$$

We now discuss the several cases that can occur. First, if both j and $2m$ are fixed points of $\phi_{\mathcal{C}}$ then it means that $\mathcal{C}$ is reduced to a vertex with 2 half-edges. We will assume in what follows that $m \geq 2$.

If j (respectively, $2m$) is a fixed point of $\phi_{\mathcal{C}} = \gamma_{\mathcal{C}}\pi_{\mathcal{C}}^{-1}$, then it means in terms of map that the face at the left of the white half-edge j (resp., $2m$) is at the left of the half-edge j (resp., $2m$) only. In terms of permutations, we have $\gamma_{\mathcal{C}}(2m) = \phi_{\mathcal{C}}\pi_{\mathcal{C}}(2m) = j$ (resp., $\gamma_{\mathcal{C}}(j) = 2m$). Then, $(j\ 2m)\gamma_{\mathcal{C}}(2m) = 2m$ (resp., $(j\ 2m)\gamma_{\mathcal{C}}(j) = j$), and when taking the trace on I_j, we remove one cycle of $(j\ 2m)\gamma_{\mathcal{C}}$. Furthermore, if we remove the disk corresponding to the face at the left of the half-edge j (resp., $2m$), the resulting map $\mathcal{C}'$ is connected.

If j and $2m$ are not fixed points of $\phi_{\mathcal{C}}$, then none of the connected component is reduced to a vertex without edges. It implies that the total number of faces and cycles of the associated permutation stays the same. We have $c(\phi_{\mathcal{C}}) = c(\phi_{\mathcal{C}'})$. The resulting map $\mathcal{C}'$ has one or two connected components: by cutting the map we either remove a handle (and decreased the genus by one) or disconnect the map.

It gives us one particular case (the degenerate case where one connected component is reduced to a vertex without edges):

(1) if j or $2m$ is a fixed point of $\phi_{\mathcal{C}}$, then $c(\phi_{\mathcal{C}'}) = c(\phi_{\mathcal{C}}) - 1$, $c(\gamma') = c(\gamma)$, and $\mathcal{C}'$ is connected. Thus, $g_{\mathcal{C}'} = g_{\mathcal{C}}$.

If both j and $2m$ are not fixed points of $\phi_{\mathcal{C}}$, we have three cases:

(2) if j and $2m$ belong to the same cycle of $\gamma_{\mathcal{C}}$ and $\mathcal{C}'$ is connected, then $c(\gamma_{\mathcal{C}'}) = c(\gamma_{\mathcal{C}}) + 1$, $c(\phi_{\mathcal{C}'}) = c(\phi_{\mathcal{C}})$, and $g' = g - 1$ (by Euler's formula),

(3) if j and $2m$ belong to the same cycle of $\gamma_{\mathcal{C}}$, and $\mathcal{C}'$ has two connected components, then $c(\gamma_{\mathcal{C}'}) = c(\gamma_{\mathcal{C}}) + 1$, $c(\phi_{\mathcal{C}'}) = c(\phi_{\mathcal{C}})$, and $g' = g$,

(4) if j and $2m$ belong to two different cycles of $\gamma_{\mathcal{C}}$, then $c(\gamma_{\mathcal{C}'}) = c(\gamma_{\mathcal{C}}) - 1$, $c(\phi_{\mathcal{C}'}) = c(\phi_{\mathcal{C}})$, and $g' = g$.

4.1.2. Second case: $\tau_r(2m) \neq 2m$. In this case, the white half-edge labeled $2m$ is connected to a black vertex. Let $j = \tau_r(2m) \in \varepsilon^{-1}(+1)$ (as the support of the transpositions τ_i is contained in $\varepsilon^{-1}(+1)$). In that case, the last black vertex has an outgoing half-edge labeled by $2m$ by Lemma 3.18. Similarly, the last black vertex has an outgoing half-edge labeled j.

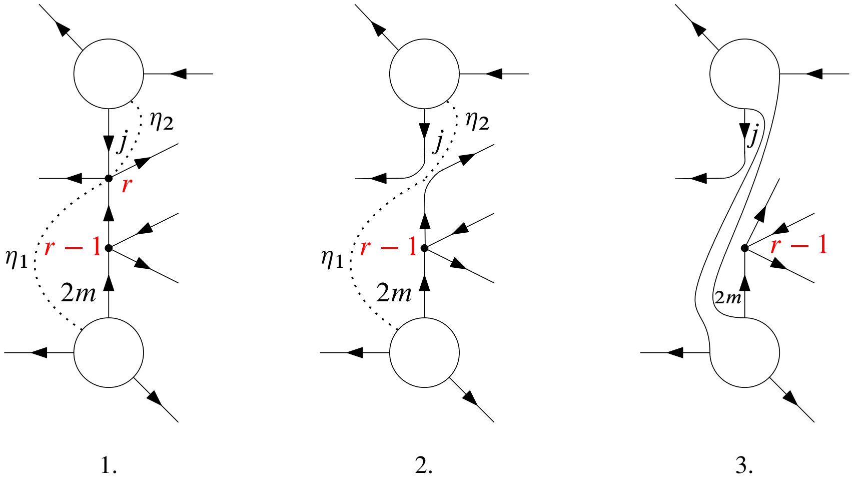

Figure 13. Second way to cut the map.

We construct a unitary map $\mathcal{C}'$ from $\mathcal{C}$ using the following procedure, depicted in Figure 13. Notice that the first step is possible as there is no loop with black edges, as indicated in Remark 3.16.

(1) We choose two paths η_1 and η_2 contained, respectively, in the face f_{2m} at the left of the half-edge $2m$, and f_j at the left of the half-edge j. The path η_1 (respectively, η_2) is chosen to start from the white vertex w_{2m} (respectively, w_j), attached in the corner at the left of the half-edge $2m$ (respectively, j), and end at the rth black vertex, attached in the corner at the left of the half-edge labeled $2m$ (respectively, j).

(2) We remove the rth black vertex, and attach each ingoing edge to the outgoing edge that follows it in the counterclockwise order.

(3) We cut the surface S along $\eta = \eta_1 \cup \eta_2$.

Lemma 4.18. *If $\mathcal{C}$ is a map of unitary type with labeling set I and with r black vertices, then $\mathcal{C}'$ is a map of unitary type with labeling set I and with $r-1$ black vertices. Furthermore, if $\mathcal{C}$ is non-decreasing, then so is $\mathcal{C}'$.*

Proof. As in the proof of Lemma 4.12, we first prove that $\mathcal{C}'$ is a map. We have to check that the faces are homeomorphic to disks. When removing the black vertex, at step 2, we may have connected two faces together, or may have connected a face to itself, thus creating a "face" homeomorphic to an annulus. However, when we cut the map, at step 3, we recover one or two faces homeomorphic to disks. Thus, $\mathcal{C}'$ is a map.

We now show that $\mathcal{C}'$ is indeed a map of unitary type. The map $\mathcal{C}'$ has $r-1$ black vertices. We removed the last black vertex and did not create any new edge linking two black vertices, thus properties 1 and 3 of Definition 3.13 are satisfied. We did not remove any white half-edge so property 2 is satisfied as well. Thus, $\mathcal{C}'$ is of unitary type. Furthermore, as we removed the last black vertex and let the other ones unchanged, if $\mathcal{C}$ is non-decreasing (recall Definition 3.20), then $\mathcal{C}'$ is non-decreasing as well. ■

Let us now compute the permutations that represent $\mathcal{C}'$.

Lemma 4.19. *Let $\pi' = (j\ 2m)\pi_{\mathcal{C}} = \tau_r \pi_{\mathcal{C}}$. We have $\pi_{\mathcal{C}'} = \pi'$.*

Proof. We only modify the edges during step 2, when we remove the black vertex. The outgoing half-edges of the rth black vertex in $\mathcal{C}$ are labeled on the left by j and $2m$. These half-edges are part of edges connected at their other end to white vertices, because the black vertex we remove is the last. These edges are connected, respectively, to the half edge $\pi_{\mathcal{C}}^{-1}(j)$ and $\pi_{\mathcal{C}}^{-1}(2m)$.

Consider the white half-edge labeled $\pi_{\mathcal{C}}^{-1}(j)$. After the surgery, it is connected to the half-edge labeled $2m$. Similarly, the white half-edge labeled $\pi_{\mathcal{C}}^{-1}(2m)$ is attached to the half-edge labeled j.

This corresponds to having $\pi_{\mathcal{C}'}(\pi_{\mathcal{C}}^{-1}(j)) = 2m$ and $\pi_{\mathcal{C}'}(\pi_{\mathcal{C}}^{-1}(2m)) = j$, and $\pi_{\mathcal{C}'} = \pi_{\mathcal{C}}$ for all other values. We can write this as $\pi_{\mathcal{C}'} = (j\ 2m)\pi_{\mathcal{C}}$. ■

Note that

$$\pi'^{(\varepsilon)} = (\tau_r \pi_{\mathcal{C}} \tau_r \pi_{\mathcal{C}})|_{\varepsilon^{-1}(+1)} = (\tau_r \pi_{\mathcal{C}}^2)|_{\varepsilon^{-1}(+1)} = \tau_r \pi_{\mathcal{C}}^{(\varepsilon)} = \tau_{r-1} \cdots \tau_1.$$

The first and third equalities are Definition 2.9, the second one follows from the fact that $\pi_{\mathcal{C}}(\varepsilon^{-1}(+1)) = \varepsilon^{-1}(-1)$ and the fact that the support of τ_r is contained in $\varepsilon^{-1}(+1)$, the fourth one is a consequence of Proposition 3.26. This is coherent with the fact that

$$\tau_{\mathcal{C}'} = (\tau_1, \dots, \tau_{r-1}). \tag{4.4}$$

Lemma 4.20. *Let $\gamma' = \gamma(j\ 2m) = \gamma\tau_r$. We have $\gamma_{\mathcal{C}'} = \gamma'$.*

Proof. The white vertices are only modified when we cut the map, at step 3. The proof is similar to the one of Lemma 4.15. We consider the two cases of j and $2m$ in the same cycle in γ or not, and we compute $\gamma_{\mathcal{C}} = \gamma\tau_r$. ■

It follows from Lemmas 3.25, 4.19, and 4.20 that

$$\phi_{\mathcal{C}'} = \tau_r \phi_{\mathcal{C}} \tau_r. \tag{4.5}$$

In particular, $c(\phi_{\mathcal{C}}) = c(\phi_{\mathcal{C}'})$.

We can now state the counterpart of Lemma 4.17.

Lemma 4.21. *If the unitary type map $\mathcal{C}$ is connected, then the map $\mathcal{C}'$ has one or two connected components. Furthermore, if j and $2m$ do not belong to the same cycle in γ, then $\mathcal{C}'$ is connected.*

Proof. The proof is almost the same as for Lemma 4.17. Alternatively, we can prove this lemma using Proposition 3.29.

Let $k \geq 1$ be the number of orbits of the action of $G' := G(\gamma', \pi', \tau')$ on $[2m]$. We notice that $G := G(\gamma, \pi_{\mathcal{C}}, \tau_{\mathcal{C}}) = \langle \gamma', \pi', \tau_1, \ldots, \tau_r \rangle$. In particular, j and $2m$ always belong to the same orbit of G. If j and $2m$ do not belong to the same orbit of G', τ_r connects two orbits of the action of G', and G has $k-1$ orbits. Conversely, if j and $2m$ belong to the same orbit of G', the actions of the two groups have the same number k of orbits.

In particular, if j and $2m$ belong to the same cycle of γ' (or equivalently to different cycles of γ), then the two groups have the same number of orbits. We assumed that $\mathcal{C}$ is connected so by Proposition 3.29, the action of G has one orbit. Thus, the action of G' has one orbit and $\mathcal{C}'$ is connected.

In the other cases, G has k or $k-1$ orbits and necessarily, k is 1 or 2. ■

Using Lemmas 4.21, 4.20, and (4.5), we can compute the genus g' of $\mathcal{C}'$ using Euler's formula (3.3). There are three cases:

- if j and $2m$ belong to the same cycle in γ and $\mathcal{C}'$ has two connected components, then $c(\gamma') = c(\gamma) + 1$ and $g' = g$,
- if j and $2m$ belong to the same cycle in γ and $\mathcal{C}'$ is connected, then $c(\gamma') = c(\gamma) + 1$ and $g' = g - 1$,
- if j, $2m$ belong to two different cycles in γ, then $\mathcal{C}'$ is connected (Lemma 4.21), $c(\gamma') = c(\gamma) - 1$, and $g = g'$.

Note that in these three cases, m is unchanged.

4.2. Proof of Theorem 4.1

We can now turn to the proof of Theorem 4.1.

Proof. Fix $\gamma = \gamma_{\boldsymbol{P}} \in \mathfrak{S}_{2m}$, $\varepsilon = \varepsilon_{\boldsymbol{P}}$, and $\boldsymbol{M} = \boldsymbol{M}_{\boldsymbol{P}}$. Assume first that $m = \frac{1}{2} \deg \boldsymbol{P} \geq 2$. We decompose the sum

$$\mathcal{M}_{0,l}^{(g),N}(P_1, \ldots, P_l u) = (-1)^m \sum_{\substack{\mathcal{C} \in \mathfrak{C}(g,[2m],\varepsilon,\gamma) \\ \mathcal{C} \text{ connected}}} (-1)^{c(\phi_{\mathcal{C}})} \operatorname{tr}_{\phi_{\mathcal{C}}}(\boldsymbol{M})$$

in two sums, each corresponding to one of the two cases of the previous construction.

We introduce the set W_{2m}^f of monotone walks whose last step τ satisfies $\tau(2m) = 2m$, and the set W_{2m}^c of monotone walks whose last step τ satisfy $\tau(2m) \neq 2m$. The functions $\mathbb{1}_{W_{2m}^f}$ and $\mathbb{1}_{W_{2m}^c}$ are the indicator functions of those sets.

The sum corresponding to the first case is thus by the previous surgery of Section 4.1

$$\begin{aligned}
&(-1)^m \sum_{\substack{\mathcal{C}\in\mathfrak{C}(g,m,\varepsilon,\gamma)\\ \mathcal{C}\text{ connected}}} (-1)^{c(\phi_{\mathcal{C}})} \operatorname{tr}_{\phi_{\mathcal{C}}}(\boldsymbol{M})\mathbb{1}_{W_{2m}^f}(\tau_{\mathcal{C}})\\
&= (-1)^m \sum_{\substack{\pi\in\mathfrak{S}_{2m}^{(\varepsilon)}\\ \tau\in\overrightarrow{\mathcal{W}}^{r(g,m,\gamma,\pi)}(\pi^{(\varepsilon)})\\ G(\gamma,\pi,\tau)\text{ is transitive}}} (-1)^{c(\gamma\pi^{-1})} \operatorname{tr}_{\gamma\pi^{-1}}(\boldsymbol{M})\mathbb{1}_{W_{2m}^f}(\tau)\\
&= (-1)^m \sum_{j\in\varepsilon^{-1}(-1)} \sum_{\substack{\pi'\in\mathfrak{S}^{(\varepsilon)}(I_j)\\ \pi=(j\ 2m)\pi'\\ \tau\in\overrightarrow{\mathcal{W}}^{r(g,m,\gamma,\pi)}(\pi'^{(\varepsilon)})\\ G(\gamma,\pi,\tau)\text{ is transitive}}} (-1)^{c(\gamma\pi^{-1})} \operatorname{tr}_{\gamma\pi^{-1}}(\boldsymbol{M}),
\end{aligned}$$

where $r(g,m,\gamma,\pi) = c(\gamma) + c(\gamma^{-1}\pi^{-1}) - m + 2g - 2$ according to (3.3), and we used the fact that in the first case π can be rewritten as $\pi'(j\ 2m)$ for some $j \in \varepsilon^{-1}(-1)$. Notice that the global factor $(-1)^m$ will account for a sign -1, when removing an edge and going from $2m$ white half-edges to $2m-2$ white half-edges.

We rewrite this as a sum of four terms, corresponding to the different ways of computing the genus, as explained in the last section. We interpret the new sums as series $\mathcal{M}_{g',l',0}^N(Q_1,\dots,Q_{l'})$, with l' (which corresponds to the number of vertices in the new map $\mathcal{C}'$) and g' (the genus of the new map $\mathcal{C}'$) two integers, and $Q_1,\dots,Q_{l'}$ monomials either in $\mathcal{X}$ or of degree 0. We introduce the notation $\boldsymbol{Q} = (Q_1,\dots,Q_{l'})$. These monomials are chosen so that the combinatorial data γ', and ε' described in the last section, and the tuple $\boldsymbol{M'}$ of appropriate monomials of degree 0 is such that $\gamma_{\boldsymbol{Q}} = \gamma'$, $\boldsymbol{\varepsilon_q} = \varepsilon'$, and $\boldsymbol{M_Q} = \boldsymbol{M}'$. The tuple $\boldsymbol{M'}$ is chosen differently depending on the subcase, but always so that $\operatorname{tr}_{\phi_{\mathcal{C}}}(\boldsymbol{M}) = \operatorname{tr}_{\mathcal{C}'}(\boldsymbol{M'})$ (except for subcase (1), see below).

There are four cases. Let us consider first the terms corresponding to subcases (1) and (3):

(1) if j or $2m$ is a fixed point of $\phi_{\mathcal{C}}$, then $c(\phi_{\mathcal{C}'}) = c(\phi_{\mathcal{C}}) - 1$, $l' = c(\gamma_{\mathcal{C}'}) = c(\gamma_{\mathcal{C}}) = l$, and $\mathcal{C}'$ is connected. Thus, $g' = g$,

(3) if j and $2m$ belong to the same cycle of $\gamma_{\mathcal{C}}$, and $\mathcal{C}'$ has two connected components, then $l' = c(\gamma_{\mathcal{C}'}) = c(\gamma_{\mathcal{C}}) + 1 = l + 1$, $c(\phi_{\mathcal{C}'}) = c(\phi_{\mathcal{C}})$, and $g' = g$.

In these two cases, the map $\mathcal{C}$ is cut into two maps, with total genus equal to g. Case (1) corresponds to the degenerate case where one of the two maps has no edges, and is reduced to a vertex. We associate to it the weight $\mathrm{tr}(M_j)$ or $\mathrm{tr}(M_{2m})$.

Together, these cases account for the term

$$-\sum_{P_l u = Qu^{-1}Ru} \sum_{\substack{g_1+g_2=g \\ I\subset[l-1]}} \mathcal{M}_{0,|I|+1}^{(g_1),N}(\boldsymbol{P}|_I, Q)\mathcal{M}_{0,|I^c|+1}^{(g_2),N}(\boldsymbol{P}|_{I^c}, R).$$

Subcase (1) corresponds to the term for which Q or R in the sum is reduced to a monomial of degree 0, and the subcase (3) to the other terms. When cutting the map, we obtain two connected components, each containing a vertex corresponding to part of P_l. This corresponds to the fact that in the argument of the series, P_l is replaced by two monomials Q and R such that $Qu^{-1}Ru$, and one u and one u^{-1} are removed, corresponding to the two removed half-edges.

Similarly, the subcase

(2) If j and $2m$ belong to the same cycle of $\gamma_{\mathcal{C}}$, and $\mathcal{C}'$ is connected, then

$$\begin{aligned} l' &= c(\gamma_{\mathcal{C}'}) = c(\gamma_{\mathcal{C}}) + 1 = l + 1, \\ c(\phi_{\mathcal{C}'}) &= c(\phi_{\mathcal{C}}), \\ g' &= g - 1 \end{aligned}$$

(by Euler's formula)

corresponds to the term

$$-\sum_{P_l u = Qu^{-1}Ru} \mathcal{M}_{0,l+1}^{(g-1),N}(P_1, \dots, P_{l-1}, Q, R).$$

The subcase

(4) If j and $2m$ belong to two different cycles of $\gamma_{\mathcal{C}}$, then

$$\begin{aligned} l' &= c(\gamma_{\mathcal{C}'}) = c(\gamma_{\mathcal{C}}) - 1 = l - 1, \\ c(\phi_{\mathcal{C}'}) &= c(\phi_{\mathcal{C}}), \\ g' &= g \end{aligned}$$

corresponds to the term

$$-\sum_{i=1}^{l-1} \sum_{P_i = Qu^{-1}R} \mathcal{M}_{0,l-1}^{(g),N}(P_1, \dots, P_{i-1}, P_{i+1}, \dots, P_{l-1}, RQP_l).$$

Here, two vertices are glued together, corresponding to replacing two polynomials in the argument of the series by one: RQP_l.

We proceed similarly for the terms that correspond to the second case, $\tau(2m) \neq 2m$. The corresponding sum is

$$
\begin{aligned}
&(-1)^m \sum_{\substack{\mathcal{C} \in \mathfrak{C}(g,[2m],\varepsilon,\gamma) \\ \mathcal{C} \text{ connected}}} (-1)^{c(\phi_{\mathcal{C}})} \operatorname{tr}_{\phi_{\mathcal{C}}}(\boldsymbol{M}) \mathbb{1}_{W_{2m}^c}(\tau_{\mathcal{C}}) \\
&= (-1)^m \sum_{\substack{\pi \in \mathfrak{S}_{2m}^{(\varepsilon)} \\ \boldsymbol{\tau} \in \overrightarrow{\mathcal{W}}^{r(g,m,\gamma,\pi)}(\pi^{(\varepsilon)}) \\ G(\gamma,\pi,\boldsymbol{\tau}) \text{ is transitive}}} (-1)^{c(\gamma\pi^{-1})} \operatorname{tr}_{\gamma\pi^{-1}}(\boldsymbol{M}) \mathbb{1}_{W_{2m}^c}(\boldsymbol{\tau}) \\
&= (-1)^m \sum_{\substack{j \in \varepsilon^{-1}(+1) \\ j \neq 2m}} \sum_{\substack{\pi' \in \mathfrak{S}_{2m}^{(\varepsilon)} \\ (\tau_1,\ldots,\tau_{r-1}) \in \overrightarrow{\mathcal{W}}^{r-1}(\pi'^{(\varepsilon)}) \\ G(\gamma',\pi',\boldsymbol{\tau}) \text{ is transitive}}} (-1)^{c(\gamma'\pi'^{-1})} \operatorname{tr}_{\gamma'\pi'^{-1}}(\boldsymbol{M}_{(j\,2m)}),
\end{aligned}
$$

where we use $\gamma' = (j\ 2m)\gamma$, $\boldsymbol{M}_{(j\,2m)} = (M_{(j\,2m)(1)}, M_{(j\,2m)(2)}, \ldots, M_{(j\,2m)(2m)})$, $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_{r-1}, (j\ 2m))$, and $r = r(g, m, \gamma, (j\ 2m)\pi')$. To go from the second to the third line, we replaced π by $\pi' = (j\ 2m)\pi$.

Following the construction from last section, we get three kinds of terms corresponding to the three subcases from last section. The first subcase is

(1) if j and $2m$ belong to the same cycle in γ and $\mathcal{C}'$ has two connected components, then $c(\gamma') = c(\gamma) + 1$ and $g' = g$, corresponding to the sum

$$
\sum_{P_l u = QuRu} \sum_{\substack{g_1+g_2=g \\ I \subset [2m]}} \mathcal{M}_{0,|I|+1}^{(g_1),N}(\boldsymbol{P}|_I, Qu) \mathcal{M}_{0,|I^c|+1}^{(g_2),N}(\boldsymbol{P}|_{I^c}, Ru).
$$

The second subcase is

(2) if j and $2m$ belong to the same cycle in γ and $\mathcal{C}'$ is connected, then $c(\gamma') = c(\gamma) + 1$ and $g' = g - 1$, corresponding to

$$
\sum_{P_l u = QuRu} \mathcal{M}_{0,l+1}^{(g-1),N}(P_1, \ldots, P_{l-1}, Qu, Ru).
$$

Finally, the last subcase is

(3) if $j, 2m$ belong to two different cycles in γ, then $\mathcal{C}'$ is connected (Lemma 4.21), $c(\gamma') = c(\gamma) - 1$, and $g = g'$, corresponding to

$$
\sum_{i=1}^{l-1} \sum_{P_i = QuR} \mathcal{M}_{0,l-1}^{(g),N}(P_1, \ldots, P_{i-1}, P_{i+1}, \ldots, P_{l-1}, RQuP_l u).
$$

Putting all the terms together, we get the induction relation of Theorem 4.1.

∎

4.3. Induction relation for the series $\mathcal{M}^{(g),N}_{V,l}$

We first prove that the series $\mathcal{M}^{(g),N}_{V,l}(\boldsymbol{P})$ exist with a radius of convergence that depends on g, l, and V. To that end, we show bounds on the series of maps for $V = 0$ that are a consequence of Theorem 4.1. Similar bounds have been obtained in the Gaussian case in [31, Lemma 4.3].

Proposition 4.22. *Assume that for all $N \geq 1$ and all $1 \leq i \leq p$ we have the bound $\|A^N_i\| \leq 1$ on the deterministic matrices (A^N_i). Let $\boldsymbol{q} = (q_1, \dots, q_k) \in \mathcal{X}^k_n$ be monomials, and $\nu = \max_{1 \leq i \leq k} \deg q_i$. We introduce the nth Catalan number $c_n = \frac{1}{n+1}\binom{2n}{n}$.*

There exist constants $A_k > 1$, $B_k > 1$, and $C_k > 1$ depend on k, and $D_{k,\nu} > 1$ that depends on k and ν such that for all $\boldsymbol{P} = P_1, \dots, P_l \in \mathcal{X}^l_n$, and all $\boldsymbol{n} = (n_1, \dots, n_k) \in \mathbb{N}^k$, we have the bound

$$\frac{1}{\boldsymbol{n}!}\Big|\mathcal{M}^{(g),N}_{0,\sum_i n_i + l}(\underbrace{q_1, \dots, q_1}_{n_1 \text{ times}}, \dots, \underbrace{q_k, \dots, q_k}_{n_k \text{ times}}, P_1, \dots, P_l)\Big|$$
$$\leq A_k^{l(2m+\nu\boldsymbol{n})} B_k^{-l} C^{g(2m+\nu\boldsymbol{n})} D^{\boldsymbol{n}} \prod_i c_{\deg P_i} \prod_{j=1}^{k} c_{n_j}, \tag{4.6}$$

where $m = \frac{1}{2}\deg \boldsymbol{P}$.

The constants can be chosen to be

$$\begin{aligned} A_k &= C_k = \sqrt{6}\pi^{1/4} 2^{k+3}, \\ B_k &= 3 \cdot 4^{k+1}, \\ D_{k,\nu} &= 4k(4\mathrm{e}^{1/\mathrm{e}})^{\nu}. \end{aligned}$$

The proof is given in Appendix A. The value of the constants can be improved. These bounds allow us to prove immediately that the series $\mathcal{M}^{(g),N}_{V,l}$ (see Definition 3.38) converge.

Corollary 4.23. *Suppose that $\boldsymbol{P} = (P_1, \dots, P_l) \in \mathcal{X}^l_n$, $\boldsymbol{q} = (q_1, \dots, q_k) \in \mathcal{X}^k_n$, and $\boldsymbol{z} = (z_1, \dots, z_k)$, and let $V = \sum_{i=1}^k z_i q_i$ be a potential.*

As a series in $\boldsymbol{z}$, $\mathcal{M}^{(g),N}_{V,l}(P_1, \dots, P_l)$ converges absolutely with radius of convergence $R_{l,g,V} \geq (4A_k^{l+g} D_{k,\nu})^{-1}$.

We can now turn to the induction relations. The induction relation from Theorem 4.1 translates to an induction relation on the series $\mathcal{M}^{(g),N}_{V,l}$.

Proposition 4.24. *Let $\boldsymbol{P} = (P_1, \dots, P_l) \in (\mathcal{X}_n)^l$, $\boldsymbol{q} = (q_1, \dots, q_k) \in (\mathcal{X}_n)^k$ and $\boldsymbol{z} = (z_1, \dots, z_k)$, and let $V = \sum_{i=1}^k z_i q_i$ be a potential. Assume that for all $1 \leq i \leq k$, $|z_i| < R_{l,g,V}$.*

For all $1 \leq i \leq n$, we have the equation

$$\sum_{\substack{g_1+g_2=g \\ I\subset[l-1]}} \mathcal{M}^{(g_1),N}_{V,|I|+1} \otimes \mathcal{M}^{(g_2),N}_{V,|I^c|+1}(\boldsymbol{P}_I \otimes \boldsymbol{P}_{I^c} \sharp \partial P_l)$$
$$+ \mathcal{M}^{(g),N}_{V,l}(P_1 \otimes \cdots \otimes P_{l-1} \otimes (\mathcal{D}V)P_l)$$
$$= -\mathcal{M}^{(g-1),N}_{V,l+1}(P_1 \otimes \cdots \otimes P_{l-1} \otimes \partial P_l)$$
$$- \sum_{j=1}^{l-1} \mathcal{M}^{(g),N}_{V,l-1}(P_1 \otimes \cdots \otimes P_{j-1} \otimes P_{j+1} \otimes \cdots \otimes P_{l-1} \otimes (\mathcal{D}P_j)P_l). \tag{4.7}$$

Proof. We sum on $\boldsymbol{n} \in \mathbb{N}^k$ the induction relations of Proposition 5.10 for

$$\mathcal{M}^{(g),N}_{0,\sum_i n_i+l}(\underbrace{q_1,\ldots,q_1}_{n_1 \text{ times}},\ldots,\underbrace{q_k,\ldots,q_k}_{n_k \text{ times}}, P_1,\ldots,P_l),$$

times $\frac{\boldsymbol{z}^{\boldsymbol{n}}}{\boldsymbol{n}!}$. ∎

5. The multimatrix case

Up to now, we have only considered integrals involving one Haar-distributed matrix U^N. The results obtained so far can be extended in a straightforward way to the case where we have $n \geq 1$ independent Haar-distributed matrices $U_1^N,\ldots,U_n^N$. The polynomials we consider in the sequel are the non-commutative polynomials in u_i, u_i^{-1}, for $1 \leq i \leq n$ and a_j, a_j^* for $1 \leq j \leq p$. We denote this ast-algebra by $\mathcal{A}_n$. Notice that $\mathcal{A} = \mathcal{A}_1$.

5.1. Weingarten calculus

As previously, we will consider a subset of monomials of $\mathcal{A}_n$, as the quantity we consider are multilinear functions which are tracial in each of their arguments. We define $\mathcal{X}_n$ the set of monomials of $\mathcal{A}_n$ of the form

$$P = M_1 u_{t_1}^{\varepsilon_1} M_2 u_{t_2}^{\varepsilon_2} \cdots M_d u_{t_d}^{\varepsilon_d}, \tag{5.1}$$

where $\varepsilon\colon [d] \to \{+1,-1\}$, $t\colon [d] \to [n]$, and $\boldsymbol{M} = (M_1,\ldots,M_d)$ is a d-tuples of monomials $M_j \in \mathcal{A}_n$, each of them being empty or a word in $a_1, a_1^*,\ldots,a_p,a_p^*$.

We define for a tuple $\boldsymbol{P} = (P_1,\ldots,P_l)$ the tuples $\boldsymbol{\varepsilon_P}, \boldsymbol{t_P}, \boldsymbol{M_P}$ obtained by concatenating the tuples corresponding to each polynomial P_j, $1 \leq j \leq l$. We also define for $1 \leq i \leq n$, $\boldsymbol{\varepsilon}_{\boldsymbol{P},i} = \boldsymbol{\varepsilon}|_{\boldsymbol{t}_{\boldsymbol{P}}^{-1}(i)}$, i.e., the tuple which encodes the exponents of the variables u_i only. We define the degree with respect to u_i of a monomial P, $\deg_i P$ as the number of occurrences of u_i or u_i^{-1} in P. The total degree of a tuple is $\deg_i \boldsymbol{P} = \sum_{j=1}^{l} \deg_i P_j$. We define $\deg P = \sum_{i=1}^{n} \deg_i P$, and $\deg \boldsymbol{P} = \sum_{j=1}^{l} \deg P_j$.

Proposition 2.11 is generalized as follows.

Definition 5.1. Let $\boldsymbol{P} = (P_1, \dots, P_l) \in (\mathcal{X}_n)^l$. We define the permutation

$$\gamma_{\boldsymbol{P}} = (1 \dots \deg P_1) \cdots \left(\sum_{j=1}^{l-1} \deg P_j + 1 \dots \deg \boldsymbol{P} \right).$$

Definition 5.2. We introduce the moment with respect to the Haar measure in the multimatrix case

$$\alpha^N_{\boldsymbol{U},0,l}(P_1, \dots, P_l) = \mathbb{E}[\mathrm{Tr}(P_1) \cdots \mathrm{Tr}(P_l)],$$

where the expectation is under the product Haar measure $\mathrm{d}U_1^N \cdots \mathrm{d}U_n^N$.

Proposition 5.3. *Let $\boldsymbol{P} = (P_1, \dots, P_l) \in (\mathcal{X}_n)^l$ and $J_i = \boldsymbol{t}_{\boldsymbol{P}}^{-1}(i)$.*

We have

$$\alpha^N_{\boldsymbol{U},0,l}(P_1, \dots, P_l) = \sum_{\substack{\pi_1 \in \mathfrak{S}^{(\boldsymbol{\varepsilon}_{\boldsymbol{P}},1)}(J_1) \\ \pi_2 \in \mathfrak{S}^{(\boldsymbol{\varepsilon}_{\boldsymbol{P}},2)}(J_2) \\ \dots \\ \pi_n \in \mathfrak{S}^{(\boldsymbol{\varepsilon}_{\boldsymbol{P}},n)}(J_n)}} \left(\prod_{i=1}^{k} \mathrm{Wg}_N \left(\pi_i^{(\varepsilon_{\boldsymbol{P}}^{(i)})} \right) \right) \mathrm{Tr}_{\gamma_{\boldsymbol{P}} \pi_1^{-1} \dots \pi_n^{-1}}(\boldsymbol{M}_{\boldsymbol{P}}). \tag{5.2}$$

This proposition is obtained by applying the following lemma n times.

Lemma 5.4. *Let $\boldsymbol{M}_{\boldsymbol{P}} = (M_1, \dots, M_{\deg \boldsymbol{P}})$ and let $\widetilde{\boldsymbol{M}} = (\widetilde{M}_i, 1 \le i \le \deg \boldsymbol{P})$, defined by $\widetilde{M}_i = M_i$ if $t_i = 1$, and $\widetilde{M}_i = M_i u_{t_i}^{\varepsilon_{\boldsymbol{P}}(i)}$ otherwise. Then, we have the expectation with respect to U_1^N only*

$$\mathbb{E}_{U_1^N}[\mathrm{Tr}(P_1) \cdots \mathrm{Tr}(P_l)] = \sum_{\pi_1 \in \mathfrak{S}^{(\varepsilon_{\boldsymbol{P}},1)}(J_1)} \mathrm{Wg}_N \left(\pi_1^{(\varepsilon_{\boldsymbol{P}},1)} \right) \mathrm{Tr}_{\gamma_{\boldsymbol{P}} \pi_1^{-1}}(\widetilde{\boldsymbol{M}}).$$

Proof of Lemma 5.4. Let $I \subset [l]$ be the subset of indices i such that P_i contains a letter u_1 or u_1^*. Denote by c_i the cycle in γ that corresponds to P_i, i.e.,

$$c_i = \left(\sum_{j=1}^{i-1} \deg P_j + 1 \cdots \sum_{j=1}^{i-1} \deg P_j + \deg P_i \right).$$

We have

$$\mathbb{E}_{U_1^N}[\mathrm{Tr}(P_1) \cdots \mathrm{Tr}(P_l)] = \left(\prod_{i \notin I} \mathrm{Tr}_{c_i}(P_i) \right) \mathbb{E}\left[\prod_{i \in I} \mathrm{Tr}(P_i) \right]. \tag{5.3}$$

Furthermore, if we set $S = \bigcup_{i \notin I} \operatorname{Supp} c_i$ and $\gamma'' = (\prod_i c_i)|_S$, we can rewrite the terms

$$\Big(\prod_{i \notin I} \mathrm{Tr}_{c_i}(P_i)\Big) = \mathrm{Tr}_{\gamma''}(\widetilde{\boldsymbol{M}}|_S).$$

Considering only the second term in the right side of (5.3) and using the cyclic property of the trace, we can assume that the last factor of each polynomial P_i is a u_1 or a u_1^*. Let $J_1 = t^{-1}(1) = \{p_1 < p_2 < \cdots < p_q\}$. We let $\gamma' = \mathrm{Tr}(\gamma_{\boldsymbol{P}}; J_1)$. Proposition 2.11 shows that

$$\mathbb{E}_{U_1^N}[\mathrm{Tr}(P_1)\cdots\mathrm{Tr}(P_l)] = \sum_{\pi_1 \in \mathfrak{S}^{(\varepsilon_{\boldsymbol{P}},1)}(J_1)} \mathrm{Wg}_N\big(\pi_1^{(\varepsilon_{\boldsymbol{P}},1)}\big)\, \mathrm{Tr}_{\gamma'\pi_1^{-1}}(\boldsymbol{M}'),$$

where $\boldsymbol{M}' = (M_i', i \in J_1)$ is defined by

$$M_i' = M_{p_{i-1}+1} u_{t(p_{i-1}+1)}^{\varepsilon_{\boldsymbol{P}}(p_{i-1}+1)} M_{p_{i-1}+2} u_{t(p_{i-1}+2)}^{\varepsilon_{\boldsymbol{P}}(p_{i-1}+2)} \cdots M_{p_i-1} u_{t(p_i-1)}^{\varepsilon_{\boldsymbol{P}}(p_i-1)} M_{p_i},$$

with the convention $p_0 = 0$.

This is equal to

$$\mathbb{E}_{U_1^N}[\mathrm{Tr}(P_1)\cdots\mathrm{Tr}(P_l)] = \sum_{\pi_1 \in \mathfrak{S}^{(\varepsilon_{\boldsymbol{P}},1)}(J_1)} \mathrm{Wg}_N\big(\pi_1^{(\varepsilon_{\boldsymbol{P}},1)}\big)\, \mathrm{Tr}_{\gamma_{\boldsymbol{P}}\pi_1^{-1}}(\widetilde{\boldsymbol{M}}). \qquad \blacksquare$$

5.2. Multicolored maps of unitary type

We now generalize the notion of a map of unitary type to address the multimatrix case.

Definition 5.5. Let I be a finite subset of $\mathbb{N}^*$. A *multicolored map of unitary type* with n colors, with labels in I, and with r_i vertices of color i for $1 \leq i \leq n$, is an oriented map with vertices colored in white or in one of n colors, and colored half-edges which can be of any of the n colors such that

- there are r_i vertices of color i for $1 \leq i \leq n$, which are alternated of degree 4 and numbered from 1 to r_i;
- the half-edges connected to a vertex of color i (which is not white) are of color i as well;
- there are $|I|$ half-edges that are connected to white vertices. Each element of I labels exactly one of these half-edges;
- each half-edge connected to a white vertex is colored in one of the n colors;
- each edge is composed of two half-edges of the same color;
- if an oriented edge connects the vertex of color i numbered l_1 to the vertex of color i numbered l_2, then $l_1 < l_2$.

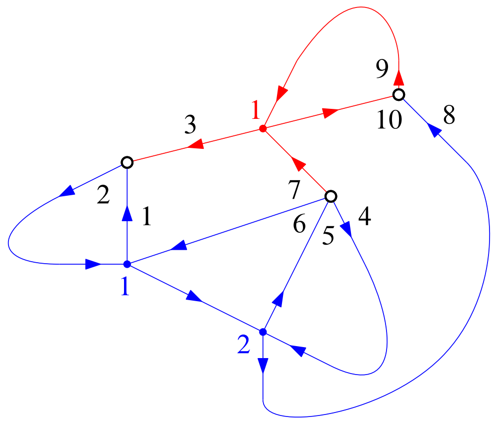

Figure 14. A muticolored map of unitary type with two colors. The integers 3, 7, 10, 9 label red half-edges, and the other elements of [10] label blue half-edges.

Remark 5.6. Notice that we can erase the colors of a multicolored unitary map to obtain a (monocolored) map of unitary type (see Figure 14). To do so, we proceed as follows:

- each colored half-edge connected to a white vertex becomes a white half-edge;
- each colored half-edge connected to a colored vertex becomes a black half-edge;
- each colored vertex becomes a black vertex.

The resulting map is a map of unitary type with $\sum_i r_i$ black vertices and labels in I.

As in Section 3.2, we define for a multicolored map $\mathcal{C}$ of unitary type, with n colors, the following permutations. We construct permutations $\gamma_{\mathcal{C}}$, $\pi_{\mathcal{C}}$, and $\phi_{\mathcal{C}}$, and the tuple $\boldsymbol{\varepsilon}_{\mathcal{C}}$, as for a monocolored map of unitary type. If the ith labeled half-edge is of color j, we set $\boldsymbol{t}_{\mathcal{C}}(i) = j$. We then define

$$J_{\mathcal{C},i} = \boldsymbol{t}_{\mathcal{C}}^{-1}(i).$$

We set

$$\boldsymbol{\varepsilon}_{\mathcal{C},i} = \boldsymbol{\varepsilon}_{\mathcal{C}}|_{J_{\mathcal{C},i}}$$

for all $1 \leq i \leq k$.

For each color $i \in [n]$, we consider the edges of this color. We then define as previously a permutation $\pi_{\mathcal{C},i} \in \mathfrak{S}^{(\boldsymbol{\varepsilon}_{\mathcal{C},i})}(J_{\mathcal{C},i})$ describing these edges and the vertices of color i. Finally, if we consider the vertices of color i, we can associate to the jth vertex of color i the transposition $\tau_{i,j}$ as previously. We set $\tau_{\mathcal{C}} = (\tau_{i,j}, i \in [n], j \in [r_i])$. Notice that by construction, we have $\pi_{\mathcal{C}} = \pi_{\mathcal{C},1}\pi_{\mathcal{C},2}\cdots\pi_{\mathcal{C},n}$.

Definition 5.7. Let $n \geq 1$ and $g \geq 0$ be integers, and I be a finite subset of the positive integers. Let $\boldsymbol{r} = (r_1, \ldots, r_n) \in \mathbb{N}^n$, $\gamma \in \mathfrak{S}(I)$, $\boldsymbol{t}: I \to [n]$ and $\boldsymbol{\varepsilon}: I \to \{\pm 1\}$.

We denote by $\mathfrak{C}^{\boldsymbol{r}}(I, \boldsymbol{\varepsilon}, \gamma, \boldsymbol{t})$ the set of multicolored maps of unitary type $\mathcal{C}$ with n colors, with label set I, and with r_i vertices of color i, for $1 \leq i \leq n$, such that $\gamma_{\mathcal{C}} = \gamma$, $\boldsymbol{\varepsilon}_{\mathcal{C}} = \boldsymbol{\varepsilon}$, and $\boldsymbol{t}_{\mathcal{C}} = \boldsymbol{t}$.

We denote by $\mathfrak{C}(g, I, \boldsymbol{\varepsilon}, \gamma, \boldsymbol{t})$ the set of multicolored maps of unitary type $\mathcal{C}$ with n colors, with label set I, and of genus g, such that $\gamma_{\mathcal{C}} = \gamma$, $\boldsymbol{\varepsilon}_{\mathcal{C}} = \boldsymbol{\varepsilon}$, and $\boldsymbol{t}_{\mathcal{C}} = \boldsymbol{t}$.

We then have the analog of Theorem 3.28

Theorem 5.8. *Let $n \geq 1$ be an integer, and I be a finite subset of the positive integers. Let $\boldsymbol{r} = (r_1, \ldots, r_n) \in \mathbb{N}^n$, $\gamma \in \mathfrak{S}(I)$, $\boldsymbol{t}: I \to [n]$ and $\boldsymbol{\varepsilon}: I \to \{\pm 1\}$. Define $J_i = \boldsymbol{t}^{-1}(i)$ and $\boldsymbol{\varepsilon}_i = \boldsymbol{\varepsilon}|_{J_i}$ for $1 \leq i \leq n$.*

The previous construction gives a bijection between $\mathfrak{C}^{\boldsymbol{r}}(I, \boldsymbol{\varepsilon}, \gamma, \boldsymbol{t})$ and

$$\bigcup_{\pi_1 \in \mathfrak{S}^{(\varepsilon_1)}(J_1), \ldots, \pi_n \in \mathfrak{S}^{(\varepsilon_n)}(J_n)} \prod_{i=1}^{n} \{\pi_i\} \times \overrightarrow{\mathcal{W}}^{r_i}\left(\pi_i^{(\varepsilon_i)}\right).$$

Proof. The proof is very similar to the one of Theorem 3.28. By considering each color, we prove that the construction does give a map

$$\mathfrak{C}^{\boldsymbol{r}}(I, \boldsymbol{\varepsilon}, \gamma, \boldsymbol{t}) \to \bigcup_{\pi_1 \in \mathfrak{S}^{(\varepsilon_1)}(J_1), \ldots, \pi_n \in \mathfrak{S}^{(\varepsilon_n)}(J_n)} \prod_{i=1}^{n} \{\pi_i\} \times \overrightarrow{\mathcal{W}}^{r_i}\left(\pi_i^{(\varepsilon_i)}\right).$$

We can construct its inverse exactly as in the proof for the case with one unitary matrix, by constructing the edges for the color 1, then for the color 2, etc. More precisely, we first consider the color 1. At this step, we leave untouched the half-edges of color $2, \ldots, n$. We construct the incidence relation for the edges of color 1 using the data of $(\tau_{1,j})_{j \in [r_1]}$ and π_1. When this is finished, all the half-edges of color 1 are part of some edge.

We then turn to the half-edges of color 2. Using the data of $(\tau_{2,j})_{j \in [r_2]}$ and π_2, we construct the incidence relation for the edges of color 2. We do this until we reach color n and obtain a multicolored map of unitary type. ∎

It follows directly by erasing the colors (see Remark 5.6) and Proposition 3.29 that we have the following proposition.

Proposition 5.9. *Let $\mathcal{C}$ be a multicolored map of unitary type with n colors. The map $\mathcal{C}$ is connected if and only if the group $\langle \gamma_{\mathcal{C}}, \pi_{\mathcal{C},1}, \ldots, \pi_{\mathcal{C},n}, \tau_{i,j}, 1 \leq i \leq n, 1 \leq j \leq r_i \rangle$ is transitive, with*

$$\tau_{\mathcal{C}} = \left(\tau_{i,j}, i \in [n], j \in [r_i]\right).$$

Using Proposition 2.14, 5.3, and Theorem 5.8, we can compute the moments with no potential (for $N \geq m = \frac{1}{2} \deg \boldsymbol{P}$). Let $\boldsymbol{P} = (P_1, \ldots, P_l) \in (\mathcal{X}_n)^l$, we have

$$\begin{aligned}
&\alpha^N_{\boldsymbol{U},0,l}(P_1, \ldots, P_l) \\
&= N^{-m} \sum_{\substack{\pi_1 \in \mathfrak{S}^{(\varepsilon_{\boldsymbol{P}},1)}(J_1) \\ \pi_2 \in \mathfrak{S}^{(\varepsilon_{\boldsymbol{P}},2)}(J_2) \\ \cdots \\ \pi_n \in \mathfrak{S}^{(\varepsilon_{\boldsymbol{P}},n)}(J_n)}} \sum_{r_1,\ldots,r_n \geq 0} \left(\frac{-1}{N}\right)^{\boldsymbol{r}} \prod_{i=1}^{n} \vec{w}^{r_i}\left(\pi_i^{(\varepsilon_i)}\right) \mathrm{Tr}_{\gamma_{\boldsymbol{P}} \pi_1^{-1} \cdots \pi_n^{-1}}(\boldsymbol{M_P}) \\
&= N^{-m} \sum_{\boldsymbol{r} \in \mathbb{N}^k} \sum_{\mathcal{C} \in \mathfrak{C}(\boldsymbol{r}_{[2m],\boldsymbol{\varepsilon_P},\gamma_{\boldsymbol{P}},\boldsymbol{t}})} \left(\frac{-1}{N}\right)^{\boldsymbol{r}} \mathrm{Tr}_{\gamma \pi_1^{-1} \cdots \pi_n^{-1}}(\boldsymbol{M_P}),
\end{aligned} \tag{5.4}$$

where we use the notation $x^{\boldsymbol{r}} = x^{\sum_i r_i}$ for any $x \in \mathbb{R}$.

We then compute the cumulants for no potential, when $N \geq m$,

$$\mathcal{W}^N_{\boldsymbol{U}^N,0,l}(P_1, \ldots, P_l) = N^{-m} \sum_{\boldsymbol{r} \in \mathbb{N}^n} \sum_{\substack{\mathcal{C} \in \mathfrak{C}^{\boldsymbol{r}}([2m],\boldsymbol{\varepsilon_P},\gamma_{\boldsymbol{P}},\boldsymbol{t}) \\ \mathcal{C} \text{ connected}}} \left(\frac{-1}{N}\right)^{\boldsymbol{r}} \mathrm{Tr}_{\phi_{\mathcal{C}}}(\boldsymbol{M_P}). \tag{5.5}$$

We now rewrite this sum using the genus of the maps rather than the number of colored vertices. In this context, the Euler formula becomes

$$2 - 2g_{\mathcal{C}} = c(\gamma_{\mathcal{C}}) + c(\phi_{\mathcal{C}}) - m - \sum_{i=1}^{k} r_i.$$

We thus get the renormalized cumulant $\widetilde{\mathcal{W}}^N_{\boldsymbol{U}^N,0,l}$,

$$\begin{aligned}
&\widetilde{\mathcal{W}}^N_{\boldsymbol{U}^N,0,l}(P_1, \ldots, P_l) \\
&= N^{l-2} \mathcal{W}^N_{\boldsymbol{U}^N,0,l}(P_1, \ldots, P_l) \\
&= (-1)^{m+l} \sum_{g \geq 0} \frac{1}{N^{2g}} \sum_{\substack{\mathcal{C} \in \mathfrak{C}(g,[2m],\boldsymbol{\varepsilon_P},\gamma_{\boldsymbol{P}},\boldsymbol{t}) \\ \mathcal{C} \text{ connected}}} (-1)^{c(\phi_{\mathcal{C}})} \mathrm{tr}_{\phi_{\mathcal{C}}}(\boldsymbol{M_P}).
\end{aligned} \tag{5.6}$$

Let us recall the relevant notation. Here, l is the number of monomials or the number of white vertices. In particular, we have $c(\gamma_{\boldsymbol{P}}) = l$. The set $\mathfrak{C}(g, [2m], \boldsymbol{\varepsilon_P}, \gamma_{\boldsymbol{P}}, \boldsymbol{t})$ was introduced in Definition 5.7. As in Section 3.4, the coefficients in the $1/N^2$ expansion are sums of maps, with a weight determined by the permutation of the faces $\phi_{\mathcal{C}}$ and the tuple of matrices $\boldsymbol{M}$.

The term of order $2g$ is then

$$\mathcal{M}^{(g),N}_{0,l}(P_1, \ldots, P_l) = (-1)^{m+l} \sum_{\substack{\mathcal{C} \in \mathfrak{C}(g,[2m],\boldsymbol{\varepsilon_P},\gamma_{\boldsymbol{P}},\boldsymbol{t}) \\ \mathcal{C} \text{ connected}}} (-1)^{c(\phi_{\mathcal{C}})} \mathrm{tr}_{\phi_{\mathcal{C}}}(\boldsymbol{M_P}).$$

We then define the formal cumulant as

$$\mathcal{M}^{(g),N}_{V,l}(P_1,\dots,P_l) = \sum_{\boldsymbol{n}\in\mathbb{N}^k} \frac{\boldsymbol{z}^{\boldsymbol{n}}}{\boldsymbol{n}!}\mathcal{M}^{(g),N}_{0,l}(\underbrace{q_1,\dots,q_1}_{n_1 \text{ times}},\dots,\underbrace{q_k,\dots,q_k}_{n_k \text{ times}},P_1,\dots,P_l),$$

as previously.

5.3. Induction relation

We will now deduce from the relations obtained in Section 4 similar relations in the multimatrix case.

Proposition 5.10. *Let $\boldsymbol{P} = (P_1,\dots,P_l) \in (\mathcal{X}_n)^l$, $i \in [n]$ and $g \geq 1$. If $\frac{1}{2}\deg_i \boldsymbol{P} \geq 2$, then we have the following equation:*

$$\begin{aligned}
&\sum_{\substack{g_1+g_2=g\\ I\subset[l-1]}} \mathcal{M}^{(g_1),N}_{V,|I|+1} \otimes \mathcal{M}^{(g_2),N}_{V,|I^c|+1}(\boldsymbol{P}_I \otimes \boldsymbol{P}_{I^c}\#\partial_i P_l)\\
&\qquad + \mathcal{M}^{(g),N}_{V,l}(P_1\otimes\cdots\otimes P_{l-1}\otimes(\mathcal{D}_i V)P_l)\\
&\quad = -\mathcal{M}^{(g-1),N}_{V,l+1}(P_1\otimes\cdots\otimes P_{l-1}\otimes\partial_i P_l)\\
&\qquad - \sum_{j=1}^{l-1}\mathcal{M}^{(g),N}_{V,l-1}(P_1\otimes\cdots\otimes P_{j-1}\otimes P_{j+1}\otimes\cdots\otimes P_{l-1}\otimes(\mathcal{D}_i P_j)P_l).
\end{aligned}\tag{5.7}$$

Here, ∂_i and $\mathcal{D}_i$ are the non-commutative and cyclic derivatives with respect to u_i for $i \in [n]$.

This proposition is proved as in Section 4. If no polynomial of $\boldsymbol{P}$ contains a u_i then the equation is trivial. Thus, we can assume by symmetry that $\deg_i P_l \geq 1$. We cut the maps from the sum $\mathcal{W}^{(g),N}_{0,l}(\boldsymbol{P})$ as in Section 4.1. Notice that in this construction, we only modify edges of the color i so we can use the exact same arguments. We thus obtain the wanted equation.

6. The Dyson–Schwinger equation and the topological expansion

We now work in the multi-matrix setting. All the maps involved will be multicolored maps. The induction equations obtained in Section 4 are related to Dyson–Schwinger equations for unitary matrices. In this section, we introduce the Dyson–Schwinger lattice of equations for the renormalized cumulants $\widetilde{\mathcal{W}}^N_{V,l} = N^{l-2}\mathcal{W}^N_{V,l}$. Together with the induction relations derived in Section 4, they allow us to show that the renormalized cumulants $\widetilde{\mathcal{W}}^N_{V,l}$ admit an asymptotic topological expansion as $N \to \infty$. The methods used in this section are heavily inspired from [23].

6.1. Scalar product and parametric norms on $\mathcal{A}_n$

Following [23], we introduce some useful notions. The vector space $\mathcal{A}$—the algebra of noncommutative polynomials—admits a countable basis, which is the set $\widehat{\mathcal{X}}_n$ of all words in the letters $u_1, u_1^*, \dots, u_n, u_n^*, a_1, a_1^*, \dots, a_p, a_p^*$, plus the empty word 1. Notice that this set contains $\mathcal{X}_n$, the set of such words that finishes by a u_i or a u_i^*. Let $\langle \cdot, \cdot \rangle$ be the scalar product that makes this basis orthonormal. In particular, $\mathcal{B}^\perp$ is the algebra generated by the polynomials with no constant term, i.e., without factors u_i or u_i^*.

Definition 6.1. Let $\xi \geq 1$. The ξ-norm is $\|\cdot\|_\xi$ defined by

$$\|P\|_\xi = \sum_{Q \in \widehat{\mathcal{X}}} |\langle P, Q \rangle| \xi^{\deg Q}$$

for $P \in \mathcal{A}$.

We write $\mathcal{B}_\xi^\perp$ the completion of the algebra $\mathcal{B}^\perp$ in the ξ-norm $\|\cdot\|_\xi$.

This norm is a deformation of the ℓ^1 norm that takes into account the degree of the basis monomials. The usual ℓ^1 norm will in many cases not be the appropriate norm, as the effect of many operators we consider in the sequel depends on the degree of the monomial it is applied to.

Example 6.2. This norm is deformation of the ℓ^1 norm, recovered when considering the 1-norm. For instance, the 1-norm of the potential we consider is

$$\|V\|_1 = \sum_{i=1}^{k} |z_i|.$$

This notion of norm allows us to define the parametric ξ-norm of a linear operator or form.

Definition 6.3. Let T be an operator on $\mathcal{A}$ and $\xi, \xi' \geq 1$. Its (ξ, ξ')-norm is

$$\|T\|_{\xi,\xi'} = \sup_{P \in \mathcal{A}} \frac{\|TP\|_{\xi'}}{\|P\|_\xi}.$$

When $\xi = \xi'$, we write $\|T\|_\xi = \|T\|_{\xi,\xi}$.

Similarly, let $\tau \colon \mathcal{A} \to \mathbb{C}$ be a linear form. Its ξ-norm is

$$\|\tau\|_\xi = \sup_{P \in \mathcal{A}} \frac{|\tau(P)|}{\|P\|_\xi}.$$

A particularly important sort of linear forms are tracial states.

Definition 6.4. Let $\mathcal{C}$ be a unital *ast*-algebra. A *tracial state* on $\mathcal{C}$ is a linear form $\tau : \mathcal{C} \to \mathbb{C}$ such that for any $P, Q \in \mathcal{C}$, we have

- $\tau(\mathbf{1}) = 1$, where $\mathbf{1}$ is the empty word;
- $\tau(PQ) = \tau(QP)$;
- $\tau(PP^*) \geq 0$.

Remark 6.5. The normalized trace tr is a tracial state on $\mathcal{B}$. Under Hypothesis 1.3, the Cauchy–Schwarz inequality implies that $\| \operatorname{tr} \|_1 \leq 1$. Furthermore, we have $\operatorname{tr}(\mathbf{1}) = \operatorname{tr}(\mathrm{Id}) = 1$, where Id is the identity matrix, and thus $\| \operatorname{tr} \|_1 = 1$. Assuming furthermore Hypothesis 1.1, we have that $\widetilde{\mathcal{W}}^N_{V,1}$ is a tracial state on $\mathcal{A}_n$, with $\| \widetilde{\mathcal{W}}^N_{V,1} \|_1 = 1$.

6.2. The Dyson–Schwinger equations for the unitary matrices

Let σ be a tracial state on $\mathcal{B}$. A tracial state μ on $\mathcal{A}$ is a solution to the Dyson–Schwinger problem with initial value σ if for all $P \in \mathcal{A}$,

$$\begin{cases} \mu \otimes \mu(\partial_i P) + \mu(\mathcal{D}_i V \cdot P) = 0 & \text{for } 1 \leq i \leq n, \\ \mu|_{\mathcal{B}} = \sigma, \end{cases} \tag{6.1}$$

where ∂_i and $\mathcal{D}_i$ are the non-commutative derivative and cyclic derivative with respect to u_i, see Definitions 4.2 and 4.5.

It has been shown in [8] that there exists a solution to this problem when $\operatorname{Tr} V = \operatorname{Tr} V^*$ (which implies that $\operatorname{Tr} V$ is real), and that the solution is unique for a potential V small enough (i.e., $\sum_{i=1}^k |z_i| < \varepsilon$ for some $\varepsilon > 0$). Notice that for all $N \geq 1$, $\mathcal{M}^{(0),N}_{V,1}$ is a solution to (6.1) with $\sigma = \operatorname{tr}_N$. In [23], a family of equations that generalize (6.1) was studied. The renormalized cumulants $\widetilde{\mathcal{W}}^N_{V,l}$ are solution to these equations. We reproduce them here.

Proposition 6.6 ([23, Proposition 20]). *Assume Hypothesis 1.1. The renormalized cumulants $\{\widetilde{\mathcal{W}}^N_{V,l}\}_{l\geq 1}$ satisfy the equation*

$$\begin{aligned} &\sum_{I \subset [l-1]} \widetilde{\mathcal{W}}^N_{V,|I|+1} \otimes \widetilde{\mathcal{W}}^N_{V,|I^c|+1}(\boldsymbol{P}_I \otimes \boldsymbol{P}_{I^c} \# \partial_i P_l) + \widetilde{\mathcal{W}}^N_{V,l}(\boldsymbol{P}_{[l-1]} \otimes (\mathcal{D}_i V \cdot P_l)) \\ &\quad = -\sum_{j=1}^{l-1} \widetilde{\mathcal{W}}^N_{V,l-1}(P_1 \otimes \cdots \otimes \check{P}_j \otimes \cdots \otimes P_{l-1} \otimes (\mathcal{D}_i P_j \cdot P_l)) \\ &\qquad - \frac{1}{N^2} \widetilde{\mathcal{W}}^N_{V,l+1}(\boldsymbol{P}_{[l-1]} \otimes \partial_i P_l), \end{aligned} \tag{6.2}$$

where $\check{P}_j$ means that the factor P_j is omitted.

The series of maps $\mathcal{M}^{(g),N}_{V,l}$ satisfy similar equations (see (4.7)).

6.3. Radius of convergence of the series $\mathcal{M}_{V,l}^{(g),N}$

Before giving the proof of Theorem 1.4, we show that all the terms $\mathcal{M}_{V,l}^{(g),N}$ have a radius of convergence greater than some $R_V > 0$. We can apply the gradient trick from [23] that we explain in Appendix B to the equations from Proposition 5.10. To do so, we introduce some notation, motivated in Appendix B.

Definition 6.7. Let $P \in \widehat{\mathcal{X}}_n$ be a monomial, we define

$$\begin{aligned}\Delta_i P &= \sum_{P=P_1u_iP_2}\Big(\sum_{P_2P_1=Q_1u_iQ_2} Q_1u_i \otimes Q_2u_i - \sum_{P_2P_1=Q_1u_i^{-1}Q_2} Q_1 \otimes Q_2\Big)\\ &\quad - \sum_{P=P_1u_i^{-1}P_2}\Big(\sum_{P_2P_1=Q_1u_iQ_2} Q_1 \otimes Q_2 - \sum_{P_2P_1=Q_1u_i^{-1}Q_2} u_i^{-1}Q_1 \otimes u_i^{-1}Q_2\Big).\end{aligned}$$

The *reduced Laplacian* $\Delta\colon \mathcal{A}_n \to \mathcal{A}_n^{\otimes 2}$ is then defined by

$$\Delta = \sum_{i=1}^{n} \Delta_i.$$

Let $(P,Q) \in \mathcal{A}_n^2$, we define the operator $\mathcal{P}^Q$ by

$$\mathcal{P}^Q P = \sum_{i=1}^{n} (\mathcal{D}_i Q)(\mathcal{D}_i P).$$

We define the operator D_i, which acts on a monomial P by $D_i P = \deg_i(P)P$, and D by

$$D = \sum_{i=1}^{n} D_i.$$

Furthermore, for an operator T, we introduce its regularization $\overline{T} = TD^{-1}$.

Definition 6.8 ([23, Definition 13]). Let Π be the orthogonal projection of the polynomials onto $\mathcal{B}^{\perp}$, the algebra of polynomials without a degree 0 term. Let $\Pi' = \mathrm{Id} - \Pi$ be the complementary projection of the polynomials onto $\mathcal{B}$.

Let τ be a tracial state. We define

$$T_\tau = (\mathrm{Id} \otimes \tau + \tau \otimes \mathrm{Id})\Delta.$$

The *master operator* is

$$\Xi_\tau^V = \mathrm{Id} + \Pi\overline{T}_\tau + \overline{\mathcal{P}}^V,$$

where $\overline{T}_\tau$ and $\overline{\mathcal{P}}^V$ are the regularization (see Definition 6.7) of T_τ and $\mathcal{P}^V$.

Applying the gradient trick, we obtain, for $(g,l) \neq (0,1)$,

$$
\begin{aligned}
&\mathcal{M}_{V,l}^{(g),N}\Big(\boldsymbol{P}_{[l-1]} \otimes \Xi^{V}_{\mathcal{M}_{V,1}^{(0),N}} P_l\Big) \\
&= -\mathcal{M}_{V,l+1}^{(g-1),N}(\boldsymbol{P}_{[l-1]} \otimes \bar{\Delta} P_l) \\
&\quad - \sum_{h=1}^{g-1} \mathcal{M}_{V,l}^{(g-h),N}\big(\boldsymbol{P}_{[l-1]} \otimes \mathcal{M}_{V,1}^{(h),N} \otimes \mathrm{Id}\big)(\bar{\Delta} P_l) \\
&\quad - \sum_{\substack{\emptyset \subsetneq I \subsetneq [l-1] \\ g_1+g_2=g}} \mathcal{M}_{V,|I|+1}^{(g_1),N} \otimes \mathcal{M}_{V,|I^c|+1}^{(g_2),N}(\boldsymbol{P}_I \otimes \boldsymbol{P}_{I^c} \# \bar{\Delta} P_l) \\
&\quad - \sum_{j=1}^{l-1} \mathcal{M}_{V,l-1}^{(g),N}(P_1, \ldots, \check{P}_j, \ldots, P_{l-1}, \bar{\mathcal{P}}^{P_j} P_l).
\end{aligned} \tag{6.3}
$$

Proposition 6.9. *Fix a potential $V = \sum_{i=1}^{k} z_i q_i$, with $q_1, \ldots, q_k \in \mathcal{X}_n$. Let $g \geq 0$, $l \geq 1$, $\boldsymbol{P} \in \mathcal{A}_n^l$.*

The radius of convergence of $\mathcal{M}_{V,l}^{(g),N}(\boldsymbol{P})$ depends only on k and $q_1, \ldots, q_k$, and is greater than

$$
R_V = \min\left(\frac{1}{2}(4A_k D_{k,\nu})^{-1}, \frac{1}{2k\nu(4A_k + \frac{2^{k+2}}{B_k})^{\nu}}\right),
$$

where $\nu = \max_{1 \leq i \leq k} \deg q_i$ and the constants A_k, B_k, and $D_{k,\nu}$ are those of Proposition 4.22.

Proof. Let $\boldsymbol{P} \in (\mathcal{X}_n)^l$ be monomials. As $\mathcal{M}_{V,l}^{(g),N}$ is linear in each polynomial P_i, the result follows from the case where the P_i are monomials. Using Proposition 4.22, the series $\mathcal{M}_{V,1}^{(0),N}(\boldsymbol{P})$ can be bounded as follows:

$$
\begin{aligned}
\left|\mathcal{M}_{V,1}^{(0),N}(\boldsymbol{P})\right| &\leq \sum_{\boldsymbol{n} \in \mathbb{N}^k} \frac{\boldsymbol{z}^{\boldsymbol{n}}}{\boldsymbol{n}!} \Big|\mathcal{M}^{(0),N}_{0,\sum_i n_i + l}(\underbrace{q_1, \ldots, q_1}_{n_1 \text{ times}}, \ldots, \underbrace{q_k, \ldots, q_k}_{n_k \text{ times}}, P_1, \ldots, P_l)\Big| \\
&\leq \frac{(4A_k)^{\deg \boldsymbol{P}}}{B_k} \sum_{\boldsymbol{n} \in \mathbb{N}^k} \boldsymbol{z}^{\boldsymbol{n}} (4A_k D_{k,\nu})^{\boldsymbol{n}} \\
&\leq \frac{(4A_k)^{\deg \boldsymbol{P}}}{B_k} \prod_{i=1}^{k} \frac{1}{1 - 4A_k D_{k,\nu} z_i},
\end{aligned}
$$

where Proposition 4.22 is used in the second line. Notice that the radius of convergence of the series does not depend on $\boldsymbol{P}$.

Assuming that $\|\boldsymbol{z}\|_\infty < \frac{1}{2}(4A_k D_{k,\nu})^{-1}$, we get

$$
\|\mathcal{M}_{V,1}^{(0),N}\|_{4A_k} \leq \frac{2^k}{B_k}.
$$

Define

$$K(\xi, V) = \frac{2^{k+1}}{B_k} \frac{4A_k}{\xi - 4A_k} + \|\Pi V\|_1 \nu \xi^{\nu},$$

where as before $\nu = \max_{1 \le i \le k} \deg q_i$. Choose $\xi = 4A_k + \frac{2^{k+2}}{B_k}$. Then, assuming that

$$\|\Pi V\|_1 = \sum_{i=1}^{k} |z_i| < \frac{1}{2\nu\xi^{\nu}},$$

where Π is the projection operator introduced in Definition 6.8, we have $K(\xi, V) < 1$ and $\Xi^{V}_{\mathcal{M}^{(0),N}_{V,1}}$ is an invertible operator $\mathcal{B}_{\xi}^{\perp} \to \mathcal{B}_{\xi}^{\perp}$. Note that this is satisfied if $\|z\|_{\infty} < \frac{1}{2k\nu\xi^{\nu}}$. We thus set $R_V = \min(\frac{1}{2}(4A_k D_{k,\nu})^{-1}, \frac{1}{2k\nu\xi^{\nu}})$.

We then proceed by induction. Assume that for all $(g', l') < (g, l)$ (with the lexicographic order), and for all $\boldsymbol{P} \in \mathcal{X}_n^{l'}$, the series $\mathcal{M}^{(g'),N}_{V,l'}(\boldsymbol{P})$ has a radius of convergence greater than R_V. Then, the right side of (6.3) is a holomorphic function that is defined on a polydisk of radius R_V. The left side is a holomorphic function defined on a polydisk of radius $R_{l,g,V}$ which coincides with the right side. Thus, it can be extended to a holomorphic function on a polydisk of radius R_V. The fact that $\Xi^{V}_{\mathcal{M}^{(0),N}_{V,1}}$ is invertible allows us to conclude. ∎

6.4. The topological expansion: proof of Theorem 1.4

We introduce the truncated formal cumulant (cf., Definition 3.38)

$$S^{(g),N}_{V,l} = \sum_{h=0}^{g} \frac{1}{N^{2h}} \mathcal{M}^{(h),N}_{V,l}.$$

We will show that the cumulants $\widetilde{\mathcal{W}}^N_{V,l}$ admit a topological expansion by bounding the errors $\delta^{(g),N}_{V,l}$ defined by

$$\delta^{(g),N}_{V,l} = \widetilde{\mathcal{W}}^N_{V,l} - S^{(g),N}_{V,l}. \tag{6.4}$$

In particular, we will set $\delta^{(-1),N}_{V,l} = \widetilde{\mathcal{W}}^N_{V,l}$.

We will derive equations on the errors $\delta^{(g),N}_{V,l}$. To make this clearer, we first consider the case $g = 0, l = 1$. In that case, when $\|z\|_{\infty} < R_V$, we have as a consequence of (4.7),

$$\frac{1}{2}\Big(S^{(0),N}_{V,1} \otimes \mathcal{M}^{(0),N}_{V,1} + \mathcal{M}^{(0),N}_{V,1} \otimes S^{(0),N}_{V,1}\Big)(\partial_i P) + S^{(0),N}_{1}((\mathcal{D}_i V)P) = 0. \tag{6.5}$$

On the other hand, Proposition 6.6 implies

$$\widetilde{\mathcal{W}}^N_{V,1} \otimes \widetilde{\mathcal{W}}^N_{V,1}(\partial_i P) + \widetilde{\mathcal{W}}^N_{V,1}(\mathcal{D}_i V \cdot P) = -\frac{1}{N^2} \widetilde{\mathcal{W}}^N_{V,2}(\partial_i P). \tag{6.6}$$

Taking the difference of (6.5) and (6.6), we get

$$\frac{1}{2}\Big(\delta^{(0),N}_{V,1} \otimes \widetilde{W}^N_{V,1} + \widetilde{W}^N_{V,1} \otimes \delta^{(0),N}_{V,1} + S^{(0),N}_{V,1} \otimes \delta^{(0),N}_{V,1} + \delta^{(0),N}_{V,1} \otimes S^{(0),N}_{V,1}\Big)(\partial_i P)$$
$$+ \delta^{(0),N}_{V,1}((\mathcal{D}_i V)P) = -\frac{1}{N^2}\widetilde{W}^N_{V,2}(\partial_i P).$$

We apply the gradient trick as in Appendix B. To do so, we replace P by $\mathcal{D}_i P$. Making use of Lemma B.1 and the fact that

$$\frac{1}{2}\Big(\delta^{(0),N}_{V,1} \otimes \widetilde{W}^N_{V,1} + \widetilde{W}^N_{V,1} \otimes \delta^{(0),N}_{V,1}\Big) \quad \text{and} \quad \frac{1}{2}\Big(S^{(0),N}_{V,1} \otimes \delta^{(0),N}_{V,1} + \delta^{(0),N}_{V,1} \otimes S^{(0),N}_{V,1}\Big)$$

are symmetric, we can make the master operator defined in Definition 6.8 appear. We get

$$\delta^{(0),N}_{V,1}\Big(\Xi^V_{\widetilde{W}^N_{V,1}/2+S^{(0),N}_{V,1}/2}\Big)(P) = -\delta^{(0),N}_{V,1} \otimes \delta^{(0),N}_{V,1}(\overline{\Delta} P) - \frac{1}{N^2}\widetilde{W}^N_{V,2}(\partial_i P). \quad (6.7)$$

We now turn to the general case. When $\|z\|_\infty < R_V$, the truncated formal cumulants satisfy the equation

$$\begin{aligned}
&\sum_{\substack{I\subset[l-1]\\ 0\le f\le g}} \frac{1}{2}\Big(\frac{1}{N^{2f}}\mathcal{M}^{(f),N}_{V,|I|+1} \otimes S^{(g-f),N}_{V,|I^c|+1} + \frac{1}{N^{2f}} S^{(g-f),N}_{V,|I|+1} \otimes \mathcal{M}^{(f),N}_{V,|I^c|+1}\Big)\\
&\qquad \times (\boldsymbol{P}_I \otimes \boldsymbol{P}_{I^c}\#\partial_i P_l)\\
&\qquad + S^{(g),N}_l(\boldsymbol{P}_{[l-1]} \otimes (\mathcal{D}_i V)P_l)\\
&\quad = -\frac{1}{N^2} S^{(g-1),N}_{V,l}(\boldsymbol{P}_{[l-1]} \otimes \partial_i P_l)\\
&\qquad -\sum_{j=1}^{l-1} S^{(g),N}_{V,l-1}(P_1 \otimes \cdots P_{j-1} \otimes P_j \otimes P_{l-1} \otimes (\mathcal{D}_i P_j)P_l).
\end{aligned} \quad (6.8)$$

These equations are obtained by summing equations (5.7) for different values of g, multiplied by $1/N^{2g}$.

Together with (6.2), these equations imply equation (6.9) on the errors defined by (6.4). Before stating the equation, we explain how this equation is derived. We subtract from (6.2) equation (6.8). To have all the terms from (6.2) simplified, we must rewrite the most complicated term:

$$\begin{aligned}
&\sum_{\substack{I\subset[l-1]\\ 0\le f\le g}} \frac{1}{2}\Big(\frac{1}{N^{2f}}\mathcal{M}^{(f),N}_{V,|I|+1} \otimes S^{(g-f),N}_{V,|I^c|+1} + \frac{1}{N^{2f}} S^{(g-f),N}_{V,|I|+1} \otimes \mathcal{M}^{(f),N}_{V,|I^c|+1}\Big)\\
&\qquad \times (\boldsymbol{P}_I \otimes \boldsymbol{P}_{I^c}\#\partial_i P_l).
\end{aligned}$$

We rewrite

$$S_{V,l}^{(g),N} = \widetilde{W}_{V,l}^{N} - \delta_{V,l}^{(g),N}$$

in the sum

$$\begin{aligned}
&\sum_{0\le f\le g} \frac{1}{N^{2f}} \mathcal{M}_{V,|I|+1}^{(f),N} \otimes S_{V,|I^c|+1}^{(g-f),N} \\
&\quad = \sum_{0\le f\le g} \frac{1}{N^{2f}} \mathcal{M}_{V,|I|+1}^{(f),N} \otimes \left(\widetilde{W}_{V,|I^c|+1}^{N} - \delta_{V,|I^c|+1}^{(g-f),N}\right) \\
&\quad = S_{V,|I|+1}^{(g),N} \otimes \widetilde{W}_{V,|I^c|+1}^{N} - \sum_{0\le f\le g} \frac{1}{N^{2f}} \mathcal{M}_{V,|I|+1}^{(f),N} \otimes \delta_{V,|I^c|+1}^{(g-f),N} \\
&\quad = \widetilde{W}_{V,|I|+1}^{N} \otimes \widetilde{W}_{V,|I^c|+1}^{N} - \delta_{V,|I|+1}^{(g),N} \otimes \widetilde{W}_{V,|I^c|+1}^{N} \\
&\qquad - \sum_{0\le f\le g} \frac{1}{N^{2f}} \mathcal{M}_{V,|I|+1}^{(f),N} \otimes \delta_{V,|I^c|+1}^{(g-f),N}.
\end{aligned}$$

We proceed similarly for

$$\sum_{0\le f\le g} \frac{1}{N^{2f}} S_{V,|I^c|+1}^{(g-f),N} \otimes \mathcal{M}_{V,|I|+1}^{(f),N}.$$

After this rewriting and subtracting (6.2), we have

$$\begin{aligned}
&\sum_{I\subset[l-1]} \frac{1}{2}\left(\delta_{V,|I|+1}^{(g),N} \otimes \widetilde{W}_{V,|I^c|+1}^{N} + \widetilde{W}_{V,|I|+1}^{N} \otimes \delta_{V,|I^c|+1}^{(g),N}\right)(\boldsymbol{P}_I \otimes \boldsymbol{P}_{I^c}\#\partial_i P_l) \\
&\quad + \sum_{\substack{I\subset[l-1]\\0\le f\le g}} \frac{1}{2}\left(\frac{1}{N^{2f}} \mathcal{M}_{V,|I|+1}^{(f),N} \otimes \delta_{V,|I^c|+1}^{(g-f),N}\right)(\boldsymbol{P}_I \otimes \boldsymbol{P}_{I^c}\#\partial_i P_l) \\
&\quad + \sum_{\substack{I\subset[l-1]\\0\le f\le g}} \frac{1}{2}\left(\frac{1}{N^{2f}} \delta_{V,|I|+1}^{(g-f),N} \otimes \mathcal{M}_{V,|I^c|+1}^{(f),N}\right)(\boldsymbol{P}_I \otimes \boldsymbol{P}_{I^c}\#\partial_i P_l) \\
&\quad + \delta_{V,l}^{(g),N}(\boldsymbol{P}_{[l-1]} \otimes (\mathcal{D}_i V) P_l) \\
&= -\frac{1}{N^2}\delta_{V,l+1}^{(g-1),N}(\boldsymbol{P}_{[l-1]} \otimes \partial_i P_l) \\
&\quad - \sum_{j=1}^{l-1} \delta_{V,l-1}^{(g),N}(P_1 \otimes \cdots P_{j-1} \otimes P_j \otimes P_{l-1} \otimes (\mathcal{D}_i P_j) P_l).
\end{aligned} \tag{6.9}$$

Using the gradient trick (see Section B), these equations can be rewritten as follows:

$$
\begin{aligned}
&\delta_{V,l}^{(g),N}\Big(\boldsymbol{P}_{[l-1]} \otimes \Xi^{V}_{\widetilde{\mathcal{W}}^N_{V,1}/2+\mathcal{M}^{(0),N}_{V,1}/2}\Big)(P_l)\\
&\quad= -\frac{1}{N^2}\delta_{l+1}^{(g-1)}(\boldsymbol{P}_{[l-1]} \otimes \bar{\Delta} P_l)\\
&\qquad - \frac{1}{2}\Big(\big[\widetilde{\mathcal{W}}^N_{V,l} + \mathcal{M}^{(0),N}_{V,l}\big](\boldsymbol{P}_{[l-1]} \otimes \mathrm{Id}) \otimes \delta_{V,1}^{(g),N}\Big)(\bar{\Delta} P_l)\\
&\qquad - \frac{1}{2}\Big(\delta_{V,1}^{(g),N} \otimes \big[\widetilde{\mathcal{W}}^N_{V,l} + \mathcal{M}^{(0),N}_{V,l}\big](\boldsymbol{P}_{[l-1]} \otimes \mathrm{Id})\Big)(\bar{\Delta} P_l)\\
&\qquad - \sum_{\emptyset \subsetneq I \subsetneq [l-1]} \frac{1}{2}\Big(\big[\widetilde{\mathcal{W}}^N_{V,l} + \mathcal{M}^{(0),N}_{V,l}\big] \otimes \delta_{V,|I^c|+1}^{(g),N}\Big)(\boldsymbol{P}_I \otimes \boldsymbol{P}_{I^c} \# \bar{\Delta} P_l)\\
&\qquad - \sum_{\emptyset \subsetneq I \subsetneq [l-1]} \frac{1}{2}\Big(\delta_{V,|I|+1}^{(g),N} \otimes \big[\widetilde{\mathcal{W}}^N_{V,l} + \mathcal{M}^{(0),N}_{V,l}\big]\Big)(\boldsymbol{P}_I \otimes \boldsymbol{P}_{I^c} \# \bar{\Delta} P_l)\\
&\qquad - \sum_{\substack{I \subset [l-1]\\ 1 \le f \le g}} \frac{1}{2}\left(\frac{1}{N^{2f}}\mathcal{M}^{(f),N}_{V,|I|+1} \otimes \delta_{V,|I^c|+1}^{(g-f),N}\right)(\boldsymbol{P}_I \otimes \boldsymbol{P}_{I^c} \# P_l)\\
&\qquad - \sum_{\substack{I \subset [l-1]\\ 1 \le f \le g}} \frac{1}{2}\left(\frac{1}{N^{2f}}\delta_{V,|I|+1}^{(g-f),N} \otimes \mathcal{M}^{(f),N}_{V,|I^c|+1}\right)(\boldsymbol{P}_I \otimes \boldsymbol{P}_{I^c} \# P_l)\\
&\qquad - \sum_{j=1}^{l-1} \delta_{l-1}^{(g)}(P_1 \otimes \cdots \otimes \check{P}_j \otimes \cdots \otimes P_{l-1} \otimes \bar{\mathcal{P}}^{P_j} P_l).
\end{aligned}
\tag{6.10}
$$

The notation Id in the third line means the identity operator, in particular terms on the third line must be read as follows:

$$
\begin{aligned}
&\big[\widetilde{\mathcal{W}}^N_{V,l} + \mathcal{M}^{(0),N}_{V,l}\big](\boldsymbol{P}_{[l-1]} \otimes \mathrm{Id}) \otimes \delta_{V,1}^{(g),N}(P \otimes Q)\\
&\quad = \big[\widetilde{\mathcal{W}}^N_{V,l} + \mathcal{M}^{(0),N}_{V,l}\big](\boldsymbol{P}_{[l-1]} \otimes P) \otimes \delta_{V,1}^{(g),N}(Q).
\end{aligned}
$$

The bounds of Proposition 4.22 imply the following results.

Lemma 6.10. *Assume that for all $N \ge 1$, $\operatorname{Tr} V$ is real and $\|A_i^N\| \le 1$ for all i (Hypotheses 1.1 and 1.3). There exist $\xi > 1$ and $\varepsilon > 0$ such that if*

$$\|z\|_\infty < \varepsilon,$$

then, for all $g \ge 0$ and $l \ge 1$, we have

$$\|\delta_1^{(0)}\|_\xi \le \frac{C}{N^2}.$$

Proof. Consider equation (6.7) for the errors with $g = 0, l = 1$:

$$\delta_1^{(0)}\Big(\Xi^V_{\widetilde{W}^N_{V,1}/2+\mathcal{M}^{(0),N}_{V,1}/2}P\Big) = -\delta_1^{(0)} \otimes \delta_1^{(0)}(\overline{\Delta}P) - \frac{1}{N^2}\widetilde{W}^N_{V,2}(\overline{\Delta}P).$$

First, notice that the series $\mathcal{M}^{(0),N}_{V,1}$ satisfies $\|\mathcal{M}^{(0),N}_{V,1}\|_{4A_k} \leq \frac{2^k}{B_k}$, with A_k and B_k the constants from Proposition 4.22. Proposition B.3 implies that

$$\Big\|\overline{T}_{\mathcal{M}^{(0),N}_{V,1}}\Big\|_\xi \leq \frac{2^{k+1}}{B_k}\frac{4A_k}{\xi + 4A_k}.$$

Let $\xi \geq 32A_k \geq 12$ and $0 < \varepsilon < R_V$ such that

$$K(\xi, V) = 2\frac{\xi+1}{\xi(\xi-1)} + \frac{2^k}{B_k}\frac{4A_k}{\xi+4A_k} + \|\Pi V\|_{1}\nu\xi^\nu < 1/2.$$

In that case, the operator $\Xi^V_{\widetilde{W}^N_{V,1}/2+\mathcal{M}^{(0),N}_{V,1}/2} : \mathcal{B}^\perp_\xi \to \mathcal{B}^\perp_\xi$ is invertible.

We get that

$$\|\delta_1^{(0)}\|_\xi \leq \|\delta_1^{(0)}\|_\xi \big\|\overline{T}_{\delta_1^{(0)}}\big\|_\xi \Big\|\Big(\Xi^V_{\widetilde{W}^N_{V,1}}\Big)^{-1}\Big\|_\xi + \frac{C'}{N^2}\|\overline{\Delta}\|_{\xi,\xi/2}\Big\|\Big(\Xi^V_{\widetilde{W}^N_{V,1}}\Big)^{-1}\Big\|_\xi,$$

where we used that there exists a constant $C' > 0$ such that $\|\widetilde{W}^N_{V,2}\|_{\xi/2} \leq C'$ by [23, Theorem 22]. Note that for this Theorem 22 to be applicable, one must show that the sequence of cumulants is ξ-uniformly bounded in the sense of [23, Definition 21]. This is shown assuming Hypothesis 1.1 in [23, Corollary 32] for all $\xi \geq 12$.

The bound on $\mathcal{M}^{(0),N}_{V,1}$ implies that $\|\delta_1^{(0)}\|_{4A_k} \leq 1 + \frac{2^k}{B_k} \leq 2$. This fact and Proposition B.3 give

$$\|\overline{T}_{\delta_1^{(0)}}\|_\xi \leq 2\Big(1 + \frac{2^k}{B_k}\Big)\frac{4A_k}{\xi - 4A_k} < 1/2.$$

With this result and [23, Proposition 19], we finally get that

$$\|\delta_1^{(0)}\|_\xi \leq \frac{C'}{N^2}\frac{\|(\Xi^V_{\widetilde{W}^N_{V,1}})^{-1}\|_\xi}{1 - \|\overline{T}_{\delta_1^{(0)}}\|_\xi\|(\Xi^V_{\widetilde{W}^N_{V,1}})^{-1}\|_\xi} \leq \frac{C'}{N^2}\frac{1}{1/2 - K(\xi, V)}.$$ ∎

Proposition 6.11. *Assume that for all $N \geq 1$, $\operatorname{Tr} V$ is real and $\|A_i^N\| \leq 1$ for all i (Hypotheses 1.1 and 1.3). There exist $\xi > 1$ and $\varepsilon > 0$ such that if*

$$\|z\|_\infty < \varepsilon,$$

then, for all $g \geq 0$ and $l \geq 1$, we have

$$\|\delta_l^{(g)}\|_{2^{l-2}\xi} = \mathcal{O}(N^{-2g-2}).$$

Proof. We proceed by induction on (g, l), with lexicographic order. For $l = 1, g = 0$, the result is given by Lemma 6.10. Assume now that for all $(g', l') < (g, l)$, we have

$$\|\delta_{l'}^{(g')}\|_\xi = \mathcal{O}(N^{-2g'-2}).$$

Then, in equations (6.10) for the errors, all the terms on the right side of the equation are of order N^{-2g}. Note that terms $\delta_l^{(-1)} = \widetilde{W}_{V,l}^N$ are bounded using [23, Theorem 22]. This gives the result. ■

We can finally prove Theorem 1.4.

Proof of Theorem 1.4. Proposition 6.11 directly implies Theorem 1.4, with Hypothesis 1.2 replaced by Hypothesis 1.3. Then, if we only assume Hypothesis 1.2, we set

$$c = \frac{1}{\sup_{N\geq 1} \sup_{1\leq i\leq p} \|A_i^N\|}.$$

We can then replace each matrix A_i^N by cA_i^N and rescale each coefficient z_i of V by an appropriate multiple of c^{-1}. When the new coefficients of V are small enough, we can apply the result obtained with Hypothesis 1.2 and obtain the result. ■

Remark 6.12. Notice that for N big enough, the series

$$\sum_{h=0}^{g} \frac{1}{N^{2h}} \mathcal{M}_{V,l}^{(h),N}(P_1, \dots, P_l)$$

is well defined for all V with $\boldsymbol{z}$ small enough, even if Hypothesis 1.1 is not satisfied. In fact, for any V with $\boldsymbol{z}$ small, provided the cumulants exist, are bounded, and satisfy the Dyson–Schwinger equations, the same method applies and the asymptotic topological expansion holds.

The complex asymptotics of the HCIZ and BGW partition functions were studied with a different method in [35].

A. Bounds for the sum of maps $\mathcal{M}_{0,l}^{(g),N}$

This appendix gives a detailed proof of Proposition 4.22.

We assume that $\nu \geq 1$ and that up to cyclic permutation of its factors we can write P_l as Pu. If P_l has no term u, a similar argument holds with $P_l = u^*P$. Furthermore, to make notation less cumbersome, we write

$$\mathcal{M}_{\boldsymbol{n},l}^{(g),N}(P_1, \dots, P_l) = \mathcal{M}_{0,\sum_i n_i+l}^{(g),N}(\underbrace{q_1, \dots, q_1}_{n_1 \text{ times}}, \dots, \underbrace{q_k, \dots, q_k}_{n_k \text{ times}}, P_1, \dots, P_l)$$

and omit the indices k and ν in the constants.

To prove the result, we do an induction on $\mathbb{N}^{k+3}$, where we endow a tuple of integers $(g, l, n_1, \ldots, n_k, m)$ with the lexicographic order. Notice that the result is obvious when $n_1 = \cdots = n_k = 0$, $g = 0$, $l = 1$ when $\deg \boldsymbol{P} = 1$ (as $\mathcal{M}_{0,1}^{(0),N}(MU^{\pm 1}) = 0$ for all $M \in \mathcal{Y}$), when $m = 1 = \frac{1}{2}\deg \boldsymbol{P}$ (as $\mathcal{M}_{0,1}^{(0),N}(M_1 U M_2 U^{-1}) = \operatorname{tr} M_1 \operatorname{tr} M_2$), as soon as $M \geq 1$. In fact, by Lemma 3.36, $|\mathcal{M}_{\boldsymbol{0},1}^{(0),N}(P)| \leq 1$ for all $P \in \mathcal{X}$.

Assuming that $m \geq 2$ or $\boldsymbol{n} \neq 0$, Theorem 4.1 yields

$$
\begin{aligned}
&\frac{1}{\boldsymbol{n}!}\mathcal{M}_{\boldsymbol{n},l}^{(g),N}(P_1, \ldots, P_{l-1}, Pu) \\
&\quad = -\frac{1}{\boldsymbol{n}!}\mathcal{M}_{\boldsymbol{n},l+1}^{(g-1),N}(P_1, \ldots, P_{l-1}, (\partial P) \times 1 \otimes u) \\
&\qquad - \sum_{\substack{I \subset [l-1] \\ \boldsymbol{n_1}+\boldsymbol{n_2}=\boldsymbol{n}}} \sum_{g_1+g_2=g} \frac{1}{\boldsymbol{n_1}!}\frac{1}{\boldsymbol{n_2}!}\mathcal{M}_{\boldsymbol{n_1},|I|+1}^{(g_1),N} \\
&\qquad \otimes \mathcal{M}_{\boldsymbol{n_2},|I^c|+1}^{(g_2),N}(\boldsymbol{P}_I \otimes \boldsymbol{P}_{I^c}\#(\partial P) \times 1 \otimes u) \\
&\qquad - \sum_{j=1}^{l-1} \frac{1}{\boldsymbol{n}!}\mathcal{M}_{\boldsymbol{n},l-1}^{(g),N}(P_1, \ldots, \check{P}_j, \ldots, P_{l-1}, (\mathcal{D}P_j)Pu) \\
&\qquad - \sum_{j=1}^{k} \frac{1}{(\boldsymbol{n}-1_j)!}\mathcal{M}_{\boldsymbol{n}-1_j,l}^{(g),N}(P_1, \ldots, P_{l-1}, (\mathcal{D}q_j)Pu),
\end{aligned}
\tag{A.1}
$$

where $\check{P}_j$ means that P_j is removed.

Assuming that the bound (4.6) holds for

$$
(g', l', n'_1, \ldots, n'_k, m') < (g, l, n_1, \ldots, n_k, m),
$$

we get four terms from (A.1). The first term is

$$
\begin{aligned}
&\frac{1}{\boldsymbol{n}!}\big|\mathcal{M}_{\boldsymbol{n},l+1}^{(g-1),N}(P_1, \ldots, P_{l-1}, (\partial P) \times 1 \otimes u)\big| \\
&\quad \leq \sum_{m'=1}^{\deg Pu} A^{(l+1)(2m+\nu\boldsymbol{n})} B^{-l-1} C^{(g-1)(2m+\nu\boldsymbol{n})} D^{\boldsymbol{n}} c_{m'} c_{\deg Pu - m'} \prod_{i=1}^{l-1} c_{\deg P_i} \prod_{j=1}^{k} c_{n_j} \\
&\quad \leq A^{(l+1)(2m+\nu\boldsymbol{n})} B^{-l-1} C^{(g-1)(2m+\nu\boldsymbol{n})} D^{\boldsymbol{n}} (c_{\deg Pu+1} - c_{\deg Pu}) \prod_{i=1}^{l-1} c_{\deg P_i} \prod_{j=1}^{k} c_{n_j} \\
&\quad \leq \frac{3}{B}\left(\frac{A}{C}\right)^{2m} A^{l(2m+\nu\boldsymbol{n})} B^{-l} C^{g(2m+\nu\boldsymbol{n})} D^{\boldsymbol{n}} \prod_{i=1}^{l} c_{\deg P_i} \prod_{j=1}^{k} c_{n_j}.
\end{aligned}
$$

In the second line, we expanded the non-commutative derivative (see Definition 4.2). In the third line we used the recurrence formula for Catalan $c_{n+1} = \sum_{i=0}^{n} c_i c_{n-i}$ and in the fourth line we used that $c_{n+1} \leq 4c_n$ for all $n \in \mathbb{N}$. We choose A and C so that $A/C \leq 1$ and $B \geq 12$.

The second term is

$$\begin{aligned}
&\sum_{\substack{I\subset[l-1]\\ \boldsymbol{n_1}+\boldsymbol{n_2}=\boldsymbol{n}}} \sum_{g_1+g_2=g} \frac{1}{\boldsymbol{n_1}!}\frac{1}{\boldsymbol{n_2}!}\left|\mathcal{M}^{(g_1),N}_{\boldsymbol{n_1},|I|+1} \otimes \mathcal{M}^{(g_2),N}_{\boldsymbol{n_2},|I^c|+1}\right|(\boldsymbol{P}_I \otimes \boldsymbol{P}_{I^c}\#(\partial P) \times 1 \otimes u) \\
&\quad \leq \sum_{m'=1}^{\deg P} \sum_{I\subset[l-1]} \sum_{g_1+g_2=g} A^{m_1(|I|+1)+m_2(|I^c|+1)} B^{-l-1} C^{g_1m_1+g_2m_2} c_{m'} c_{\deg Pu-m'} \\
&\quad \times \sum_{\boldsymbol{n_1}+\boldsymbol{n_2}=\boldsymbol{n}} A^{l\nu\boldsymbol{n}} C^{g\nu\boldsymbol{n}} D^{\boldsymbol{n}} \prod_{i=1}^{l-1} c_{\deg P_i} \prod_{i=1}^{k} c_{n_{1,i}} c_{n_{2,i}},
\end{aligned}$$

where we used the notation $m_1 = \sum_{i\in I} \deg P_i + m'$ and $m_2 = \sum_{i\in I^c} \deg P_i + \deg Pu - m'$. With this notation, we get as soon as $C \geq 2$,

$$\begin{aligned}
\sum_{g_1+g_2=g} C^{g_1m_1+g_2m_2} &= C^{2mg} \sum_{h=0}^{g} \left(\frac{1}{C^{m_2}}\right)^h \left(\frac{1}{C^{m_1}}\right)^{g-h} \\
&\leq C^{2mg} \sum_{h=0}^{g} 2^{-g} \\
&\leq C^{2mg}.
\end{aligned}$$

In the second line, we used that $m_1, m_2 \geq 1$. Similarly, we have when $A \geq 2$,

$$\begin{aligned}
\sum_{I\subset[l-1]} A^{m_1(|I|+1)+m_2(|I^c|+1)} &= A^{2ml} \sum_{I\subset[l-1]} A^{-m_1|I^c|-m_2|I|} \\
&\leq A^{2ml} \sum_{i=0}^{l-1} \binom{l-1}{i} \left(\frac{1}{A^{\deg Pu-m'}}\right)^i \left(\frac{1}{A^{m'}}\right)^{l-i-i} \\
&= A^{2ml} \left(\frac{1}{A^{\deg Pu-m'}} + \frac{1}{A^{m'}}\right)^{l-1} \\
&\leq A^{2ml}.
\end{aligned}$$

We finally get

$$\begin{aligned}
&\sum_{\substack{I\subset[l-1]\\ \boldsymbol{n_1}+\boldsymbol{n_2}=\boldsymbol{n}}} \sum_{g_1+g_2=g} \frac{1}{\boldsymbol{n_1}!}\frac{1}{\boldsymbol{n_2}!}\left|\mathcal{M}^{(g_1),N}_{\boldsymbol{n_1},|I|+1} \otimes \mathcal{M}^{(g_2),N}_{\boldsymbol{n_2},|I^c|+1}\right|(\boldsymbol{P}_I \otimes \boldsymbol{P}_{I^c}\#(\partial P) \times 1 \otimes u) \\
&\quad \leq \frac{6 \cdot 4^k}{B} A^{l(2m+\nu\boldsymbol{n})} B^{-l} C^{g(2m+\nu\boldsymbol{n})} D^{\boldsymbol{n}} \prod_{i=1}^{l} c_{\deg P_i} \prod_{i=1}^{k} c_{n_i}.
\end{aligned}$$

Thus, we choose $B \geq 6 \cdot 4^{k+1}$.

The third term is

$$\begin{aligned}
&\sum_{j=1}^{l-1} \frac{1}{\boldsymbol{n}!} \left| \mathcal{M}_{\boldsymbol{n},l-1}^{(g),N}(P_1, \ldots, \check{P}_j, \ldots, P_{l-1}, (\mathcal{D} P_j) P u) \right| \\
&\leq \sum_{j=1}^{l-1} (\deg P_j) A^{(l-1)(2m+\nu \boldsymbol{n})} B^{-l+1} C^{g(2m+\nu \boldsymbol{n})} D^{\boldsymbol{n}} c_{\deg P_j + \deg Pu} \prod_{\substack{i=1 \\ i \neq j}}^{l-1} c_{\deg P_i} \prod_{j=1}^{k} c_{n_j} \\
&\leq \frac{B}{A^{2m+\nu \boldsymbol{n}}} \left(\sum_{j=1}^{l-1} (\deg P_j) \frac{c_{\deg P_j + \deg Pu}}{c_{\deg P_j} c_{\deg Pu}} \right) \\
&\quad \times A^{l(2m+\nu \boldsymbol{n})} B^{-l} C^{g(2m+\nu \boldsymbol{n})} D^{\boldsymbol{n}} \prod_{i=1}^{l} c_{\deg P_i} \prod_{j=1}^{k} c_{n_j}.
\end{aligned}$$

To bound this term, we use the following estimate for the Catalan numbers, a consequence of the Stirling bound:

$$\begin{aligned}
\frac{4^n}{(n+1)\sqrt{\pi n}} \exp\left(\frac{1}{24n+1} - \frac{1}{24n} \right) &\leq c_n \\
&\leq \frac{4^n}{(n+1)\sqrt{\pi n}} \exp\left(\frac{1}{24n} - \frac{1}{24n+2} \right),
\end{aligned}$$

which implies

$$\frac{4^n}{\sqrt{\pi}(n+1)^{3/2}} \leq c_n \leq \frac{4^n}{\sqrt{\pi} n^{3/2}}.$$

It implies that for $p, q \in \mathbb{N}^*$,

$$\frac{c_{p+q}}{c_p c_q} \leq \pi^{1/2} \left(\frac{(p+1)(q+1)}{p+q} \right)^{3/2} \leq \pi^{1/2} (p+1)^{3/2}.$$

Thus,

$$\begin{aligned}
&\frac{B}{A^{2m+\nu \boldsymbol{n}}} \left(\sum_{j=1}^{l-1} (\deg P_j) \frac{c_{\deg P_j + \deg Pu}}{c_{\deg P_j} c_{\deg Pu}} \right) \\
&\qquad \leq \frac{\pi^{1/2} B}{A^{2m+\nu \boldsymbol{n}}} (\deg Pu + 1)^{3/2} (2m - \deg Pu).
\end{aligned}$$

As we can assume that $m \geq 1$ (else this term could be bounded by 0), it suffices to choose $A \leq 2 B^{1/2} \pi^{1/4} 2^{3/2}$. Notice that for all $n \geq 1$, $(n+1)^{3/2} \leq 2^{3n/2}$.

The fourth term is

$$\begin{aligned}
&\sum_{j=1}^{k} \frac{1}{(\boldsymbol{n}-1_j)!} \mathcal{M}^{(g),N}_{\boldsymbol{n}-1_j,l}(P_1,\dots,P_{l-1},(\mathcal{D}q_j)Pu) \\
&\leq \frac{1}{D} \sum_{j=1}^{k} (\deg q_k) A^{l(2m+\nu\boldsymbol{n})} B^{-l} C^{g(2m+\nu\boldsymbol{n})} D^{\boldsymbol{n}} \frac{c_{\deg Pu+\deg q_j} c_{n_j-1}}{c_{\deg Pu} c_{n_j}} \prod_{i=1}^{l} c_{\deg P_i} \prod_{i=1}^{k} c_{n_i} \\
&\leq \frac{1}{D} \sum_{j=1}^{k} 4^{\deg q_k} (\deg q_k) A^{l(2m+\nu\boldsymbol{n})} B^{-l} C^{g(2m+\nu\boldsymbol{n})} D^{\boldsymbol{n}} \prod_{i=1}^{l} c_{\deg P_i} \prod_{i=1}^{k} c_{n_i} \\
&\leq \frac{1}{D} \sum_{j=1}^{k} (4\mathrm{e}^{1/\mathrm{e}})^{\deg q_k} A^{l(2m+\nu\boldsymbol{n})} B^{-l} C^{g(2m+\nu\boldsymbol{n})} D^{\boldsymbol{n}} \prod_{i=1}^{l} c_{\deg P_i} \prod_{i=1}^{k} c_{n_i}.
\end{aligned}$$

We choose $D = 4k(4\mathrm{e}^{1/\mathrm{e}})^{\nu}$ to get the result. Notice that we can thus choose

$$\begin{aligned}
A = C &= 2^{k+3}\sqrt{6}\pi^{1/4}, \\
B &= 3 \cdot 4^{k+1}, \\
D &= 4k(4\mathrm{e}^{1/\mathrm{e}})^{\nu}.
\end{aligned}$$

B. The gradient trick

We use several times the gradient trick, previously introduced in [23]. The main idea of the gradient trick is to replace the polynomial P (or P_l) the equations of Proposition 6.6 (or in the Dyson–Schwinger problem (6.1), see Section 6) by its cyclic derivative $\mathcal{D}_i P$. An operator—the master operator introduced below—naturally appears in the equations. When the potential V is small enough, this operator is invertible. The gradient trick was introduced in [23] to study the Dyson–Schwinger lattice of equations.

B.1. The trick

The gradient trick allows us to simplify quadratic terms. We take as an example the equation for the sums of maps for $g = 0, l = 2$

$$\sum_{I \subset [l-1]} \mathcal{M}^{(0),N}_{0,|I|+1} \otimes \mathcal{M}^{(0),N}_{0,|I^c|+1}(\boldsymbol{P}_I \otimes \boldsymbol{P}_{I^c} \# \partial_i P_2) = -\mathcal{M}^{(g),N}_{0,1}((\mathcal{D}_i P_1) P_2).$$

We can rewrite it as

$$\mathcal{M}^{(0),N}_{0,2}\big(P_1 \otimes \mathrm{Id} \otimes \mathcal{M}^{(0),N}_{0,1} + P_1 \otimes \mathcal{M}^{(0),N}_{0,1} \otimes \mathrm{Id}\big)(\partial_i P_2) = -\mathcal{M}^{(g),N}_{0,1}((\mathcal{D}_i P_1) P_2),$$

where Id denote the identity operator. In particular, the notation

$$P_1 \otimes \mathrm{Id} \otimes \mathcal{M}_{0,1}^{(0),N}(\partial_i P_2)$$

must be understood as follows. For $(Q_1, Q_2) \in \mathcal{A}_n^2$,

$$P_1 \otimes \mathrm{Id} \otimes \mathcal{M}_{0,1}^{(0),N}(Q_1 \otimes Q_2) = \big(\mathcal{M}_{0,1}^{(0),N}(Q_2)\big) \cdot (P_1 \otimes Q_1) \in \mathcal{A}_n^{\otimes 2}.$$

We now replace P_2 by its cyclic derivative $\mathcal{D}_i P_2$ and obtain

$$\begin{aligned}&\mathcal{M}_{0,2}^{(0),N}\big(P_1 \otimes \mathrm{Id} \otimes \mathcal{M}_{0,1}^{(0),N} + P_1 \otimes \mathcal{M}_{0,1}^{(0),N} \otimes \mathrm{Id}\big)(\partial_i \mathcal{D}_i P_2)\\ &\quad = -\mathcal{M}_{0,1}^{(g),N}((\mathcal{D}_i P_1)(\mathcal{D}_i P_2)).\end{aligned}$$

Lemma B.1. *Let $\mu_2 \colon \mathcal{A}_n \times \mathcal{A}_n \to \mathbb{C}$ be a bilinear form, tracial in each of its variables. For a monomial $P \in \mathcal{X}_n$, write $\deg_i^+(P)$ for the number of factors u_i in P and $\deg_i^-(P)$ for the number of factors u_i^* in P. We have for any monomial $P \in \mathcal{A}_n$,*

$$\mu_2(\partial_i \mathcal{D}_i P) = \deg_i^+(P)\mu_2(P \otimes 1) + \deg_i^-(P)\mu_2(1 \otimes P) + \mu_2(\Delta_i P),$$

with the operator Δ_i introduced in Definition 6.7. In particular, if μ_2 is symmetric, we get

$$\mu_2(\partial_i \mathcal{D}_i P) = \deg_i(P) \cdot \mu_2(1 \otimes P) + \mu_2(\Delta_i P).$$

This lemma allows us to rewrite the above expression as

$$\begin{aligned}&\mathcal{M}_{0,2}^{(0),N}\big(\deg_i(P_2) P_1 \otimes \mathrm{Id} + (P_1 \otimes \mathrm{Id} \otimes \mathcal{M}_{0,1}^{(0),N} + P_1 \otimes \mathcal{M}_{0,1}^{(0),N} \otimes \mathrm{Id})(\Delta_i P_2)\big)\\ &\quad = -\mathcal{M}_{0,1}^{(g),N}((\mathcal{D}_i P_1)(\mathcal{D}_i P_2)).\end{aligned}$$

Introducing the operator

$$\mathcal{P}_i^q P = (\mathcal{D}_i q)(\mathcal{D}_i P),$$

for $P, Q \in \mathcal{A}_n$, we get

$$\begin{aligned}&\mathcal{M}_{0,2}^{(0),N}\big(\deg_i(P_2) P_1 \otimes \mathrm{Id} + (P_1 \otimes \mathrm{Id} \otimes \mathcal{M}_{0,1}^{(0),N} + P_1 \otimes \mathcal{M}_{0,1}^{(0),N} \otimes \mathrm{Id})(\Delta_i P_2)\big)\\ &\quad = -\mathcal{M}_{0,1}^{(g),N}(\mathcal{P}_i^{P_1} P_2)\end{aligned}$$

for $1 \leq i \leq n$.

Using the operators defined in Definition 6.7. The sum of maps $\mathcal{M}_{0,2}^{(0),N}$ satisfies

$$\begin{aligned}&\mathcal{M}_{0,2}^{(0),N}\big(P_1 \otimes \mathrm{Id} + (P_1 \otimes \mathrm{Id} \otimes \mathcal{M}_{0,1}^{(0),N} + P_1 \otimes \mathcal{M}_{0,1}^{(0),N} \otimes \mathrm{Id})(\bar{\Delta} P_2)\big)\\ &\quad = -\mathcal{M}_{0,1}^{(g),N}(\bar{\mathcal{P}}^{P_1} P_2)\end{aligned}$$

for all $P_1, P_2 \in \mathcal{A}_n$. This computation justifies the introduction of the master operator Ξ_τ^V of Definition 6.8.

Thus, we have

$$\mathcal{M}_{0,2}^{(0),N}\left(P_1 \otimes \Xi^0_{\mathcal{M}_{0,1}^{(0),N}} P_2\right) = -\mathcal{M}_{0,1}^{(g),N}(\bar{\mathcal{P}}^{P_1} P_2)$$

for all $P_1, P_2 \in \mathcal{A}_n$. This will be called the secondary form of equation (5.7). Notice that in this particular case $V = 0$. In the sequel, we will derive secondary equation with a potential.

B.2. Operator norm estimates

We now give some bounds on the norms of the different operators. These bounds and more were derived in [23, Section 3.2]. In particular, it was shown that under some hypotheses the master operator is invertible.

Proposition B.2 ([23, Section 3.3]). *Let $\xi \geq 1$, $V \in \mathcal{A}$ and τ be a tracial state satisfying $\|\tau\| \leq 1$. Introduce*

$$K(\xi, V) = 4\frac{\xi + 1}{\xi(\xi - 1)} + \|V\|_1 \xi^{\deg V} \deg V,$$

and assume that $K(\xi, V) < 1$. Then, the operator Ξ_τ^V extends to an operator $\mathcal{B}_\xi^\perp \to \mathcal{B}_\xi^\perp$ ($\mathcal{B}_\xi^\perp$ is defined in Definition 6.1) which is invertible, with inverse satisfying

$$\left\|\left(\Xi_\tau^V\right)^{-1}\right\|_\xi \leq \frac{1}{1 - K(\xi, V)}.$$

We use a slightly modified version of [23, Proposition 17].

Proposition B.3. *Let $1 \leq \xi_1 < \xi_2$, and τ be a linear form $\mathcal{A}_n \to \mathbb{C}$. We have*

$$\|\bar{T}_\tau\|_{\xi_2} \leq 2\|\tau\|_{\xi_1} \frac{\xi_1}{\xi_2 - \xi_1}.$$

Proof. We proceed as in [23]. Let P be a monomial of degree $d \geq 1$. We have

$$\begin{aligned}
&T_\tau P \\
&= \sum_{i=1}^n \sum_{P = P_1 u_i P_2} \Bigg(\sum_{P_2 P_1 u_i = Q_1 u_i Q_2 u_i} (Q_1 u_i \tau(Q_2 u_i) + \tau(Q_1 u_i) Q_2 u_i) \Bigg) \\
&- \sum_{i=1}^n \sum_{P = P_1 u_i P_2} \Bigg(\sum_{P_2 P_1 u_i = Q_1 u_i^{-1} Q_2 u_i} (Q_1 \tau(Q_2) + \tau(Q_1) Q_2) \Bigg)
\end{aligned}$$

$$
\begin{aligned}
&-\sum_{i=1}^{n}\sum_{P=P_1u_i^{-1}P_2}\Bigg(\sum_{u_i^{-1}P_2P_1=u_i^{-1}Q_1u_iQ_2}(Q_1\tau(Q_2)+\tau(Q_1)Q_2)\Bigg)\\
&+\sum_{i=1}^{n}\sum_{P=P_1u_i^{-1}P_2}\Bigg(\sum_{u_i^{-1}P_2P_1=u_i^{-1}Q_1u_i^{-1}Q_2}(u_i^{-1}Q_1\tau(u_i^{-1}Q_2)\\
&\qquad\qquad + \tau(u_i^{-1}Q_1)u_i^{-1}Q_2)\Bigg).
\end{aligned}
$$

We now take the norm $\|\cdot\|_{\xi_2}$ (recall Definition 6.1). Using the triangle inequality and $|\tau(P)|\leq\|\tau\|_{\xi_1}\xi_1^{\deg P}$, we get (as Q_1, Q_2 are monomials)

$$
\begin{aligned}
&\frac{\|T_\tau P\|_{\xi_2}}{\|\tau\|_{\xi_1}}\\
&\leq \sum_{i=1}^{n}\sum_{P=P_1u_iP_2}\Bigg(\sum_{P_2P_1u_i=Q_1u_iQ_2u_i}\xi_2^{\deg_i Q_1u_i}\xi_1^{\deg_i Q_2u_i}+\xi_1^{\deg_i Q_1u_i}\xi_2^{\deg_i Q_2u_i}\Bigg)\\
&\quad+\sum_{i=1}^{n}\sum_{P=P_1u_iP_2}\Bigg(\sum_{P_2P_1u_i=Q_1u_i^{-1}Q_2u_i}\xi_2^{\deg_i Q_1}\xi_1^{\deg_i Q_2}+\xi_1^{\deg_i Q_1}\xi_2^{\deg_i Q_2}\Bigg)\\
&\quad+\sum_{i=1}^{n}\sum_{P=P_1u_i^{-1}P_2}\Bigg(\sum_{u_i^{-1}P_2P_1=u_i^{-1}Q_1u_iQ_2}\xi_2^{\deg_i Q_1}\xi_1^{\deg_i Q_2}+\xi_1^{\deg_i Q_1}\xi_2^{\deg_i Q_2}\Bigg)\\
&\quad+\sum_{i=1}^{n}\sum_{P=P_1u_i^{-1}P_2}\Bigg(\sum_{u_i^{-1}P_2P_1=u_i^{-1}Q_1u_i^{-1}Q_2}\xi_2^{\deg_i u_i^{-1}Q_1}\xi_1^{\deg_i u_i^{-1}Q_2}\\
&\qquad\qquad+\xi_1^{\deg_i u_i^{-1}Q_1}\xi_2^{\deg_i u_i^{-1}Q_2}\Bigg)\\
&\leq 2\sum_{i=1}^{n}\deg_i P\left(\sum_{k=1}^{\deg_i P-1}\xi_2^k\xi_1^{\deg_i P-k}\right).
\end{aligned}
$$

The main step of the computation is the second inequality. The factor $\deg_i P$ accounts for the first sum, on the decompositions of P as

$$P = P_1u_iP_2 \quad\text{or}\quad P = P_1u_i^*P_2.$$

The sum on k accounts from the decompositions of P_2P_1 as

$$P_2P_1 = Q_1u_iQ_2 \quad\text{or}\quad P_2P_1 = Q_1u_i^*Q_2.$$

Finally, there is a factor 2 as we count those decompositions at most twice. For instance, we have

$$
\begin{aligned}
&\sum_{i=1}^{n}\sum_{P=P_1u_iP_2}\Bigg(\sum_{P_2P_1u_i=Q_1u_iQ_2u_i}\big(\xi_2^{\deg_i Q_1u_i}\xi_1^{\deg_i Q_2u_i}+\xi_1^{\deg_i Q_1u_i}\xi_2^{\deg_i Q_2u_i}\big)\Bigg)\\
&\quad+\sum_{i=1}^{n}\sum_{P=P_1u_iP_2}\Bigg(\sum_{P_2P_1u_i=Q_1u_i^{-1}Q_2u_i}\big(\xi_2^{\deg_i Q_1}\xi_1^{\deg_i Q_2}+\xi_1^{\deg_i Q_1}\xi_2^{\deg_i Q_2}\big)\Bigg)\\
&\le\sum_{i=1}^{n}\sum_{P=P_1u_iP_2}\sum_{P_2P_1=Q_1u_i^{\pm1}Q_2}2\xi_2^{\deg_i Q_1u_i^{\pm1}}\xi_1^{\deg_i Q_2u_i^{\pm1}}\\
&\le\sum_{i=1}^{n}\sum_{P=P_1u_iP_2}\sum_{k=1}^{\deg_i P-1}2\xi_2^{k}\xi_1^{\deg_i P-1}\\
&\le 2\sum_{i=1}^{n}\deg_i^{+}P\sum_{k=1}^{\deg_i P-1}\xi_2^{k}\xi_1^{\deg_i P-1}.
\end{aligned}
$$

In the first inequality, we abused notation and wrote $u_i^{\pm1}$ to mean either of u_i or u_i^*. In the last line, $\deg_i^+ P$ denotes the number of letter u_i in P.

We can then conclude that

$$
\begin{aligned}
\frac{\|T_\tau P\|_{\xi_2}}{\|\tau\|_{\xi_1}}&=2\sum_{i=1}^{n}(\deg_i P)\xi_2^{\deg_i P}\sum_{k=1}^{\deg_i P-1}\left(\frac{\xi_1}{\xi_2}\right)^{\deg_i P-k}\\
&\le 2\sum_{i=1}^{n}(\deg_i P)\xi_2^{\deg_i P}\frac{\xi_1}{\xi_2}\frac{1}{1-\xi_1/\xi_2}\\
&\le 2\sum_{i=1}^{n}(\deg_i P)\xi_2^{\deg_i P}\frac{\xi_1}{\xi_2-\xi_1}\\
&\le 2d\frac{\xi_1}{\xi_2-\xi_1}\|P\|_{\xi_2}.
\end{aligned}
$$

In the last line, d is the total degree of P. ∎

Acknowledgments. I thank my PhD advisors, Alice Guionnet and Grégory Miermont, for countless discussions, tips and advice during this project. I thank two anonymous referee for their many helpful recommendations.

Funding. This project was supported by ERC Project LDRAM: ERC-2019-ADG Project 884584.

References

[1] D. Bessis, C. Itzykson, and J. B. Zuber, Quantum field theory techniques in graphical enumeration. *Adv. in Appl. Math.* **1** (1980), no. 2, 109–157 Zbl 0453.05035 MR 0603127

[2] M. Bona (ed.), *Handbook of enumerative combinatorics*. Chapman and Hall/CRC, New York, 2015 Zbl 1314.05001

[3] J. Bouttier, P. Di Francesco, and E. Guitter, Census of planar maps: from the one-matrix model solution to a combinatorial proof. *Nuclear Phys. B* **645** (2002), no. 3, 477–499 Zbl 0999.05052 MR 1938319

[4] E. Brézin and S. Hikami, *Random matrix theory with an external source*. SpringerBriefs Math. Phys. 19, Springer, Singapore, 2016 Zbl 1455.60002 MR 3822737

[5] E. Brézin, C. Itzykson, G. Parisi, and J. B. Zuber, Planar diagrams. *Comm. Math. Phys.* **59** (1978), no. 1, 35–51 Zbl 0997.81548 MR 0471676

[6] L. Chekhov and B. Eynard, Matrix eigenvalue model: Feynman graph technique for all genera. *J. High Energy Phys.* (2006), no. 12, article no. 026 Zbl 1226.81138 MR 2276715

[7] B. Collins, Moments and cumulants of polynomial random variables on unitary groups, the Itzykson–Zuber integral, and free probability. *Int. Math. Res. Not.* (2003), no. 17, 953–982 Zbl 1049.60091 MR 1959915

[8] B. Collins, A. Guionnet, and E. Maurel-Segala, Asymptotics of unitary and orthogonal matrix integrals. *Adv. Math.* **222** (2009), no. 1, 172–215 Zbl 1184.15024 MR 2531371

[9] B. Collins, R. Gurau, and L. Lionni, The tensor Harish–Chandra–Itzykson–Zuber integral I: Weingarten calculus and a generalization of monotone Hurwitz numbers. *J. Eur. Math. Soc. (JEMS)* **26** (2024), no. 5, 1851–1897 Zbl 07855098 MR 4735823

[10] B. Collins, S. Matsumoto, and J. Novak, The Weingarten calculus. *Notices Amer. Math. Soc.* **69** (2022), no. 5, 734–745 Zbl 1527.60007 MR 4415894

[11] B. Collins and P. Śniady, Integration with respect to the Haar measure on unitary, orthogonal and symplectic group. *Comm. Math. Phys.* **264** (2006), no. 3, 773–795 Zbl 1108.60004 MR 2217291

[12] P. Di Francesco, P. Ginsparg, and J. Zinn-Justin, 2D gravity and random matrices. *Phys. Rep.* **254** (1995), no. 1-2, 1–133 MR 1320471

[13] G. Dubach and Y. Peled, On words of non-Hermitian random matrices. *Ann. Probab.* **49** (2021), no. 4, 1886–1916 Zbl 1467.60007 MR 4260470

[14] J. R. Edmonds, A combinatorial representation for polyhedral surfaces. *Notices Am. Math. Soc.* **7** (1960), no. 48, 1–646

[15] N. M. Ercolani and K. D. T.-R. McLaughlin, Asymptotics of the partition function for random matrices via Riemann–Hilbert techniques and applications to graphical enumeration. *Int. Math. Res. Not.* (2003), no. 14, 755–820 Zbl 1140.82307 MR 1953782

[16] B. Eynard, Formal matrix integrals and combinatorics of maps. In *Random matrices, random processes and integrable systems*, pp. 415–442, CRM Ser. Math. Phys., Springer, New York, 2011 Zbl 1257.81052 MR 2858440

[17] B. Eynard, *Counting surfaces*. Prog. Math. Phys. 70, Birkhäuser/Springer, Cham, 2016 Zbl 1338.81005 MR 3468847

[18] B. Eynard and N. Orantin, Algebraic methods in random matrices and enumerative geometry. 2008, arXiv:0811.3531v1

[19] I. P. Goulden, M. Guay-Paquet, and J. Novak, Monotone Hurwitz numbers and the HCIZ integral. *Ann. Math. Blaise Pascal* **21** (2014), no. 1, 71–89 Zbl 1296.05202 MR 3248222

[20] D. J. Gross, T. Piran, and S. Weinberg, Two-dimensional quantum gravity and random surfaces. In *Proceedings of the Eighth Winter School held in Jerusalem, December 27, 1990–January 4, 1991*, pp. xiv+240, Jerusalem Winter School for Theoretical Physics 8, World Scientific, River Edge, NJ, 1992 MR 1165557

[21] A. Guionnet and E. Maurel-Segala, Combinatorial aspects of matrix models. *ALEA Lat. Am. J. Probab. Math. Stat.* **1** (2006), 241–279 Zbl 1110.15021 MR 2249657

[22] A. Guionnet and E. Maurel-Segala, Second order asymptotics for matrix models. *Ann. Probab.* **35** (2007), no. 6, 2160–2212 Zbl 1129.15020 MR 2353386

[23] A. Guionnet and J. Novak, Asymptotics of unitary multimatrix models: the Schwinger–Dyson lattice and topological recursion. *J. Funct. Anal.* **268** (2015), no. 10, 2851–2905 Zbl 1335.46057 MR 3331788

[24] J. Harer and D. Zagier, The Euler characteristic of the moduli space of curves. *Invent. Math.* **85** (1986), no. 3, 457–485 Zbl 0616.14017 MR 0848681

[25] Harish-Chandra, Differential operators on a semisimple Lie algebra. *Amer. J. Math.* **79** (1957), 87–120 Zbl 0072.01901 MR 0084104

[26] C. Itzykson and J. B. Zuber, The planar approximation. II. *J. Math. Phys.* **21** (1980), no. 3, 411–421 Zbl 0997.81549 MR 0562985

[27] P. Johnson, Hurwitz numbers, ribbon graphs, and tropicalization. In *Tropical geometry and integrable systems*, pp. 55–72, Contemp. Math. 580, American Mathematical Society, Providence, RI, 2012 Zbl 1317.14118 MR 2985387

[28] M. Kontsevich, Intersection theory on the moduli space of curves and the matrix Airy function. *Comm. Math. Phys.* **147** (1992), no. 1, 1–23 Zbl 0756.35081 MR 1171758

[29] G. Kreweras, Sur les partitions non croisées d'un cycle. *Discrete Math.* **1** (1972), no. 4, 333–350 Zbl 0231.05014 MR 0309747

[30] S. K. Lando and A. K. Zvonkin, *Graphs on surfaces and their applications*. Encyclopaedia Math. Sci. 141, Springer, Berlin, 2004 Zbl 1040.05001 MR 2036721

[31] E. Maurel-Segala, High order expansion of matrix models and enumeration of maps. 2006, arXiv:math/0608192v1

[32] M. L. Mehta, A method of integration over matrix variables. *Comm. Math. Phys.* **79** (1981), no. 3, 327–340 Zbl 0471.28007 MR 0627056

[33] J. A. Mingo and A. Nica, Annular noncrossing permutations and partitions, and second-order asymptotics for random matrices. *Int. Math. Res. Not.* (2004), no. 28, 1413–1460 Zbl 1071.05006 MR 2052516

[34] J. A. Mingo, P. Śniady, and R. Speicher, Second order freeness and fluctuations of random matrices. II. Unitary random matrices. *Adv. Math.* **209** (2007), no. 1, 212–240 Zbl 1122.46045 MR 2294222

[35] J. Novak, On the complex asymptotics of the HCIZ and BGW integrals. 2020, arXiv:2006.04304v1

[36] J. I. Novak, Jucys–Murphy elements and the unitary Weingarten function. In *Noncommutative harmonic analysis with applications to probability II*, pp. 231–235, Banach Center Publ. 89, Institute of Mathematics of the Polish Academy of Sciences, Warsaw, 2010 Zbl 1219.05196 MR 2730867

[37] S. Samuel, $\mathrm{U}(N)$ integrals, $1/N$, and the De Wit–'t Hooft anomalies. *J. Math. Phys.* **21** (1980), no. 12, 2695–2703 MR 0597583

[38] T. P. Speed, Cumulants and partition lattices. *Austral. J. Statist.* **25** (1983), no. 2, 378–388 Zbl 0538.60023 MR 0725217

[39] G. 't Hooft, A planar diagram theory for strong interactions. *Nucl. Phys. B* **72** (1974), no. 3, 461–473

[40] W. T. Tutte, On the enumeration of planar maps. *Bull. Amer. Math. Soc.* **74** (1968), 64–74 Zbl 0157.31101 MR 0218276

[41] D. Voiculescu, The analogues of entropy and of Fisher's information measure in free probability theory. VI. Liberation and mutual free information. *Adv. Math.* **146** (1999), no. 2, 101–166 Zbl 0956.46045 MR 1711843

[42] D. Weingarten, Asymptotic behavior of group integrals in the limit of infinite rank. *J. Mathematical Phys.* **19** (1978), no. 5, 999–1001 Zbl 0388.28013 MR 0471696

[43] P. Zinn-Justin and J.-B. Zuber, On some integrals over the $\mathrm{U}(N)$ unitary group and their large N limit. *J. Phys. A* **36** (2003), no. 12, 3173–3193 Zbl 1074.82013 MR 1986413

[44] A. Zvonkin, Matrix integrals and map enumeration: an accessible introduction. *Math. Comput. Modelling* **26** (1997), no. 8-10, 281–304 Zbl 1185.81083 MR 1492512

Communicated by Adrian Tanasă

Received 8 August 2023; revised 13 September 2024.

Thomas Buc-d'Alché
UMPA UMR 5669, CNRS, 46, allée d'Italie, 69007 Lyon, France;
thomas.buc-dalche@ens-lyon.fr