Available online at www.sciencedirect.com

ScienceDirect

Advances in Mathematics 222 (2009) 172–215

ADVANCES IN Mathematics

www.elsevier.com/locate/aim

Asymptotics of unitary and orthogonal matrix integrals

Benoît Collins [a,b], Alice Guionnet [c], Edouard Maurel-Segala [d,*]

[a] *CNRS, ICJ, Lyon 1 Claude Bernard University, 43 bd du 11 Novembre 1918, 69622 Villeurbanne, France*
[b] *Dept. Math. Stat., University of Ottawa, Canada*
[c] *UMPA, CNRS UMR 5669, ENS Lyon, 46 allée d'Italie, 69007 Lyon, France*
[d] *LPMA, Universités Paris VI, 175 rue du Chevaleret, 75013 Paris, France*

Received 9 November 2007; accepted 24 March 2009

Available online 14 May 2009

Communicated by Dan Voiculescu

Abstract

In this paper, we prove that in small parameter regions, arbitrary unitary matrix integrals converge in the large N limit and match their formal expansion. Secondly we give a combinatorial model for our matrix integral asymptotics and investigate examples related to free probability and the HCIZ integral. Our convergence result also leads us to new results of smoothness of microstates. We finally generalize our approach to integrals over the orthogonal group.

MSC: 15A52; 46L54

Keywords: Matrix integrals; HCIZ integral; Schwinger–Dyson equation

0. Introduction

Matrix integrals provide models for physical systems (2D quantum gravitation, gauge theory, renormalization, etc.), and generating series for a wide family of combinatorial objects (see e.g. [20,28]).

Gaussian integrals are the most studied. It was shown by Brézin, Itzykson, Parisi and Zuber [7] that perturbations of Gaussian matrix integrals expand formally as a generating function of maps,

* Corresponding author.

E-mail addresses: collins@math.univ-lyon1.fr (B. Collins), aguionne@umpa.ens-lyon.fr (A. Guionnet), edouard.maurel-segala@upmc.fr (E. Maurel-Segala).

0001-8708/$ – see front matter
doi:10.1016/j.aim.2009.03.019

sorted by their genus when the dimension N of the matrices is regarded as a parameter. Such 'topological' expansions were also shown to hold in the large N limit, and then to match with the formal expansion on a mathematical level of rigor by two authors [16,17,23] and previously in the one matrix case in [1,2] and [12]. The relation of Gaussian matrices with the enumeration of maps is a consequence of Wick calculus—or equivalently, Feynman diagrams—see [28] for a good introduction. According to 't Hooft [20], such topological expansion should hold in a more general context. On the other hand, random matrices also provide finite approximations to free operators. Then, unitary matrices following the Haar measure appear to be even more closely related to the notion of freeness than Gaussian matrices, see [26,30,33] and [21].

In this article, we focus on matrix integrals given by

$$I_N\big(V, A_i^N\big) := \int_{\mathcal{U}_N^m} e^{N \operatorname{Tr}(V(U_i, U_i^*, A_i^N, 1 \leqslant i \leqslant m))}\, dU_1 \cdots dU_m \tag{1}$$

where $(A_i^N, 1 \leqslant i \leqslant m)$ are $N \times N$ deterministic uniformly bounded matrices, dU denotes the Haar measure on the unitary group $\mathcal{U}_N$ (normalized so that $\int_{\mathcal{U}_N} dU = 1$) and V is a polynomial function in the non-commutative variables $(U_i, U_i^*, A_i^N, 1 \leqslant i \leqslant m)$. Tr denotes the usual trace on $N \times N$ matrices given by $\operatorname{Tr}(A) = \sum_{i=1}^N A_{ii}$.

We will assume without loss of generality that the matrices $(A_i^N, 1 \leqslant i \leqslant m)$ are Hermitian matrices, up to rewrite the potential V in terms of their Hermitian and anti-Hermitian parts. We will study in this article the first order asymptotics of matrix integrals given by (1) when the joint distribution of the $(A_i^N, 1 \leqslant i \leqslant m)$ converges; namely for all polynomial function P in m non-commutative indeterminates

$$\lim_{N \to \infty} \frac{1}{N} \operatorname{Tr}\big(P\big(A_i^N, 1 \leqslant i \leqslant m\big)\big) = \tau(P) \tag{2}$$

for some linear functional τ on the set of polynomials.

For technical reasons, we assume that the polynomial V satisfies $\operatorname{Tr}(V(U_i, U_i^*, A_i^N, 1 \leqslant i \leqslant m)) \in \mathbb{R}$, for all $U_i \in \mathcal{U}_N$, all Hermitian matrices A_i^N, for all $i \in \{1, \dots, m\}$ and $N \in \mathbb{N}$.

Under those very general assumptions, the only result proved so far is the formal convergence of these matrix integrals. Namely, it was proved in [8] by one author that for each k, the quantity

$$\frac{\partial^k}{\partial z^k} N^{-2} \log \int_{\mathcal{U}_N^m} e^{zN \operatorname{Tr}(V(U_i, U_i^*, A_i^N, 1 \leqslant i \leqslant m))}\, dU_1 \cdots dU_m \bigg|_{z=0}$$

converges towards a constant $f_k(V, \tau)$ depending only on the limiting distribution of the A_i^N's and V. Besides, if V is polynomial with integer coefficients, then $f_k(V, \tau)$ is a polynomial function with integer coefficients of the limit moments of the A_i^N's.

In this paper we will answer affirmatively to the following, previously open questions:

(1) Does the limit of the matrix integrals exist for small parameters z?
(2) Does the power series $\sum_k \frac{z^k}{k!} f_k(V, \tau)$ have a strictly positive radius of convergence?
(3) Is the limit of the matrix integral equal to the sum of the power series?

The following theorem is a precise description of our results:

Theorem 0.1. *Under the above hypotheses and if we further assume that the spectral radius of the matrices $(A_i^N, 1 \leqslant i \leqslant m, N \in \mathbb{N})$ is uniformly bounded (by say M), then there exists $\varepsilon = \varepsilon(M, V) > 0$ so that for $z \in [-\varepsilon, \varepsilon]$, the limit*

$$F_{V,\tau}(z) := \lim_{N\to\infty} \frac{1}{N^2} \log \int_{\mathcal{U}_N^m} e^{zN \operatorname{Tr}(V(U_i, U_i^*, A_i^N, 1\leqslant i\leqslant m))}\, dU_1 \cdots dU_m$$

exists. Moreover, $F_{V,\tau}(z)$ is an analytic function of $z \in \mathbb{C} \cap B(0, \varepsilon) = \{z \in \mathbb{C}\colon |z| \leqslant \varepsilon\}$ and for all $k \in \mathbb{N}$,

$$\frac{\partial^k}{\partial z^k} F_{V,\tau}(z)\bigg|_{z=0} = f_k(V, \tau).$$

This also implies that the series $F_{V,\tau}(z)$ has a positive radius of convergence, a result which had not been proved by the techniques of [8] based on Weingarten functions.

Our approach is based on non-commutative differential calculus (in particular on the resulting Schwinger–Dyson or Master loop equations) and perturbation analysis as developed in the context of Gaussian matrices in [16,17,23]. Another possibility to prove the equality between real and formal limits would have been to show convergence of the integrals for complex parameters z. We have not yet been able to follow this line successfully, and this remains an open question.

An important example of unitary matrix integral is the so-called spherical integral, studied by Harish-Chandra and by Itzykson and Zuber,

$$\mathit{HCIZ}(A, B) := \int_{U\in\mathcal{U}_N} e^{N \operatorname{Tr}(U^* A U B)}\, dU.$$

This integral is of fundamental importance in analytic Lie theory and was computed for the first time by Harish-Chandra in [19]. In the last two decades it has also become an issue to study its large dimension asymptotics [11,15,18,35].

Theorem 0.1 holds true for the HCIZ integral. It thus relates the results of [8] (which computed the formal limit of the HCIZ integral) and those of [18] (where the limit of $\mathit{HCIZ}(A, B)$ was obtained (regardless of any small parameters assumptions) by using large deviations techniques). Let us recall the limit found in [18], when A and B are Hermitian matrices. Let us define

$$I(\mu) = \frac{1}{2}\mu\big(x^2\big) + \frac{1}{2}\iint \log|x-y|\, d\mu(x)\, d\mu(y).$$

If $\mu_A \in \mathcal{P}(\mathbb{R})$ (resp. μ_B) denote the limiting spectral measure of A (resp. B), assume that $I(\mu_A)$ and $I(\mu_B)$ are finite. Then, according to [18], the limit of $N^{-2} \log \mathit{HCIZ}(A, B)$ is given by

$$\begin{aligned} I(\mu_A, \mu_B) &:= \lim_{N\to\infty} \frac{1}{N^2} \log \mathit{HCIZ}(A, B) \\ &= -I(\mu_A) - I(\mu_B) - \frac{1}{2} \inf_{\rho,m} \left\{ \int_0^1 \int \left(\frac{m_t(x)^2}{\rho_t(x)} + \frac{\pi^2}{3}\rho_t(x)^3 \right) dx\, dt \right\} \end{aligned} \tag{3}$$

where the infimum is taken over m, ρ so that the measure-valued process $\mu_t(dx) = \rho_t(x)\,dx$ is a continuous process, $\mu_0 = \mu_A$, $\mu_1 = \mu_B$ and

$$\partial_t \rho_t(x) + \partial_x m_t(x) = 0.$$

The inf over (ρ_t, m_t) is taken (see [14]) at the solution of an Euler equation for isentropic flow with negative pressure $-\frac{\pi^2}{3}\rho^3$.

Theorem 0.1 shows that for some $\beta_0 > 0$, $I(\mu_{\beta A}, \mu_{\beta B})$ is real analytic for $0 \leqslant \beta < \beta_0$, a result which is not obvious from formula (3). Moreover, the coefficients of this expansion count certain planar graphs (see Section 5), as summarized in the following theorem.

Theorem 0.2. *Denote $\sqrt{\beta} \sharp \mu$ the probability measure*

$$\beta \sharp \mu(f) = \int f(\beta x)\,d\mu(x).$$

Assume that μ_A and μ_B are two compactly supported probability measures. Then, there exists $\beta_0 > 0$ such that for all $\beta \in [0, \beta_0]$,

$$I(\beta \sharp \mu_A, \beta \sharp \mu_B) = \sum_{n \geqslant 0} \beta^n \mathbb{M}_n(\mu_A, \mu_B)$$

converges absolutely. Moreover, we have

$$\mathbb{M}_n(\mu_A, \mu_B) = \sum_{m \text{ admissible maps of } \Sigma_n} M_m(\mu_A, \mu_B).$$

Σ_n is the set of planar maps drawn above n vertices defined as stars of type U^*AUB by gluing pairwise oriented arrows and possibly rings and $M_m(\mu_A, \mu_B)$ is the weight of the map.

Moreover, this result extends when A and B are not Hermitian as follows.

Theorem 0.3. *Assume that (A_N, B_N) is a sequence of matrices with spectral radius bounded by one, such that $\operatorname{Tr}(U_N^* A_N U_N B_N)$ is real for any $N \times N$ unitary matrix U_N and the joint moments of A_N, A_N^* and B_N, B_N^* converge:*

$$\lim_{N \to \infty} \frac{1}{N} \operatorname{Tr}\bigl(P\bigl(A_N, A_N^*\bigr)\bigr) =: \tau_A(P), \qquad \lim_{N \to \infty} \frac{1}{N} \operatorname{Tr}\bigl(P\bigl(B_N, B_N^*\bigr)\bigr) =: \tau_B(P).$$

Then, there exists $\beta_0 > 0$ such that for all $\beta \in [0, \beta_0]$,

$$\frac{1}{N^2} \log HCIZ(\beta A_N, \beta B_N)$$

converges as N goes to infinity to a limit $I(\tau_{\beta A}, \tau_{\beta B})$.

We refer the reader to Section 5 for the definitions of stars, admissible maps and weights. Our definition of planar maps is more complicated than those arising in the topological expansion of Gaussian matrix models (and which are directly related with Wick Gaussian calculus and Feynman diagrams): indeed, the sums are signed and we have a notion of admissibility. However it was an open question in mathematical physics to have a graphical model for unitary integrals (see [35]). Moreover, this graphical interpretation gives a new understanding of cumulants formulae (see Section 6).

The convergence of other integrals was still unknown and it is one of the points of this paper to show their convergence. We use it to study Voiculescu's microstates entropy evaluated at a set of laws which are small perturbations of the law of free variables, and prove regularity of microstates

Theorem 0.4. *For tracial states μ satisfying suitable assumptions described in Theorem* 8.1 *and with $\Gamma_R(\mu,\varepsilon,k)$ a microstates of μ*

$$\begin{aligned}\chi(\mu) &:= \liminf_{\substack{\varepsilon\downarrow 0\\ k\uparrow\infty}} \liminf_{N\to\infty} \frac{1}{N^2}\log\mu_N^{\otimes m}\big(\Gamma_R(\mu,\varepsilon,k)\big)\\ &= \limsup_{\substack{\varepsilon\downarrow 0\\ k\uparrow\infty}} \limsup_{N\to\infty} \frac{1}{N^2}\log\mu_N^{\otimes m}\big(\Gamma_R(\mu,\varepsilon,k)\big)\end{aligned}$$

and a formula for $\chi(\mu)$ can be given.

This result generalizes Section 4 in [16].

The paper is organized as follows: after setting our working framework (Section 1), we study the action of perturbations upon the integral $I_N(V,A_i^N)$ and deduce some properties of the related Gibbs measure; namely that the so-called empirical distribution of the matrices under this Gibbs measure satisfies asymptotically an equation called the Schwinger–Dyson equation (Section 2). Then, we study this equation and obtain uniqueness for parameters of the potential V small enough (Section 3) and analyticity (Section 4). We also describe a (new) combinatorial solution of the Schwinger–Dyson equation (Section 5) and therefore of the first order of unitary matrix integrals. We deduce applications of these results to free probability (Section 6) and to the convergence of matrix integrals $I_N(V,A_i^N)$ (Section 7). Moreover, we point out some consequence of our result for free entropy (Section 8). Finally, in Section 9, we consider the case where the integration is over the orthogonal group instead of the unitary group, and we show that the first order of such integrals is the same, up to a rescaling of the potential.

1. Notations

Let $\mathcal{U}_N$ be the set of $N\times N$ unitary matrices, $\mathcal{M}_N$ the set of $N\times N$ matrices with complex entries, $\mathcal{H}_N$ the subset of Hermitian matrices of $\mathcal{M}_N$ and $\mathcal{A}_N$ the subset of anti-Hermitian matrices of $\mathcal{M}_N$. Throughout this article, m will be a fixed integer. We denote by $(A_i^N)_{1\leqslant i\leqslant m}$ an m-tuple of $N\times N$ Hermitian matrices. We shall assume that the sequence $(A_i^N)_{1\leqslant i\leqslant m}$ is uniformly bounded for the operator norm, and without loss of generality that they are bounded by one,

$$\sup_{N,i}\|A_i^N\|_\infty = \sup_{N,i}\lim_{p\to\infty}\big(\mathrm{Tr}\big(\big(A_i^N\big)^{2p}\big)\big)^{\frac{1}{2p}} \leqslant 1.$$

1.1. Free $*$-algebra

Let $\mathbb{C}\langle (U_i, U_i^*, A_i)_{1\leqslant i\leqslant m}\rangle$ be the set of polynomials in the non-commutative indeterminates $(U_i, U_i^*, A_i)_{1\leqslant i\leqslant m}$ with the relation

$$U_i U_i^* = U_i^* U_i = 1.$$

Note that in general we may want to consider models with a number of "deterministic" indeterminates A_i different from the number of "random unitary" indeterminates U_i, but this general case can be obtained from the previous one by looking only at a sub-algebra and our convention simplifies a bit the notations. The algebra $\mathbb{C}\langle (U_i, U_i^*, A_i)_{1\leqslant i\leqslant m}\rangle$ is equipped with the involution $*$ so that $A_i^* = A_i$, $(U_i)^* = U_i^*$; $(U_i^*)^* = U_i$ and for any $X_1,\dots,X_n \in (U_i, U_i^*, A_i)_{1\leqslant i\leqslant m}$, any $z\in\mathbb{C}$,

$$(zX_1X_2\cdots X_{n-1}X_n)^* = \bar{z}X_n^*X_{n-1}^*\cdots X_2^*X_1^*.$$

Note that for any $U_i \in \mathcal{U}_N$, $A_i\in\mathcal{H}_N$, and $P\in\mathbb{C}\langle (U_i, U_i^*, A_i)_{1\leqslant i\leqslant m}\rangle$,

$$\big(P\big(U_i, U_i^*, A_i, 1\leqslant i\leqslant m\big)\big)^* = P^*\big(U_i, U_i^*, A_i, 1\leqslant i\leqslant m\big)$$

where in the left-hand side $*$ denotes the standard involution on $\mathcal{M}_N$. We denote $\mathbb{C}\langle (U_i, U_i^*, A_i)_{1\leqslant i\leqslant m}\rangle_{sa}$ the set of self-adjoint polynomials; $P = P^*$, and $\mathbb{C}\langle (U_i, U_i^*, A_i)_{1\leqslant i\leqslant m}\rangle_{a}$ the set of anti-self-adjoint polynomials; $P^* = -P$. In the sequel, except when something different is explicitly assumed, we shall make the hypothesis that the potential V belongs to $\mathbb{C}\langle (U_i, U_i^*, A_i)_{1\leqslant i\leqslant m}\rangle_{sa}$, which insures that $\mathrm{Tr}(V((U_i, U_i^*, A_i^N)_{1\leqslant i\leqslant m}))$ is real-valued for all $U_i\in\mathcal{U}_N$ and $A_i^N\in\mathcal{H}_N$. Conversely, any potential V such that $\mathrm{Tr}(V((U_i, U_i^*, A_i^N)_{1\leqslant i\leqslant m}))$ is real-valued for all $U_i\in\mathcal{U}_N$ and $A_i^N\in\mathcal{H}_N$ is self-adjoint up to the addition of some commutators (which does not change the trace). Indeed, this implies that $\mathrm{Tr}((V-V^*)((U_i, U_i^*, A_i^N)_{1\leqslant i\leqslant m}))$ vanishes for all $U_i\in\mathcal{U}_N$. This insures that $V - V^* = \sum_l P_lQ_l - Q_lP_l$ for some polynomials P_l, Q_l, cf. [9, Lemma 2.9] for a probabilistic proof or [22, Proposition 2.3] for a direct proof (in the real symmetric case, but directly adaptable to the Hermitian case). Then, $W := V + \sum_l (Q_lP_l - P_lQ_l)/2$ is self-adjoint.

1.2. Non-commutative derivatives

On $\mathbb{C}\langle (U_i, U_i^*, A_i)_{1\leqslant i\leqslant m}\rangle$, we define the non-commutative derivatives ∂_i, $1\leqslant i\leqslant m$, given by the linear form such that

$$\partial_i A_j = 0, \qquad \partial_i U_j = 1_{i=j} U_j\otimes 1, \qquad \partial_i U_j^* = -1_{i=j} 1\otimes U_j^*, \quad \forall j,$$

and satisfying the Leibnitz rule, namely, for $P, Q\in\mathbb{C}\langle (U_i, U_i^*, A_i)_{1\leqslant i\leqslant m}\rangle$,

$$\partial_i(PQ) = \partial_i P\times(1\otimes Q) + (P\otimes 1)\times\partial_i Q. \tag{4}$$

Here, $\times$ denotes the product $P_1\otimes Q_1\times P_2\otimes Q_2 = P_1P_2\otimes Q_1Q_2$. We also let D_i be the corresponding *cyclic* derivatives such that if $m(A\otimes B) = BA$, then $D_i = m\circ\partial_i$.

If q is a monomial in $\mathbb{C}\langle (U_i, U_i^*, A_i)_{1\leqslant i\leqslant m}\rangle$, we more specifically have

$$\partial_i q = \sum_{q=q_1 U_i q_2} q_1 U_i \otimes q_2 - \sum_{q=q_1 U_i^* q_2} q_1 \otimes U_i^* q_2, \tag{5}$$

$$D_i q = \sum_{q=q_1 U_i q_2} q_2 q_1 U_i - \sum_{q=q_1 U_i^* q_2} U_i^* q_2 q_1. \tag{6}$$

We set, for $P, Q, R \in \mathbb{C}\langle (U_i, U_i^*, A_i)_{1 \leqslant i \leqslant m} \rangle$, $P \otimes Q \sharp R := PRQ$. Note that if R is an anti-self-adjoint polynomial and we denote by

$$\mathbf{U}^\varepsilon := \big(U_1, \ldots, U_{i-1}, U_i e^{\varepsilon R}, U_{i+1}, \ldots, U_m\big),$$

we have

$$\mathrm{Diff}_i\, P.R := \lim_{\varepsilon \to 0} \varepsilon^{-1}\big(P\big(\mathbf{U}^\varepsilon\big) - P(\mathbf{U})\big) = \partial_i P \sharp R. \tag{7}$$

1.3. Bounded tracial states

Let $\mathcal{T}$ be the set of tracial states on the algebra generated by the variables $(U_i, U_i^*, A_i)_{1 \leqslant i \leqslant m}$, i.e. the set of linear forms on $\mathbb{C}\langle (U_i, U_i^*, A_i)_{1 \leqslant i \leqslant m} \rangle$ such that for all $P, Q \in \mathbb{C}\langle (U_i, U_i^*, A_i)_{1 \leqslant i \leqslant m} \rangle$,

$$\mu\big(PP^*\big) \geqslant 0, \qquad \mu(PQ) = \mu(QP), \qquad \mu(1) = 1.$$

Throughout this article, we restrict ourselves to tracial states $\mu \in \mathcal{T}$ such that

$$\mu\big((A_i)^{2n}\big) \leqslant 1 \quad \forall n \in \mathbb{N},\ \forall i \in \{1, \ldots, m\}.$$

We denote $\mathcal{M}$ this subset of $\mathcal{T}$.

Note that for any monomial $q \in \mathbb{C}\langle (U_i, U_i^*, A_i)_{1 \leqslant i \leqslant m} \rangle$, Hölder's inequality implies that for any $\mu \in \mathcal{M}$,

$$\mu\big(qq^*\big) \leqslant 1. \tag{8}$$

We endow $\mathcal{M}$ with its weak topology: μ_n converges to μ if and only if for all $P \in \mathbb{C}\langle (U_i, U_i^*, A_i)_{1 \leqslant i \leqslant m} \rangle$,

$$\lim_{n \to \infty} \mu_n(P) = \mu(P).$$

If we give to the set of polynomials the norm l^1 (i.e. the norm of polynomial is the sum of the modulus of its coefficients) then Eq. (8) proves that $\mathcal{M}$ is the unit ball of $\mathcal{T}$ for the weak* topology. Thus by Banach Alaoglu's theorem, $\mathcal{M}$ is a compact metric space.

We denote $\hat{\mu}^N$ the empirical distribution of matrices $A_i^N \in \mathcal{H}_N$ and $U_i \in \mathcal{U}_N$ which is given for all $P \in \mathbb{C}\langle (U_i, U_i^*, A_i)_{1 \leqslant i \leqslant m} \rangle$ by

$$\hat{\mu}^N(P) = \frac{1}{N} \mathrm{Tr}\big(P\big(U_i, U_i^*, A_i^N, 1 \leqslant i \leqslant m\big)\big).$$

This object will be of crucial interest for us.

The notation $\mathcal{M}|_{(A_i)_{1\leqslant i\leqslant m}}$ stands for the set of tracial states of $\mathcal{M}$ restricted to the algebra generated by the $(A_i)_{1\leqslant i\leqslant m}$. In particular, the limiting distribution τ given by (2) belongs to $\mathcal{M}|_{(A_i)_{1\leqslant i\leqslant m}}$.

1.4. Tracial power states

Let $V\in\mathbb{C}\langle(U_i,U_i^*,A_i)_{1\leqslant i\leqslant m}\rangle_{sa}$ and μ_V^N be the probability distribution on $\mathcal{U}_N^m$ given by

$$\mu_V^N(dU_1,\dots,dU_m)=I_N\big(V,A_i^N\big)^{-1}\exp\big(N\operatorname{Tr}(V)\big)\,dU_1\cdots dU_m.$$

We define, for all $P\in\mathbb{C}\langle(U_i,U_i^*,A_i)_{1\leqslant i\leqslant m}\rangle$,

$$\bar{\mu}_V^N(P):=E_{\mu_V^N}\big[\hat{\mu}^N(P)\big]:=\frac{\int\frac{1}{N}\operatorname{Tr}Pe^{N\operatorname{Tr}V}\,dU_1\cdots dU_n}{\int e^{N\operatorname{Tr}V}\,dU_1\cdots dU_n}.$$

In the following, an n-tuple of monomials $(q_i)_{1\leqslant i\leqslant n}$ in $\mathbb{C}\langle(U_i,U_i^*,A_i)_{1\leqslant i\leqslant m}\rangle$ will be fixed and we shall take $V=V_{\mathbf{t}}=\sum_{i=1}^n t_iq_i$. Then, $\bar{\mu}_{V_{\mathbf{t}}}^N(P)$ can be expanded as a power series in the t_i's;

$$\bar{\mu}_{V_{\mathbf{t}}}^N(P):=\sum_{\mathbf{k}\in\mathbb{N}^n}\frac{\mathbf{t}^{\mathbf{k}}}{\mathbf{k}!}\frac{\partial^{|\mathbf{k}|}}{\prod_i\partial t_i^{k_i}}\bigg|_{t_i=0}\frac{E[\hat{\mu}^N(P)e^{N^2\hat{\mu}^N(V_{\mathbf{t}})}]}{E[e^{N^2\hat{\mu}^N(V_{\mathbf{t}})}]}. \tag{9}$$

We will call μ a 'tracial power state' of $\mathcal{M}$ if and only if it is a linear map

$$\mu:\mathbb{C}\big\langle\big(U_i,U_i^*,A_i\big)_{1\leqslant i\leqslant m}\big\rangle\to\mathbb{C}[[\mathbf{t}]]$$

with for all a,b, $\mu(ab)=\mu(ba)$. Here $\mathbb{C}[[\mathbf{t}]]$ is the algebra of power series in the variables $t_1,\dots,t_n$. In particular, we may view $\mu_{V_{\mathbf{t}}}^N$ as a tracial power state of $\mathcal{M}$. The space of tracial power states is equipped with the topology of convergence coefficient by coefficient.

1.5. Cumulants

The classical cumulants $\{C_k\}_{k\geqslant 0}$ are defined via their formal generating function:

$$\log E\big(e^{tX}\big)=\sum_{k\geqslant 0}t^kC_k(X,\dots,X)/k!.$$

This equality holds also for t in a complex neighborhood of 0 if X is bounded. We also define the cumulants $C_{\mathbf{k}}$ for $\mathbf{k}$ in $\mathbb{N}^n$:

$$\log E\big(e^{t_1X_1+\cdots+t_nX_n}\big)=\sum_{\mathbf{k}\in\mathbb{N}^n}\mathbf{t}^{\mathbf{k}}C_{\mathbf{k}}(X_1,\dots,X_n)/\mathbf{k}!$$

where $\mathbf{k}=(k_1,\dots,k_n)$, $\mathbf{k}!=\prod_i k_i!$, $|\mathbf{k}|=\sum_i k_i$ and $\mathbf{t}^{\mathbf{k}}=\prod_i t_i^{k_i}$. Note that:

$$C_{\mathbf{k}}(X_1,\dots,X_k)=C_{|\mathbf{k}|}(X_1,\dots,X_1,\dots,X_n,\dots,X_n)$$

where in the previous list the variable X_i appears k_i times.

Let us recall some properties of these cumulants.

Proposition 1.1. *The following two statements hold true:*

(1)
$$\frac{E(Ye^{t_1X_1+\cdots+t_nX_n})}{E(e^{t_1X_1+\cdots+t_nX_n})}=\sum_{\mathbf{k}\in\mathbb{N}^n}\mathbf{t}^{\mathbf{k}}C_{1,\mathbf{k}}(Y,X_1,\dots,X_n)/\mathbf{k}!;$$

(2)
$$\frac{E(YZe^{t_1X_1+\cdots+t_nX_n})}{E(e^{t_1X_1+\cdots+t_nX_n})}-\frac{E(Ye^{t_1X_1+\cdots+t_nX_n})}{E(e^{t_1X_1+\cdots+t_nX_n})}\frac{E(Ze^{t_1X_1+\cdots+t_nX_n})}{E(e^{t_1X_1+\cdots+t_nX_n})}$$
$$=\sum_{k\geqslant 0}\mathbf{t}^{\mathbf{k}}C_{1,1,\mathbf{k}}(Y,Z,X_1,\dots,X_n)/\mathbf{k}!.$$

Proof. Item (1) is obtained by replacing $t_1X_1+\cdots+t_nX_n$ by $yY+t_1X_1+\cdots+t_nX_n$ and differentiating the generating function of the cumulants in y at $y=0$.

Item (2) is obtained by replacing tX by $yY+zZ+tX$ and differentiating the equality defining the cumulants in y and z at $y,z=0$. □

2. Matrix models

We first investigate the asymptotic behavior of the random state $\hat{\mu}^N$ under μ_V^N as a random tracial state. We then consider $\bar{\mu}_V^N=\mu_V^N(\hat{\mu}^N)$ evaluated at a polynomial and study its convergence when N goes to infinity as a power series in the parameters of the potential V. We show that they satisfy asymptotically the same type of equations called the Schwinger–Dyson (or Master loop) equations.

2.1. Behavior of $\hat{\mu}^N$

The main result of this section is the following

Theorem 2.1. *Assume that V is self-adjoint. For all polynomial $P\in\mathbb{C}\langle(U_i,U_i^*,A_i)_{1\leqslant i\leqslant m}\rangle$,*

$$\lim_{N\to\infty}\left\{\hat{\mu}^N\otimes\hat{\mu}^N(\partial_iP)+\hat{\mu}^N(D_iVP)\right\}=0\quad \mu_V^N\ a.s.$$

In particular, any limit point $\mu\in\mathcal{M}$ of $\hat{\mu}^N$ under μ_V^N satisfies the Schwinger–Dyson equation

$$\mu\otimes\mu(\partial_iP)+\mu(D_iVP)=0 \tag{10}$$

for all $P\in\mathbb{C}\langle(U_i,U_i^*,A_i)_{1\leqslant i\leqslant m}\rangle$ and $\mu|_{(A_i)_{1\leqslant i\leqslant m}}=\tau$.

The idea of the proof, rather common in quantum field theory and successfully used in [16,17,23], is to obtain equations on $\hat{\mu}^N$ by performing an infinitesimal change of variables in $I_N(V,A_i^N)$. More precisely we make the change of variables $\mathbf{U}=(U_1,\dots,U_m)\in\mathcal{U}_N^m\to\Psi(\mathbf{U})=(\Psi_1(\mathbf{U}),\dots,\Psi_m(\mathbf{U}))\in\mathcal{U}_N^m$ with

$$\Psi_j(\mathbf{U})=U_je^{\frac{\lambda}{N}P_j(\mathbf{U})}$$

where the P_j are anti-self-adjoint polynomials (i.e. $P_j^* = -P_j$). This change of variables becomes very close to the identity as N goes to infinity, thus justifying the terminology "infinitesimal."

Lemma 2.1. *The function Ψ is a local diffeomorphism and its Jacobian J_Ψ has the following expansion when N goes to infinity*

$$\left|\det J_\Psi(\mathbf{U})\right| = e^{\frac{\lambda}{N}\sum_i \mathrm{Tr}\otimes\mathrm{Tr}(\partial_i P_i(U_i, U_i^*, A_i, 1\leqslant i\leqslant m)) + O(1)}$$

where $O(1)$ is uniform on the unitary group (but depends on P).

Proof. Let us first recall the following two elementary results of differential geometry:

(1) The map $\exp : \mathcal{M}_N \longrightarrow \mathcal{M}_N$ is differentiable and:

$$\mathrm{Diff}_M \exp .H := \lim_{\varepsilon\to 0} \varepsilon^{-1}\left(e^{M+\varepsilon H} - e^M\right) = \left(\sum_{k=0}^{+\infty} \frac{(\mathrm{Ad}_M)^k}{(k+1)!} H\right) e^M$$

where Ad_M is the operator defined by $\mathrm{Ad}_M H = MH - HM$.

(2) Note that if $P \in \mathbb{C}\langle (U_i, U_i^*, A_i)_{1\leqslant i\leqslant m}\rangle$ is considered as a function of the U_i's, then it is differentiable. Denoting by $A \otimes B \sharp C = ACB$ for $A, B, C \in \mathcal{M}_N$, we obtain similarly as in (7) that its differential with respect to the ith variable in the direction A, for A in $\mathcal{A}_N$, is given by

$$\mathrm{Diff}_i\, P.R = \partial_i P(\mathbf{U}) \sharp A.$$

As a consequence, if we fix A in $\mathcal{A}_N$ and $i \in \{1, \dots, m\}$, we have

$$\begin{aligned}
\mathrm{Diff}_i\, \Psi_j(\mathbf{U}).A &= 1_{i=j} U_j A + U_j\, \mathrm{Diff}_{\frac{\lambda}{N} P_j(U)} \exp .\left(\frac{\lambda}{N} \partial_i P_j \sharp A\right) \\
&= 1_{i=j} U_j A + \frac{\lambda}{N} \sum_{k=0}^{+\infty} U_j \frac{(\mathrm{Ad}_{\frac{\lambda}{N} P_j(U)})^k}{(k+1)!} (\partial_i P_j \sharp A) e^{\frac{\lambda}{N} P_j(U)} \\
&= 1_{i=j} U_j A + U_j \frac{\lambda}{N} \Phi_{ij}(\mathbf{U}) A
\end{aligned}$$

with $\Phi_{ij}(\mathbf{U})$ the linear map from $\mathcal{A}_N$ into $\mathcal{M}_N$ given by

$$\Phi_{ij}(\mathbf{U})A := \sum_{k=0}^{+\infty} \frac{(\mathrm{Ad}_{\frac{\lambda}{N} P_j(\mathbf{U})})^k}{(k+1)!} (\partial_i P_j \sharp A) e^{\frac{\lambda}{N} P_j(U)}.$$

We can factorize the term U_j to obtain

$$\mathrm{Diff}\, \Psi(\mathbf{U}) = U \circ \left(\mathrm{Id}_{\mathcal{A}_N^m} + \frac{\lambda}{N} \Phi(\mathbf{U})\right) \tag{11}$$

with $U \circ (M_1, \dots, M_m) = (U_1 M_1, \dots, U_m M_m)$ and Φ the linear operator from $\mathcal{A}_N^m$ to $\mathcal{M}_N^m$ whose blocks are the $\Phi_{ij}(\mathbf{U})$.

Since the operator norms of the A_i's and the U_i's are uniformly bounded in N, the operator norm of $\mathrm{Ad}_{\frac{\lambda}{N} P_j(\mathbf{U})}$ as an operator on $(\mathcal{M}_N, \|.\|_\infty)$ is also bounded. Thus, $\Phi_{ij}(\mathbf{U})$ is a uniformly bounded operator from $\mathcal{A}_N$ to $\mathcal{M}_N$, and the norm of $\frac{\lambda}{N}\Phi(\mathbf{U})$ is less than $1/2$ for N large enough. For those N, Ψ is a local diffeomorphism.

We can now compute the Jacobian of Ψ

$$\left|\det J_\Psi(\mathbf{U})\right| := \left|\det \mathrm{Diff}\, \Psi(\mathbf{U})\right| = |\det U| \left|\det\left(I + \frac{\lambda}{N}\Phi(\mathbf{U})\right)\right|.$$

Clearly, $|\det U| = 1$. Besides, the positivity of the eigenvalues of $I + \lambda\Phi(\mathbf{U})/N$ allows us to replace the determinant by the exponential of a trace:

$$\left|\det J_\Psi(\mathbf{U})\right| = \exp\left(\mathrm{Tr} \log\left(I + \frac{\lambda}{N}\Phi(\mathbf{U})\right)\right) = \exp\left(-\sum_{p\geqslant 1} \frac{(-\lambda)^p}{pN^p} \mathrm{Tr}\left(\Phi(\mathbf{U})^p\right)\right).$$

Note that since Φ is a bounded operator on $\mathcal{A}_N$, which is a space of dimension N^2, the pth term in the previous sum is at most of order N^{2-p}. We only look at the terms up to the order $O(N)$. A quick computation shows that if

$$\varphi: \begin{array}{rcl} \mathcal{A}_N & \to & \mathcal{A}_N, \\ X & \to & \sum_l A_l X B_l \end{array}$$

is considered as a real endomorphism, $\mathrm{Tr}\,\varphi = \sum_l \mathrm{Tr}\, A_l \,\mathrm{Tr}\, B_l$. Indeed, if we consider $E(kl)$, $1 \leqslant k, l \leqslant N$, the canonical basis of $\mathcal{A}_N$, given by

$$E(kl)_{rj} := \sqrt{-1}\, \frac{1_{r=k, j=l} + 1_{r=l, j=k}}{\sqrt{2(1 + 1_{k=l})}}$$

for $k \leqslant l$ and

$$E(kl)_{rj} := \frac{1_{r=k, j=l} - 1_{r=l, j=k}}{\sqrt{2}}$$

for $k \geqslant l$, $\mathrm{Tr}\,\varphi = \sum_{k,l} \mathrm{Tr}(E(kl)^* \varphi(E(kl))) = \sum_l \mathrm{Tr}\, A_l \,\mathrm{Tr}\, B_l$. This is sufficient to obtain the first term of the Jacobian:

$$\frac{\lambda}{N} \mathrm{Tr}\big(\Phi(\mathbf{U})\big) = \frac{\lambda}{N} \sum_i \mathrm{Tr}\big(\Phi_{ii}(\mathbf{U})\big) = \frac{\lambda}{N} \sum_i \mathrm{Tr} \otimes \mathrm{Tr}\big(\partial_i P_i\big(U_j, U_j^*, A_j\big)\big) + O(1)$$

with $O(1)$ uniformly bounded on $\mathcal{U}_N^m$ (since the operator norm of $\mathrm{Ad}_{\frac{\lambda}{N} P_j(\mathbf{U})}$ is uniformly small). $\square$

Before making the change of variables we show that Ψ is a bijection.

Lemma 2.2. *For N large enough, Ψ is a diffeomorphism of $\mathcal{U}_N^m$.*

Proof. First observe that since Ψ is a local diffeomorphism, its image is open in $\mathcal{U}_N^m$. Besides, since $\mathcal{U}_N^m$ is compact and Ψ is continuous, the image is compact and therefore closed. Thus by connectedness of $\mathcal{U}_N^m$, and since $\Psi(\mathcal{U}_N^m)$ is closed, open and non-empty, Ψ is surjective.

The only property we still need to prove is the injectivity of Ψ. If $\Psi(U) = \Psi(V)$, then, for all $j \in \{1, \dots, m\}$, we have

$$U_j^* V_j - I = e^{\frac{\lambda}{N} P_j(U)} e^{-\frac{\lambda}{N} P_j(V)} - I.$$

Thus, if N is sufficiently large so that $\frac{\lambda}{N} P_j(U)$ is in a domain where the function exp is 2-Lipschitz, we obtain

$$\begin{aligned}\|U_j - V_j\|_\infty &= \left\|U_j V_j^* - 1\right\|_\infty = \left\|e^{\frac{\lambda}{N} P_j(U)} e^{\frac{-\lambda}{N} P_j(V)} - 1\right\|_\infty \\ &= \left\|e^{\frac{\lambda}{N} P_j(U)} - e^{\frac{\lambda}{N} P_j(V)}\right\|_\infty \leqslant \frac{2|\lambda|}{N} \left\|P_j(U) - P_j(V)\right\|_\infty\end{aligned}$$

with $\|.\|_\infty$ the operator norm. Since $(P_j, 1 \leqslant j \leqslant m)$ are uniformly Lipschitz on $\mathcal{U}_N^m$, we conclude that $\sum_{j=1}^m \|U_j - V_j\|_\infty$ vanishes for sufficiently large N. □

We can now prove Theorem 2.1.

Proof of Theorem 2.1. Let us define, for $P = (P_1, \dots, P_m)$ with P_i anti-self-adjoint,

$$Y^N(P) = \sum_i \left(\frac{1}{N} \operatorname{Tr}(D_i V P_i) + \frac{1}{N} \operatorname{Tr} \otimes \frac{1}{N} \operatorname{Tr}(\partial_i P_i) \right).$$

One can check easily that adjunction and derivation anticommute: for all polynomial P, $D_i(P^*) = -(D_i P)^*$ and similarly if we define the adjunction on tensors by the linear map such that $(P \otimes Q)^* = Q^* \otimes P^*$ then $\partial_i(P^*) = -(\partial_i P)^*$. From there we deduce:

$$\begin{aligned}\operatorname{Tr}(D_i V P) &= \overline{\operatorname{Tr}\big((D_i V P)^*\big)} \\ &= \overline{\operatorname{Tr}\big(P^*(D_i V)^*\big)} = \overline{\operatorname{Tr}\big(P D_i\big(V^*\big)\big)} = \overline{\operatorname{Tr}(D_i V P)}\end{aligned}$$

so that $\operatorname{Tr}(D_i V P)$ is real. A similar computation shows that $\operatorname{Tr} \otimes \operatorname{Tr}(\partial_i P_i)$ is always real for an anti-self-adjoint polynomial.

We can expand $\operatorname{Tr} V(\Psi(\mathbf{U})_i, \Psi(\mathbf{U})_i^*, A_i, 1 \leqslant i \leqslant m)$ as

$$\begin{aligned}&\operatorname{Tr}\big(V\big(\Psi(\mathbf{U})_i, \Psi(\mathbf{U})_i^*, A_i, 1 \leqslant i \leqslant m\big)\big) - \operatorname{Tr}\big(V\big(U_i, U_i^*, A_i, 1 \leqslant i \leqslant m\big)\big) \\ &\quad = \frac{\lambda}{N} \sum_j \operatorname{Tr}\big(D_j V P_j\big(U_i, U_i^*, A_i, 1 \leqslant i \leqslant m\big)\big) + O\big(N^{-1}\big) \qquad (12)\end{aligned}$$

and perform the change of variables $\mathbf{U} \to \Psi(\mathbf{U})$ in $I_N(V, A_i^N)$;

$$
\begin{aligned}
I_N\big(V, A_i^N\big) &:= \int e^{N\,\mathrm{Tr}(V(U_i, U_i^*, A_i, 1\leqslant i\leqslant m))}\, dU_1 \cdots dU_m \\
&= \int e^{N\,\mathrm{Tr}(V(\Psi(\mathbf{U})_i, \Psi(\mathbf{U})_i^*, A_i, 1\leqslant i\leqslant m))}\big|\det J_\Psi(\mathbf{U})\big|\, dU_1 \cdots dU_m \\
&= \int e^{NY^N(P)+0(1)} e^{N\,\mathrm{Tr}(V(U_i, U_i^*, A_i, 1\leqslant i\leqslant m))}\, dU_1 \cdots dU_m
\end{aligned}
$$

where we used (12) and Lemma 2.1. $O(1)$ is of order one independently of N and uniformly on the unitary matrices $(U_1, \dots, U_m)$. Thus we have proved that

$$\int e^{NY^N(P)}\, d\mu_V^N(\mathbf{U}) = O(1).$$

Borel–Cantelli's lemma thus insures that

$$\limsup_{N\to\infty} Y^N(P) \leqslant 0 \quad \text{a.s.}$$

and the converse inequality holds by changing P into $-P$ since Y^N is linear in P. This proves the first statement of Theorem 2.1 for any anti-self-adjoint polynomials $(P_i)_{1\leqslant i\leqslant m}$. Multiplying these polynomials by $\sqrt{-1}$ gives it for self-adjoint polynomials and then for all polynomials by linearity. The last result is simply based on the compactness of $\mathcal{M}$ and the fact that any limit point must then satisfy the same asymptotic equations as $\hat{\mu}^N$. □

Another consequence of this convergence is the existence of solutions to (10) for any self-adjoint potential V (since any limit point of $\hat{\mu}^N$ in the compact metric space $\mathcal{M}$ will satisfy it) a fact already proved in [6]. Moreover, since these solutions are limit points of $\hat{\mu}^N$, they belong to $\mathcal{M}$ and in particular $|\mu(q)| \leqslant 1$ for any monomial q.

2.2. *Moments of $\hat{\mu}^N$*

In this section, we denote by E the expectation with respect to the Haar measure on the unitary group. The goal of this section is to show (see Proposition 2.1) that cumulants also satisfy a formal version of the Schwinger–Dyson equation. We start with the following lemma:

Lemma 2.3. *For all i all N, all monomials $q_1, \dots, q_n$ and all $\mathbf{k} = (k_1, \dots, k_n)$ in $\mathbb{N}^n$, we have*

$$
\begin{aligned}
&N^2 E\big(\hat{\mu}^N \otimes \hat{\mu}^N(\partial_i P) \cdot \big(\hat{\mu}^N(q_1)\big)^{k_1} \cdots \big(\hat{\mu}^N(q_n)\big)^{k_n}\big) \\
&\quad + \sum_j k_j E\big(\big(\hat{\mu}^N(q_1)\big)^{k_1} \cdots \big(\hat{\mu}^N(q_j)\big)^{k_j-1} \cdots \big(\hat{\mu}^N(q_n)\big)^{k_n} \hat{\mu}^N(D_i q_j \cdot P)\big) = 0.
\end{aligned}
$$

Proof. Following Lemma 2.1, we write down the change of variables

$$\Psi_i : \mathbf{U} \to \big(U_1, \dots, U_{i-1}, U_i e^{\lambda P_i(\mathbf{U})}, U_{i+1}, \dots, U_m\big)$$

in the integral $\int((\hat{\mu}^N q_1)^{k_1}\cdots(\hat{\mu}^N q_n)^{k_n})\,dU_1\cdots dU_m$, where the integration is performed with respect to the Haar measure. The Jacobian J_i of Ψ_i satisfies

$$\left|\det J_i(\mathbf{U})\right| = 1 + \frac{\lambda}{N}\operatorname{Tr}\otimes\operatorname{Tr}(\partial_i P) + o(\lambda)$$

and we have the expansion

$$\begin{aligned}\operatorname{Tr}\bigl(q_j\bigl(\Psi(\mathbf{U})_i, \Psi(\mathbf{U})_i^*, A_i, 1\leqslant i\leqslant m\bigr)\bigr) &= \operatorname{Tr}\bigl(q_j\bigl(U_i, U_i^*, A_i, 1\leqslant i\leqslant m\bigr)\bigr)\\ &\quad + \lambda\operatorname{Tr}\bigl(D_i q_j\cdot P\bigl(U_i, U_i^*, A_i, 1\leqslant i\leqslant m\bigr)\bigr) + \lambda^2 o(\lambda)\end{aligned}$$

where the $o(\lambda)$'s are for a given P uniform bounds in N. The first order of the Taylor expansion of this change of variables around $\lambda = 0$ proves the claim. □

Proposition 2.1. *For all i, we have the following identity of power series:*

$$E\bigl[\hat{\mu}^N\otimes\hat{\mu}^N(\partial_i P)e^{N^2\hat{\mu}^N(V_{\mathbf{t}})}\bigr] + E\bigl[\hat{\mu}^N(D_i V_{\mathbf{t}}\cdot P)e^{N^2\hat{\mu}^N(V_{\mathbf{t}})}\bigr] = 0.$$

Proof. For all $\mathbf{k}$, the left-hand side of the equality of Lemma 2.3 multiplied by $N^{2|\mathbf{k}|-2}/\mathbf{k}!$ is the coefficient of $\mathbf{t}^{\mathbf{k}}$ in the series

$$E\bigl[\hat{\mu}^N\otimes\hat{\mu}^N(\partial_i P)e^{N^2\hat{\mu}^N(V_{\mathbf{t}})}\bigr] + E\bigl[\hat{\mu}^N(D_i V_{\mathbf{t}}\cdot P)e^{N^2\hat{\mu}^N(V_{\mathbf{t}})}\bigr].$$

Therefore this series vanishes as all its coefficients equal zero by Lemma 2.3. □

Finally we study the large N limit μ^f of these tracial power states (the exponent f stands for "formal").

Theorem 2.2. *Let $V_{\mathbf{t}}$ be the polynomial $\sum_{j=1}^n t_j q_j$. The sequence of tracial power state $\bar{\mu}_{V_{\mathbf{t}}}^N$ converges when N goes to infinity to some limit μ^f in the sense that, for all P each coefficient of the power series $\bar{\mu}_{V_{\mathbf{t}}}^N(P)$ converges towards $\mu^f(P)$ in $\mathbb{C}[[N^{-1}]]$. Besides, μ^f satisfies the following family of equations in $\mathbb{C}[[N^{-1}]]$:*

$$\mu^f\otimes\mu^f(\partial_i P) + \mu^f(D_i V_{\mathbf{t}}\cdot P) = 0,$$

for all i and for all P.

Proof. First, we prove the existence of a limit. By the first item of Proposition 1.1, we can express $\bar{\mu}_{V_{\mathbf{t}}}^N(P)$ as a sum over cumulants,

$$\bar{\mu}_{V_{\mathbf{t}}}^N(P) = \sum_{\mathbf{k}\in\mathbb{N}^n}\mathbf{t}^{\mathbf{k}}C_{1,\mathbf{k}}\left(\frac{1}{N}\operatorname{Tr}P, N\operatorname{Tr}q_1,\ldots,N\operatorname{Tr}q_n\right)/\mathbf{k}!.$$

The limit in N, of the $C_{1,\mathbf{k}}(\frac{1}{N}\operatorname{Tr}P, N\operatorname{Tr}q_1,\ldots,N\operatorname{Tr}q_n)$ was proved to exist in [8] so that μ^f is well defined.

Item (2) from Proposition 1.1 implies that:

$$\frac{E(\frac{1}{N}\operatorname{Tr}P_1\frac{1}{N}\operatorname{Tr}P_2e^{N\operatorname{Tr}V})}{E(e^{N\operatorname{Tr}V})}-\frac{E(\frac{1}{N}\operatorname{Tr}P_1e^{N\operatorname{Tr}V})}{E(e^{N\operatorname{Tr}V})}\frac{E(\frac{1}{N}\operatorname{Tr}P_2e^{N\operatorname{Tr}V})}{E(e^{N\operatorname{Tr}V})}$$
$$=\sum_{k\geqslant 0}\frac{\mathbf{t}^{\mathbf{k}}}{\mathbf{k}!}C_{1,1,\mathbf{k}}\left(\frac{1}{N}\operatorname{Tr}P_1,\frac{1}{N}\operatorname{Tr}P_2,N\operatorname{Tr}q_1,\dots,N\operatorname{Tr}q_n\right).$$

Now, it follows from [8] that each coefficient of the series in the right-hand side decays like N^{-2} so that the coefficientwise limit is zero.

The proof of the theorem follows from this observation and from Proposition 2.1. $\square$

3. Study of the Schwinger–Dyson equation

We have shown that the limit points of the matrix model satisfy the Schwinger–Dyson equation (10). The aim of this section is to study this equation and show that it has a unique solution under appropriate boundedness assumptions.

Definition 3.1. Let τ be an element of $\mathcal{M}|_{(A_i)_{1\leqslant i\leqslant m}}$. A tracial state $\mu\in\mathcal{M}$ is said to satisfy the Schwinger–Dyson equation $\mathbf{SD}[V,\tau]$ if and only if for all $P\in\mathbb{C}\langle (A_i)_{1\leqslant i\leqslant m}\rangle$,

$$\mu(P)=\tau(P)$$

and for all $P\in\mathbb{C}\langle (U_i,U_i^*,A_i)_{1\leqslant i\leqslant m}\rangle$, all $i\in\{1,\dots,m\}$,

$$\mu\otimes\mu(\partial_i P)+\mu(D_i VP)=0.$$

Let V be in $\mathbb{C}\langle (U_i,U_i^*,A_i)_{1\leqslant i\leqslant m}\rangle$. The polynomial V can be written as a sum

$$V=\sum_{i=1}^{n}t_iq_i\big(U_j,U_j^*,A_j,1\leqslant j\leqslant m\big)$$

with monomial functions q_i and complex numbers t_i. We let D be the maximal degree of the monomials q_i.

Here we prove that if the parameters $(t_i,1\leqslant i\leqslant m)$ are small enough this equation has a unique solution μ.

Theorem 3.1. *Let D be an integer and τ a tracial state in $\mathcal{M}|_{(A_i)_{1\leqslant i\leqslant m}}$. There exists $\varepsilon=\varepsilon(D,m)>0$ such that if $|t_i|\leqslant\varepsilon$, there exists at most one solution μ to* $\mathbf{SD}[V,\tau]$.

From this and Theorem 2.1 we deduce the following

Corollary 3.1. *Assume that V is self-adjoint. Let D be an integer and τ a tracial state in $\mathcal{M}|_{(A_i)_{1\leqslant i\leqslant m}}$. There exists $\varepsilon=\varepsilon(D,m)>0$ such that if $|t_i|\leqslant\varepsilon$, $\hat\mu^N$ converges almost surely to the unique solution μ of the Schwinger–Dyson equation. Moreover, $\bar\mu_V^N=\mu_V^N(\hat\mu^N)$ converges as well to this solution as N goes to infinity.*

This result is obvious since Theorems 2.1 and 3.1 show that $\hat{\mu}^N$ has a unique limit point, and thus converges almost surely. The convergence of $\bar{\mu}_V^N$ is then a direct consequence of bounded convergence theorem since $\hat{\mu}^N \in \mathcal{M}$.

We would like to draw the attention of the reader on the fact that Theorem 2.1 and Corollary 3.1 do not use the assumption that the matrices $(A_i^N, 1 \leqslant i \leqslant m)$ are deterministic, but only that they are bounded and have a converging joint distribution. Therefore these two results extend to the case where these matrices are random, independent of the $(U_i, 1 \leqslant i \leqslant m)$, and satisfy the above two conditions almost surely. This observation implies that our result can also encompass the case of the truncated GUE or other classical bounded matrix models.

We are now ready to prove Theorem 3.1:

Proof of Theorem 3.1. Let μ be a solution to $\mathbf{SD}[V,\tau]$. Note that if q is a monomial, then either q does not depend on $(U_j, U_j^*, 1 \leqslant j \leqslant m)$ and $\mu(q) = \tau(q)$ defines μ on this polynomial or q can be written as $q = q_1 U_i^a q_2$ for some $i \in \{1, \ldots, m\}$, $a \in \{-1, +1\}$ and monomials q_1, q_2. Then, by the traciality assumption, $\mu(q) = \mu(q_2 q_1 U_i^a) = \mu(U_i^a q')$ with $q' = q_2 q_1$. Without loss of generality we assume that the last letter of q' is not U_i^{-a}. We next use $\mathbf{SD}[V,\tau]$ to compute $\mu(U_i^a q)$ for some monomial q. We assume first that $a = -1$. Then, by (4),

$$\partial_i \big(U_i^* q\big) = -1 \otimes \big(U_i^* q\big) + U_i^* \otimes 1 \times \partial_i q.$$

Taking the expectation, we thus find by (5), since $\mu(1) = 1$, that

$$\begin{aligned}
\mu\big(U_i^* q\big) &= \mu \otimes \mu\big(U_i^* \otimes 1 \times \partial_i q\big) + \mu(D_i V q) \\
&= \sum_{q = q_1 U_i q_2} \mu(q_1)\mu(q_2) - \sum_{q = q_1 U_i^* q_2} \mu\big(U_i^* q_1\big)\mu\big(U_i^* q_2\big) \\
&\quad + \sum_j t_{ij} \mu(q_{ij} q)
\end{aligned} \tag{13}$$

where $D_i V = \sum_j t_{ij} q_{ij}$ is a decomposition of $D_i V$ in monomials q_{ij}. Note that the sum runs at most on Dn terms and that all the t_{ij} are bounded by $\max |t_{ij}|$. A similar formula is found when $a = +1$ by differentiating qU_i (or by using $\overline{\mu(qU_i))} = \mu((qU_i)^*) = \mu(U_i^* q^*)$.

We next show that (13) and its equivalent for $a = +1$ characterize uniquely $\mu \in \mathcal{M}$ when the t_{ij} are small enough. It will be crucial here that $\mu(q)$ is bounded independently of the t_i's (here by the constant 1).

Let $\mu, \mu' \in \mathcal{M}$ be two solutions to $\mathbf{SD}[V,\tau]$ and set

$$\Delta(\ell) = \sup_{\deg(q) \leqslant \ell} \big|\mu(q) - \mu'(q)\big|$$

where the supremum holds over monomials of $\mathbb{C}\langle (U_i, U_i^*, A_i)_{1 \leqslant i \leqslant m}\rangle$ with total degree in the U_j and U_j^* less than ℓ. Namely, if the monomial (or word) q contains a_j^+ times U_j and a_j^- times U_j^*, we assume $\sum_{j=1}^m (a_j^+ + a_j^-) \leqslant \ell$. Note that by traciality of μ,

$$\Delta(\ell) = \max_{\substack{1 \leqslant i \leqslant m \\ a \in \{+1,-1\}}} \sup_{\deg_q \leqslant \ell - 1} \big|\mu\big(U_i^a q\big) - \mu'\big(U_i^a q\big)\big| \tag{14}$$

and that by (13), we find that, for q with degree less than $\ell - 1$,

$$\begin{aligned}\bigl|\mu\bigl(U_i^* q\bigr) - \mu'\bigl(U_i^* q\bigr)\bigr| \leqslant{}& \sum_{q=q_1 U_i q_2} \bigl|(\mu-\mu')(q_1)\bigr| + \sum_{q=q_1 U_i q_2} \bigl|(\mu-\mu')(q_2)\bigr| \\ &+ \sum_{q=q_1 U_i^* q_2} \bigl|(\mu-\mu')\bigl(U_i^* q_1\bigr)\bigr| + \sum_{q=q_1 U_i^* q_2} \bigl|(\mu-\mu')\bigl(U_i^* q_2\bigr)\bigr| \\ &+ \sum_j t_{ij} \bigl|(\mu-\mu')(q_{ij} q)\bigr|.\end{aligned}$$

A similar formula holds for $|\mu(U_i q) - \mu'(U_i q)|$ by conjugation, therefore

$$\Delta(\ell) \leqslant 2\sum_{p=1}^{\ell-2} \Delta(p) + 2\sum_{p=1}^{\ell-1} \Delta(p) + nD\varepsilon\Delta(\ell+D-1)$$

where we used that $\deg(q_1) \in \{0,\dots,\ell-2\}$, $\deg(q_2) \in \{0,\dots,\ell-2\}$ (but $\Delta(0)=0$) and $\deg(q_{ij}) \leqslant D$ and assumed $|t_i| \leqslant \varepsilon$. Hence, we have proved that

$$\Delta(\ell) \leqslant 4\sum_{p=1}^{\ell-1} \Delta(p) + nD\varepsilon\Delta(\ell+D).$$

Multiplying these inequalities by γ^ℓ we get, since $H(\gamma) := \sum_{\ell\geqslant 1} \gamma^\ell \Delta(\ell)$ is finite for $\gamma < 1$,

$$H(\gamma) \leqslant \frac{\gamma}{1-\gamma} H(\gamma) + \frac{nD\varepsilon}{\gamma^D} H(\gamma)$$

resulting with $H(\gamma) = 0$ for γ so that $1 > \frac{\gamma}{1-\gamma} + \frac{nD\varepsilon}{\gamma^D}$. Such a $\gamma > 0$ exists when ε is small enough. This proves the uniqueness. □

As a corollary, we characterize asymptotic freeness by the Schwinger–Dyson equation, a result which was already obtained in [32, Proposition 5.17].

Corollary 3.2. *A tracial state μ satisfies* **SD**$[0,\tau]$ *if and only if, under μ, the algebras generated by $(A_i, 1 \leqslant i \leqslant m)$ and $(U_i, U_i^*, 1 \leqslant i \leqslant m)$ are free and the U_i's is a family of free variables such that*

$$\mu\bigl(U_i^a\bigr) = 0 \quad \forall a \in \mathbb{Z}\backslash\{0\}.$$

Proof. By the previous theorem, it is enough to verify that the law μ of free variables $(A_i, U_i, U_i^*)_{1\leqslant i\leqslant m}$ satisfies **SD**$[0,\tau]$. So take $P = U_{i_1}^{a_1} B_1 \cdots U_{i_p}^{a_p} B_p$ with some B_k's in the algebra generated by $(A_i, 1 \leqslant i \leqslant m)$. We wish to show that for all $i \in \{1,\dots,m\}$,

$$\mu \otimes \mu(\partial_i P) = 0.$$

Note that by linearity, it is enough to prove this equality when $\mu(B_j)=0$ for all j. Now, by definition, we have

$$\partial_i P = \sum_{k:i_k=i,a_k>0}\sum_{l=1}^{a_k} U_{i_1}^{a_1}B_1\cdots B_{k-1}U_i^l\otimes U_i^{a_k-l}B_k\cdots U_{i_p}^{a_p}B_p$$
$$-\sum_{k:i_k=i,a_k<0}\sum_{l=0}^{a_k-1} U_{i_1}^{a_1}B_1\cdots B_{k-1}U_i^{-l}\otimes U_i^{a_k+l}B_k\cdots U_{i_p}^{a_p}B_p.$$

Taking the expectation on both sides, since $\mu(U_j^i)=0$ and $\mu(B_j)=0$ for all $i\neq 0$ and j, we see that freeness implies that the right-hand side is null (recall here that in the definition of freeness, two consecutive elements have to be in free algebras but the first and the last element can be in the same algebra). Thus, $\mu\otimes\mu(\partial_i P)=0$ which proves the claim. □

4. Formal solution and analyticity

We have shown in Theorem 2.2 that the limit points of the formal model also satisfy an equation similar to the Schwinger–Dyson equation. The only difference with Definition 3.1 is that the Schwinger–Dyson equation is on the space of tracial states while for the formal model, the equation holds on the space of tracial power states. In order to prove that the formal model matches the matrix model we need to study this formal equation and show that the series have a positive radius of convergence, hence providing a solution to $\mathbf{SD}[V,\tau]$ as defined in Definition 3.1.

Definition 4.1. Let $V_{\mathbf{t}}=\sum_i t_i q_i$ be a polynomial. Let τ be a tracial power state in $\mathcal{M}|_{(A_i)_{1\leqslant i\leqslant m}}$. A tracial power state $\mu\in\mathcal{M}$ is said to satisfy the Schwinger–Dyson equation $\mathbf{SD}^f[V_{\mathbf{t}},\tau]$ if and only if for all $P\in\mathbb{C}\langle(A_i)_{1\leqslant i\leqslant m}\rangle$,

$$\mu(P)=\tau(P)$$

and for all $P\in\mathbb{C}\langle(U_i,U_i^*,A_i)_{1\leqslant i\leqslant m}\rangle$, all $i\in\{1,\dots,m\}$, the equation

$$\mu\otimes\mu(\partial_i P)+\mu(D_iV_{\mathbf{t}}P)=0$$

holds in $\mathbb{C}[[\mathbf{t}]]$.

We already know, due to Theorem 2.2, that there exists a solution to this equation. We now prove that this solution is unique.

Theorem 4.1. *There exists a unique tracial power state $\mathbf{t}\to\mu_{\mathbf{t}}$ which satisfies the Schwinger–Dyson equation $\mathbf{SD}^f[V_{\mathbf{t}},\tau]$.*

Proof. Let $\mu_{\mathbf{t}}$ be a tracial power state solution of $\mathbf{SD}^f[V_{\mathbf{t}},\tau]$. There exists a family $\mu^{\mathbf{k}},\mathbf{k}=(k_1,\dots,k_n)\in\mathbb{N}^n$ in the algebraic dual of $\mathbb{C}\langle(U_i,U_i^*,A_i)_{1\leqslant i\leqslant m}\rangle$ such that for all P,

$$\mu_{\mathbf{t}}(P)=\sum_{\mathbf{k}\in\mathbb{N}^n}\prod_{i=1}^n\frac{t_i^{k_i}}{k_i!}\mu^{\mathbf{k}}(P).$$

We will now show that the $\mu_{\mathbf{k}}$ are uniquely inductively defined by the relation given by $\mathbf{SD}^f[V_{\mathbf{t}},\tau]$. Let us define e_j the canonical basis of $\mathbb{R}^n$. We get the following equalities, for all $\mathbf{k}$,

(1) If P is in $\mathbb{C}\langle (A_i)_{1\leqslant i\leqslant m}\rangle$, $\mu^{\mathbf{k}}(P)=\tau(P)1_{\mathbf{k}=0}$.
(2) If $P=RU_iS$ with S in $\mathbb{C}\langle (A_i)_{1\leqslant i\leqslant m}\rangle$, $\mu^{\mathbf{k}}(P)=\mu^{\mathbf{k}}(SRU_i)$.
(3) If $P=RU_i^*S$ with R in $\mathbb{C}\langle (A_i)_{1\leqslant i\leqslant m}\rangle$ and S does not contain any U_j (but may contain the U_j^*), $\mu^{\mathbf{k}}(P)=\mu^{\mathbf{k}}(U_i^*SR)$.
(4) If q does not contain any U_j,

$$\mu^{\mathbf{k}}\big(U_i^*q\big)=-\sum_{q=q_1U_i^*q_2}\binom{\mathbf{k}}{\mathbf{k}'}\sum_{\mathbf{k}'+\mathbf{k}''=\mathbf{k}}\mu^{\mathbf{k}'}\big(U_i^*q_1\big)\mu^{\mathbf{k}''}\big(U_i^*q_2\big)+\sum_j k_j\mu^{\mathbf{k}-e_j}\big(U_i^*qD_iq_j\big).$$

(5) And for all q,

$$\mu^{\mathbf{k}}(qU_i)=-\sum_{q=q_1U_iq_2}\sum_{\mathbf{k}'+\mathbf{k}''=\mathbf{k}}\binom{\mathbf{k}}{\mathbf{k}'}\mu^{\mathbf{k}'}(q_1U_i)\mu^{\mathbf{k}''}(q_2U_i)+\sum_{q=q_1U_i^*q_2}\sum_{\mathbf{k}'+\mathbf{k}''=\mathbf{k}}\binom{\mathbf{k}}{\mathbf{k}'}\mu^{\mathbf{k}'}(q_1)\mu^{\mathbf{k}''}(q_2)-\sum_j k_j\mu^{\mathbf{k}-e_j}(D_iq_jqU_i).$$

This allows to compute uniquely any $\mu^{\mathbf{k}}(P)$. The first relation takes care of the non-random case, the relations (2) and (3) use the traciality to place a variable U in a convenient place. Finally relations (4) and (5) allow to compute $\mu^{\mathbf{k}}(P)$ as a function which depends on the $\mu^{\mathbf{k}'}(Q)$ with $\deg Q<\deg P$ and $\mathbf{k}'\leqslant\mathbf{k}$ (first terms) or on the $\mu^{\mathbf{k}'}(Q)$ with $\mathbf{k}'<\mathbf{k}$ (last term). This is a well founded induction. Thus the $\mu^{\mathbf{k}}$ are uniquely defined. □

We next show that this solution is not only a tracial power state but that if we evaluate it with some t_i's in $\mathbb{C}$ we obtain a family of solutions $\mu_{\mathbf{t}}$ of the non-formal equation $\mathbf{SD}[V_{\mathbf{t}},\tau]$, which depends analytically on the parameters $(t_i)_{1\leqslant i\leqslant n}$.

Theorem 4.2. *There exists $\varepsilon>0$ such that for $\mathbf{t}\in\mathbb{C}^n$, $\max_{1\leqslant i\leqslant n}|t_i|\leqslant\varepsilon$, the tracial power state $\mu_{\mathbf{t}}$ solution of $\mathbf{SD}^f[V_{\mathbf{t}},\tau]$ is a convergent series. For all polynomials P, $\mathbf{t}\in B(0,\varepsilon)=\{t\in\mathbb{C}^n\colon \max_{1\leqslant i\leqslant n}|t_i|\leqslant\varepsilon\}\longrightarrow\mu_{\mathbf{t}}(P)$ is analytic, and there exists a family ($\mu^{\mathbf{k}},\mathbf{k}=(k_1,\dots,k_n)\in\mathbb{N}^n$) in the algebraic dual of $\mathbb{C}\langle (U_i,U_i^*,A_i)_{1\leqslant i\leqslant m}\rangle$ such that for all P,*

$$\mu_{\mathbf{t}}(P)=\sum_{\mathbf{k}\in\mathbb{N}^n}\prod_{i=1}^n\frac{t_i^{k_i}}{k_i!}\mu^{\mathbf{k}}(P)$$

converges absolutely for $\max_{1\leqslant i\leqslant n}|t_i|\leqslant\varepsilon$.

An immediate consequence of this result is that the tracial power state solution of $\mathbf{SD}^f[V_{\mathbf{t}},\tau]$ is also after taking the t_i's in a small parameters region an actual solution of $\mathbf{SD}[V_{\mathbf{t}},\tau]$, and therefore by Theorem 3.1, equals the real solution. This will be a key to prove Theorem 0.1 (see Section 7).

Corollary 4.1. *The tracial power state solution of the Schwinger–Dyson equation* $\mathbf{SD}^f[V_{\mathbf{t}}, \tau]$ *is a convergent series for small* $\mathbf{t}$. *In addition it matches the real solution of* $\mathbf{SD}[V_{\mathbf{t}}, \tau]$ *which thus depends analytically in the parameters* $\mathbf{t}$ *of the potential in a neighborhood of the origin.*

Let us now prove Theorem 4.2.

Proof of Theorem 4.2. According to the proof of Theorem 4.1 the $\mu^{\mathbf{k}}$ are uniquely defined by the family of relations (1)–(5). We only need to control the growth of the coefficients $\mu^{\mathbf{k}}(P)$ to show that $\mu_{\mathbf{t}}(P)$ is indeed convergent for small enough parameters.

To bound these quantities, we use the Catalan numbers

$$C_0 = 1, \qquad C_{k+1} = \sum_{0 \leqslant p \leqslant k} C_p C_{k-p}$$

and the fact that they satisfy the exponential growth inequality $C_{k+1} \leqslant 4C_k$. We denote $C_{\mathbf{k}} := \prod_i C_{k_i}$ and for $A > 0$, $D_k^A := A^{k-1} C_{k-1}$ for $k \geqslant 1$, $D_0^A := 0$. The two key properties of this sequence is first that it is sub-geometric ($D_{k+1}^A \leqslant 4AD_k^A$) and secondly it satisfies $D_k^A = A \sum_{0<p<k} D_p^A D_{k-p}^A$. Now our induction hypothesis is that there exists $A, B > 0$ such that for all $\mathbf{k}$, for all monomial P of degree p,

$$\frac{|\mu^{\mathbf{k}}(P)|}{\mathbf{k}!} \leqslant C_{\mathbf{k}} B^{\mathbf{k}} D_p^A. \tag{15}$$

We prove this bound by induction, and the relations (1)–(5) which define the $\mu^{\mathbf{k}}$. For $\mathbf{k} = (0, \dots, 0)$ this bound is satisfied since $D_p^A \geqslant 1$. We will check the induction for a polynomial of the form qU_i since the case qU_i^* is obtain by taking the complex conjugate.

$$\begin{aligned}\frac{|\mu^{\mathbf{k}}(qU_i)|}{\mathbf{k}!} \leqslant{}& \sum_{\substack{q=q_1U_iq_2\\ \mathbf{k}'+\mathbf{k}''=\mathbf{k}}} \frac{|\mu^{\mathbf{k}'}(q_1U_i)|}{\mathbf{k}'!}\frac{|\mu^{\mathbf{k}''}(q_2U_i)|}{\mathbf{k}''!}\\ &+ \sum_{\substack{q=q_1U_i^*q_2\\ \mathbf{k}'+\mathbf{k}''=\mathbf{k}}} \frac{|\mu^{\mathbf{k}'}(q_1)|}{\mathbf{k}'!}\frac{|\mu^{\mathbf{k}''}(q_2)|}{\mathbf{k}''!} + \sum_{k_j \neq 0} \frac{|\mu^{\mathbf{k}-1_j}(D_i^A q_j q)|}{(\mathbf{k}-1_j)!}.\end{aligned}$$

Now we use the induction hypothesis. If q is of degree $p-1$,

$$\begin{aligned}\frac{|\mu^{\mathbf{k}}(qU_i)|}{\mathbf{k}! C_{\mathbf{k}} B^{\mathbf{k}} D_p^A} &\leqslant 2 \sum_{\substack{0<r<p\\ \mathbf{k}'+\mathbf{k}''=\mathbf{k}}} \frac{C_{\mathbf{k}'} B^{\mathbf{k}'} D_r^A C_{\mathbf{k}''} B^{\mathbf{k}''} D_{p-r}^A}{C_{\mathbf{k}} B^{\mathbf{k}} D_p^A} + D \sum_j \frac{C_{\mathbf{k}-1_j} B^{\mathbf{k}-1} D_{p+D}^A}{C_{\mathbf{k}} B^{\mathbf{k}} D_p^A}\\ &\leqslant 2 \prod_i \frac{C_{k_i+1}}{C_{k_i}} \frac{1}{A} + nD \frac{(4A)^D}{B}.\end{aligned}$$

The point is that we can choose $A, B > 0$ such that this last quantity is less than 1. For example take $A > 4^{n+1}$ and then $B > 2nD(4A)^D$.

Thus, for $\|t\| := \max_i |t_i| < 1/4B$, for all P in $\mathbb{C}\langle (U_i, U_i^*, A_i)_{1 \leqslant i \leqslant m}\rangle$, the series $\sum_{\mathbf{k}} \prod_i \frac{t_i^{k_i}}{k_i!} \mu^{\mathbf{k}}(P)$ is absolutely convergent. □

5. Combinatorics

The purpose of this section is to provide a graphical approach to the solution of the Schwinger–Dyson equation, and therefore to the computation of unitary matrix integrals and free entropy (see Sections 6, 7 and 8). Actually, the proof of Theorem 4.1 gives a recursive way of computing a tracial power state solution to the formal Schwinger–Dyson equation, and in turn, numerical solutions with arbitrary precision.

Before giving a detailed description of our combinatorial model, we start with an overview. We need the notions of a **star**, which is a pictorial encoding of a monomial of $\mathbb{C}\langle (U_i, U_i^*, A_i)_{1 \leqslant i \leqslant m}\rangle$, of **root star**, which is a distinguished star, and of a **map**, which is a specific planar decoration over a set of stars and one root star.

The goal of this section is to show that the limits of integrals on the space of unitary matrices are generating function of the number of some maps as described above. However we are not interested in all maps, but rather in some that arise from an admissible construction, which leads us to the concept of **admissible maps**. Last, we need the notion of **weight** of a map, and our result will be in terms of sum over admissible maps of weights.

For the sake of clarity, although our usual playground is the algebra $\mathbb{C}\langle (U_i, U_i^*, A_i)_{1 \leqslant i \leqslant m}\rangle$ and our definitions work in full generality, we restrict ourselves in the examples to the case of one single unitary matrix U and two variables $A_1 =: A$ and $A_2 =: B$. We first start with the definition of stars and root stars, in the spirit of [16,17].

Definition 5.1.

(1) A **star** is a circle endowed with the clockwise orientation, decorated with elements such as colored incoming or outgoing arrows, and colored diamonds. One of the element is marked.

(2) To each letter X in the alphabet $(A_i, U_i, U_i^*)_{1 \leqslant i \leqslant m}$, we associate bijectively an **element** as follows: a diamond of color i if $X = A_i$ and a ring of color i if $X = U_i$ or U_i^*. In the case of U_i (resp. U_i^*) we attach before the ring an outgoing arrow of color i (resp. we attach after the ring an incoming arrow of color i) outside of the circle.

(3) To a monomial $q \in \mathbb{C}\langle (U_i, U_i^*, A_i)_{1 \leqslant i \leqslant m}\rangle$, we associate in a canonical way a **star of type** q by drawing on the clockwise oriented circle the elements associated to the successive letters of q, while the element corresponding to the first letter of q is marked (or distinguished).

(4) A **root star of type** q is a star with a distinguished first element. Although the maps are on the sphere, in the graphical representation of this section we will draw them on the plane and, to highlight the role of the root star we will draw it in this section such that it contains all the other stars. Thus on drawings, the root star will be the border of the outer face. It contains the point of the sphere which was send to infinity in order to make a planar representation. Therefore, its orientation is counterclockwise. Besides, on a root star we will distinguish a root element. If q contains no U_i nor U_i^*, there are no root element. If q contains a U_i, the ring associated to the last $(U_i, 1 \leqslant i \leqslant m)$ is the root element. If q contains no U_i but some U_i^*, the ring associated to the first $(U_i^*, 1 \leqslant i \leqslant m)$ (the order being determined by the marked element determined by q and the orientation) is called the root element.

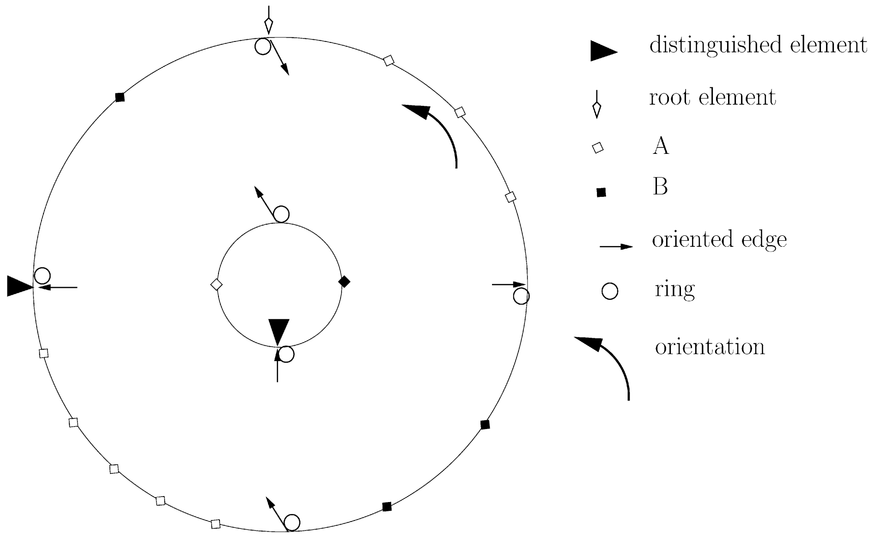

Fig. 1. Star of type U^*AUB and root star of type $U^*A^5UB^2U^*A^3UB$.

(5) A **multistar** is a set of stars inside a root star drawn on the same plane with a coherent orientation.

Fig. 1 shows a concrete example of a multistar. In the middle of the picture there is a star of type U^*AUB and, surrounding it, a root star of type $U^*A^5UB^2U^*A^3UB$.

We are now ready to introduce the main objects in our combinatorial model, namely, maps:

Definition 5.2. A **map** is a decoration of a multistar into a connected graph embedded in the plane by drawing two species of edges between rings or arrows:

(1) A first category of edges, called "colored dotted edges," can be drawn between two different rings of the same color either attached to two outgoing arrows or to two incoming arrows. These edges can only have rings as endpoints, not diamonds nor arrows. Rings can have any number of dotted edges going out of them, possibly none.
(2) A second category of edges, called "colored oriented edge" arises from the connection of an outgoing arrow with an incoming arrow of the same color. The edge takes the color of its arrows. These colored oriented edges form a pairing between elements representing one variable U_i and a variable U_i^* for some i: exactly one incoming arrow is glued to each outgoing arrow.

In addition, no crossing among the above edges is allowed, all arrows must be paired but rings can be attached to any number of dotted edges (including to none).

In the remainder of this section we keep considering pictures drawn on the sphere (and in fact on the plane). They therefore give rise to graphs with vertices, edges and faces—together with additional decoration. For our forthcoming definitions, we need to clarify the notion of 'face': we consider that faces of a graph are the connected components of the complement of the graph on the sphere. However, we take the convention that the original stars are ‘fattened vertices.’ Therefore the interior of stars will not be considered as faces (neither is the exterior of the root star).

Each ‘face’ component of a map is isomorphic to a disc. This is due to the fact that our map is embedded into a sphere. This condition would not be granted in the case of an embedding into a higher genus oriented 2D compact manifold. In this case it would have to stand in the definition of a map of ‘higher genus’: this will be of use for future work but for the sake of simplicity we do not emphasize this notion in this paper.

Next, we define the weight of a map. The boundary of a face is homeomorphic to a circle, it is given an orientation (the orientation of the sphere) and is decorated with diamonds (note that all arrows have been paired); it thus has the structure of a star except for the distinguished element.

Definition 5.3. Assume we are given the tracial state τ of Eq. (2).

- First we define the weight of a face of a map. The boundary of a face has the structure of a star, i.e. it has the topology of a circle with some diamonds on it. We can therefore associate to each of these boundaries a monomial in the A_i's, given up to cyclic permutation (or equivalently up to knowing its first letter). The weight of a face is the trace $\tau(q)$ (which does not depend on cyclic permutations) of the monomial q associated with its boundary.
- The weight of the map m, denoted by $M_m(\tau)$, is the product of the weights of its faces multiplied by $(-1)^{\text{number of dotted edges}}$.

As we said before, not all maps will contribute and we need to define now the notion of admissible maps. Admissibility can be checked by an inductive procedure **IP**, which looks like Tutte's surgery [29]. The idea is to define a procedure which examines one by one each edge of the map. Once an edge has been checked to be correct we will declare it “frozen” and proceed. Thus at each step of this inductive procedure our map contains a certain number of frozen edges which are exactly the part of the map which has been checked. Each step of the procedure amounts to froze some new edges. A map will be declared admissible if we can continue this procedure until all edges are frozen.

Inductive Procedure IP.

Case (a). The root star has no root element. Then it cannot be connected to any other star. Hence, the graph cannot be a map unless there is no other star in which case the map is just the trivial graph with no edges. The trivial graph is declared admissible.

Case (b). The root star has a root element which is associated to a U_i (resp. a U_i^*), for some $i \in \{1, \dots, m\}$.

(1) Then, we first check the admissibility of the dotted edges starting from this root element. These dotted edges are naturally ordered from the nearest of the arrow of the root element to the farthest. We first consider the non-frozen dotted edge which is the farthest and declare it admissible if its other vertex is a ring of an outgoing arrow (resp. ingoing arrow if the root element is attached to an ingoing arrow) and if *there is no other dotted edge* attached to this ring which is farther from its arrow and not frozen. Once this condition is verified, we freeze this dotted edge and the root element remains the root element. We check all dotted edges of the root element inductively. Once a dotted edge has been checked to be admissible, it is frozen and we go on checking the others (starting with the non-frozen one farthest from the root). Once all the

dotted edges of the root element have been checked we look at the second step which examine the arrow.

(2) When all dotted edges are frozen (or when there was no dotted edges on the root element), we check that the arrow of the root element is paired with an arrow of the opposite direction with no unfrozen dotted edges attached on the ring next to it (note that if the root element comes from a U_i^*, it can only be paired with an element of another star since by definition there is no more outgoing arrow on the root star).

Now that we have frozen all the edges attached to our root element, the map may have been cut by these frozen edges into disjoint subgraphs whose boundary (which may contain frozen dotted edges) is homeomorphic to a disc (in the case where it has edges glued with an internal star, we see these other stars as part of the external star by following all the graph connected to the external boundary). In each of these subgraphs, we declare the first (following the orientation of the plane) element (corresponding to a U_i or a U_i^*) after the last frozen dotted edge of its boundary as distinguished. We then define the root element of the boundaries of these subgraphs by the same procedure as for the root star. The boundary of each subgraph is then a root star and these subgraphs have now the structure of a map; we will call them submaps.

For instance, in Fig. 2, once the two dotted edges and the arrow of the root element have been frozen, the map is cut into three disjoint submaps, one which is right of the arrow and which is trivial, one between the oriented edge and its closest dotted edge dotted edge which is also a trivial map and the third one which is left of the rightmost arrow and whose boundary contains the remaining boundary, the two dotted edges and the internal star (since it links the two star we have to visit it twice when we explore the new border). The boundary of this left subgraph is now seen as a star of type $q = UB^5U^*AUBU^*A$. This left subgraph has the same distinguished element as before but a new root element (here the outgoing arrow on its boundary corresponding to the first U in q).

Case (c). We continue the inductive procedure on the submaps until all edges have been checked to be admissible and have been frozen.

Definition 5.4. Assume we are given the tracial state τ of (2).

Let P and $r_1, \dots, r_n$ be monomials, we define the weight of the multistar containing a root star of type P and for each i a star of type r_i by:

$$\mathbb{M}_{r_1,\dots,r_n}(P) = \sum M_m(\tau)$$

where the sum runs over all admissible maps m constructed above this multi-star. Assuming that $V_{\mathbf{t}} = t_1 q_1 + \dots + t_n q_n$ where q_i are monomials, we define the power series in the formal parameter $\mathbf{t}$:

$$\mathbb{M}_{\mathbf{t}}(P) = \sum_{\mathbf{k}\in\mathbb{N}^n} \frac{\mathbf{t}^{\mathbf{k}}}{\mathbf{k}!} \mathbb{M}_{\mathbf{k}}(P)$$

with $\mathbb{M}_{k_1,\dots,k_n}(P) = \mathbb{M}_{q_1,\dots,q_1,\dots,q_n,\dots,q_n}(P)$ where the monomial q_j appears in k_j successive position and $\mathbf{t}^{\mathbf{k}} = \prod t_i^{k_i}$, $\mathbf{k}! = \prod k_i!$.

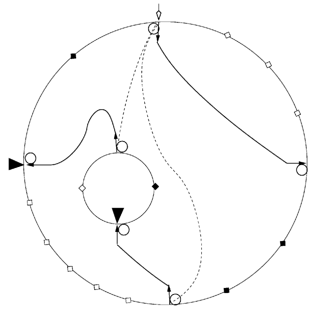

Fig. 2. A possible map. Its weight is $\tau^{\otimes 5}(A^6 \otimes B \otimes B^2 \otimes A^3 \otimes B)$.

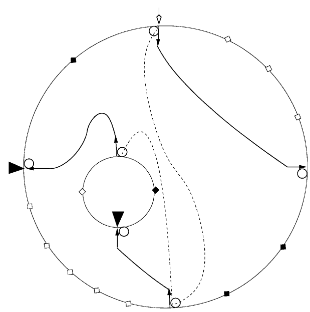

Fig. 3. Another one. Its weight is $\tau^{\otimes 5}(A^6 \otimes B \otimes B^2 \otimes A^3 \otimes B)$.

Remark that we do not count all the maps which contain the stars $r_1, \ldots, r_n$ but only those that are constructed using our inductive rules; they for instance forbid to glue the two same rings more than twice.

However, a given map is counted at most once since there is only one way to decompose it using the procedure **IP**. Indeed, it is easy to check that at each step we have only one possibility for the next step since the dotted edges have to be drawn one after the other following the orientation and no new dotted edge can be drawn after the arrow of the root has been glued.

Example. Let us show some examples. We start from one root star and a star on the sphere (see Fig. 1). We want to construct maps above these stars with our rules, starting with the root element shown by the arrow outside the root star. Figs. 2, 3 and 5 are examples of such maps. Note that the weights of the maps of Figs. 2 and 3 are the same, the only difference is the way the three rings are glued. There is a third way to glue those three rings shown in Fig. 4 which is a map but cannot be obtained by our construction rule (and thus is not admissible).

We now come to the main theorem of this section, namely the graphical expansion result for $\mathbb{M}_{\mathbf{t}}$:

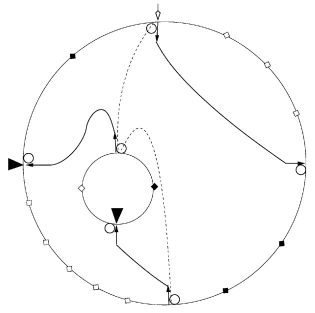

Fig. 4. A counterexample: **IP** is violated because the dotted edge of the root element is linked to the other star by a dotted edge which is not the farthest from the arrow.

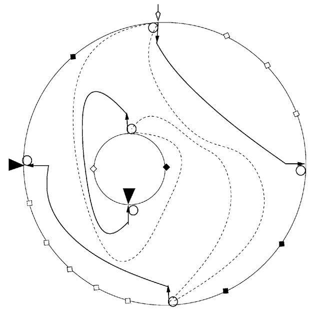

Fig. 5. An admissible map. Its weight is $-\tau^{\otimes 6}(A^5 \otimes A \otimes B \otimes B^2 \otimes A^3 \otimes B)$.

Theorem 5.1. *Let $V = \sum_{1 \leqslant i \leqslant n} t_i q_i$ be a polynomial. Let $\mu_{\mathbf{t}}$ be a solution of* $\mathbf{SD}[V_{\mathbf{t}}, \tau]$ *and $\mathbb{M}_{\mathbf{t}}$ be the power series defined for monomials P by*

$$\mathbb{M}_{\mathbf{t}}(P) = \sum_{\mathbf{k} \in \mathbb{N}^n} \prod_{i=1}^{n} \frac{t_i^{k_i}}{k_i!} \mathbb{M}_{\mathbf{k}}(P)$$

where $\mathbb{M}_{\mathbf{k}}(P)$ is the weighted sum of planar maps with one root star of type P and k_i stars of type q_i. If we extend the definition of $\mathbb{M}_{\mathbf{t}}$ by linearity to any polynomial P then the series $\mathbb{M}_{\mathbf{t}}(P)$ is absolutely convergent in a neighborhood of the origin and,

$$\mathbb{M}_{\mathbf{t}}(P) = \mu_{\mathbf{t}}(P).$$

Proof. For the sake of clarity we first prove the case $V = 0$ and show that $\mathbb{M}(P) := \mathbb{M}_{\mathbf{0}}(P) = \mu_{\mathbf{0}}(P)$ for a monomial P.

We proceed by induction on the total degree in U_i, $1 \leqslant i \leqslant m$, in q.

Suppose that there is no variable U_i in P. Then either there is no variable U_i^* and both sides of the equality are equal to $\tau(P)$, or there is a U_i^* and both sides vanish: the left-hand side by freeness between U_i and the A_i's and the fact that all non-trivial moments of U_i are 0 and the right-hand side because one cannot glue the arrow coming out from this U_i^* anywhere.

We assume our identification proved when the degree of P in the U_i's is less than k. We next take q with degree in the U_i's equal to $k+1$. Thus we can assume that there is a U_i in P, and we consider the last one in P so that $P = pU_i b$ with b a polynomial in the U_j^* and the A_j's, $1 \leqslant j \leqslant m$. By definition, $\mathbb{M}(pU_i b) = \mathbb{M}(bpU_i)$ since it depends only on the position of the last U_i (which corresponds to the root element). Thus, we may assume that P is of the form QU_i with Q of degree k. We apply the Schwinger–Dyson equation to this quantity:

$$\mu(QU_i) = -\sum_{Q=RU_iS} \mu(RU_i) \otimes \mu(SU_i) + \sum_{Q=RU_i^*S} \mu(R) \otimes \mu(S). \tag{16}$$

Now, we can apply our induction hypothesis since all polynomials appearing in the right-hand side have degree strictly smaller than $k+1$.

We need to show that this is exactly the induction relation for maps. To construct a map above a star of type QU_i, we first look at the root element U_i and we have to decide what to do first with the dotted edges. There are two possibilities:

(1) The first possibility is that there is no dotted edge going outside of the ring of the root. In such a case, we can glue the arrow to any other arrow of opposite direction and of the same color (corresponding to a variable U_i^*). This implies that Q decomposes into RU_i^*S and we construct an oriented edge between U_i and U_i^*. Thus we separate the map into two parts and we have to construct a map above the R part and another one above the S part (this is the case 2 of **IP**). This gives

$$\mathbb{M}(R)\mathbb{M}(S)$$

possibilities which is exactly the possibilities counted by the second term in the right-hand side of (16).

(2) The second possibility is that we glue the root ring to another ring with a dotted edge. Thus Q must decompose into RU_iS and the creation of the dotted edge amounts to decompose the map into RU_i and SU_i and again to continue the construction of the map we will have to construct a map above the RU_i part and another one above the SU_i part (note here that when a dotted edge is attached to a circle of a U_i, the arrow and the circle keep their structure and live on the right of the dotted edge). In this procedure, we have fixed one dotted edge and thus multiplied the contribution of the resulting map by -1 (this is the case 1 of **IP**). The resulting contribution to $\mathbb{M}$ is therefore $-\mathbb{M}(RU_i)\mathbb{M}(SU_i)$. Thus, the first term in (16) computes the operation of gluing rings by dotted edges.

Putting these two possibilities together we see that the state μ and the enumeration of maps $\mathbb{M}$ satisfy the same induction so that they are equal; $\mathbb{M}(pU_ib) = \mu(pU_ib)$ for any b monomial which do no contain any of the $(U_i, 1 \leqslant i \leqslant m)$. Note here that no dotted edges between rings of incoming arrows can be drawn since if there are no outgoing arrows in a map, but some U_i^*, there is no contribution. By traciality of μ, we deduce as well that $\mathbb{M}$ is tracial (and therefore $\mathbb{M}$ does not depend on the choice of the root element). Indeed, if we decompose p, q into $p = p_1U_{i_1}p_2U_{i_2}\cdots p_{n-1}U_{i_{n-1}}p_n$ and $q = q_1U_{j_1}q_2U_{j_2}\cdots q_{r-1}U_{j_{r-1}}q_r$ with monomials p_i, q_i which does no contain any of the $(U_i, 1 \leqslant i \leqslant m)$, then

$$\begin{aligned}\mathbb{M}(pq) &= \mathbb{M}\big((pq_1U_{j_1}q_2U_{j_2}\cdots U_{j_{r-2}}q_{r-1})U_{j_{r-1}}q_r\big)\\ &= \mu(pq_1U_{j_1}q_2U_{j_2}\cdots U_{j_{r-2}}q_{r-1}U_{j_{r-1}}q_r) = \mu(pq)\end{aligned}$$

$$= \mu(qp) = \mu\big((qp_1U_{i_1}p_2U_{i_2}\cdots p_{n-1})U_{i_{n-1}}p_n\big) = \mathbb{M}(qp).$$

Now we turn to the general V case.

We first check the induction relation when the root star P contains a U_i for some $i \in \{1,\dots,m\}$ so that we can write $P = QU_i$. Let us denote for n-tuples $\mathbf{k} = (k_1,\dots,k_n)$ and $\boldsymbol{\ell} = (l_1,\dots,l_n)$, $\binom{\mathbf{k}}{\boldsymbol{\ell}} = \prod_i \binom{k_i}{l_i}$. We check the formal equality by considering the induction relation, now given by:

$$\begin{aligned}
\mu^{\mathbf{k}+\mathbf{e_j}}(QU_i) = &- \sum_{\boldsymbol{\ell}\leqslant \mathbf{k}+\mathbf{e_j}} \sum_{Q=RU_iS} \binom{\mathbf{k}+e_j}{\boldsymbol{\ell}} \mu^{\boldsymbol{\ell}}(RU_i) \otimes \mu^{\mathbf{k}+\mathbf{e_j}-\boldsymbol{\ell}}(SU_i) \\
&+ \sum_{\boldsymbol{\ell}\leqslant \mathbf{k}+\mathbf{e_j}} \sum_{Q=RU_i^*S} \binom{\mathbf{k}+e_j}{\boldsymbol{\ell}} \mu^{\boldsymbol{\ell}}(R) \otimes \mu^{\mathbf{k}+\mathbf{e_j}-\boldsymbol{\ell}}(S) \\
&- \sum_{q_j=RU_iS} k_j\mu^{\mathbf{k}}(QU_iSRU_i) - \sum_{q_j=RU_i^*S} k_j\mu^{\mathbf{k}}(QSR).
\end{aligned} \tag{17}$$

We need to show that the enumeration of maps satisfies the same relation. We start by putting stars of type $(q_j, 1 \leqslant j \leqslant n)$ inside a root star of type QU_i and we wonder what happens to the root element U_i. We apply one step of **IP**. Two things can happen. Either we link U_i to another part of Q and in that case we have already shown that the possibilities are enumerated by the first two terms of the induction relation. Here, note that the product of $\binom{k_i}{\ell_i}$ corresponds to the possible distribution of stars in each part (or submap) of the map, since all the stars are labeled.

Thus we need to show that the two other terms take into account the case where U_i is linked to another star of type q_j. According to our construction rules we have two possibilities:

(1) Starting from U_i we glue the arrow to an arrow of the same color entering a star of type q. This rule forbids any other gluing from U_i, this is counted by

$$\sum_{q_j=RU_i^*S} k_j\mu(QSR).$$

The coefficient k_j counts the number of choices for the star of type q_j since they are all labeled.

(2) The other possibility is to glue the ring to a ring of the same color. This leads to

$$-\sum_{q_j=RU_iS} k_j\mu(QU_iSRU_i)$$

possibilities.

In the case where P does not contain any U_i, $1 \leqslant i \leqslant m$ but still some U_i^*, the root of the root star can only be glued by a dotted edge to any other U_i^*, or by a directed edge to a U_i of a star. The resulting induction relation is exactly given by the formula obtained by conjugation of (17), hence again $M_{\mathbf{k}}(P) = \mu^{\mathbf{k}}(P)$. This completes the proof. □

This theorem gives a combinatorial interpretation in term of maps to the unitary integrals. The fact that we do not take the sum on all maps but only on admissible ones makes this interpretation less transparent than the one for the Gaussian case found in [7]. However, now that we know that the series can be identified to the matrix integral, we obtain some combinatorial identities which show that **IP** is less rigid than it looks like.

Corollary 5.1. *Let $V = \sum t_i q_i$ be a polynomial.*

(1) *For all P, Q,*

$$\mathbb{M}_{\mathbf{t}}(PQ) = \mathbb{M}_{\mathbf{t}}(QP).$$

(2) *For all monomials $r_1, \ldots, r_n, r_{n+1}$, and all permutation σ of $n+1$ elements,*

$$\mathbb{M}_{r_1,\ldots,r_n}(r_{n+1}) = \mathbb{M}_{r_{\sigma(1)},\ldots,r_{\sigma(n)}}(r_{\sigma(n+1)}).$$

(3) *Assume that we define another procedure to define the root element of the root star (for example we pick the root element to be the second ring available if possible, or we pick a ring at random, or any other choice which may change during* **IP** *for the root stars that are created during the procedure when new faces are added). This will change the notion of admissible maps and we can define a new weighted sum $\mathbb{M}'_{r_1,\ldots,r_n}(P)$ and a new series $\mathbb{M}'_{\mathbf{t}}(P)$ where the sum occurs on these new maps. For all $r_1, \ldots, r_n, P$,*

$$\mathbb{M}_{r_1,\ldots,r_n}(P) = \mathbb{M}'_{r_1,\ldots,r_n}(P),$$
$$\mathbb{M}_{\mathbf{t}}(P) = \mathbb{M}'_{\mathbf{t}}(P).$$

Note that due to the definition of admissible maps via the procedure **IP**, those properties are far from being obvious from a purely combinatorial point of view. Still they will appear as an easy consequence of the identification with the matrix model.

Obviously different roots lead to a different procedure **IP**, and thus potentially to different maps. It is actually possible to see through examples that this phenomenon actually happens.

However, it follows from the second point of the corollary that the choice of the root does not affect the weighted sum. The first and third points show that the choice of the root element and of the root star does not affect the final series. We were not able to give a more direct combinatorial proof of that result.

To be more specific on the impact of the choice of the roots on the maps, let us call clusters the equivalence class of rings for the equivalence relation generated by $a \sim b$ if the ring a is glued to the ring b by a dotted edge. Changing the choices of the roots will lead to different admissible maps since it will allow different positions for the dotted edges. For example, they were three choices for the starting root in Fig. 1. For each of these choices, two of the three maps represented in Figs. 2, 3 and 4 would have been reachable by the inductive construction **IP** but not the third one. The one who is not constructible depends on the choice of the first root. It seems that if the maps are different, nevertheless the clusters are the same and in that simple case, knowing this cluster is sufficient to define the faces created by the dotted edges and thus the weight of the maps.

Proof. Changing the root element of a star is the same thing as making a circular permutation of the variable of the associated monomial. The theorem shows that weighted sums are equal to the limit of the empirical measure of the matrix model which are tracial. The first and third items are a direct consequence of this identification.

For the second item, observe that permuting the first n monomials doesn't change the sum by its definition. Thus we only need to show that

$$\mathbb{M}_{r_1,\dots,r_n}(P) = \mathbb{M}_{P,r_2,\dots,r_n}(r_1).$$

Let us define $V = \sum_i u_i r_i + tP$. We will again use the identification with the matrix model but now we will use the formal version. The coefficient $\mathbb{M}_{r_1,\dots,r_n}(P)$ appears as the coefficient of the limit tracial power state μ^f by Corollary 3.1 and Theorem 5.1. More precisely,

$$\mathbb{M}_{r_1,\dots,r_n}(P) = \lim_N \frac{\partial^n}{\partial u_1 \cdots \partial u_n} \mu^f(P)\bigg|_{u_i=0}.$$

We now use the fact that μ^f is the limit coefficientwise of the formal model defined in (9). Thus,

$$\begin{aligned}\mathbb{M}_{r_1,\dots,r_n}(P) &= \lim_N \frac{\partial^n}{\prod_i \partial u_i} \frac{E[\hat{\mu}^N(P) e^{N^2 \hat{\mu}^N(V)}]}{E[e^{N^2 \hat{\mu}^N(V)}]}\bigg|_{u_i=0,t=0} \\ &= \lim_N \frac{\partial^{n+1}}{\partial t \prod_i \partial u_i} \frac{1}{N^2} \ln E\big[e^{N^2 \hat{\mu}^N(V)}\big]\bigg|_{u_i=0,t=0}.\end{aligned}$$

We conclude by noticing that this last expression is symmetric in the monomials $r_1, \dots, r_n, P$. □

6. Application to free probability

In this section we show how one can recover some classical results of free probability by using the combinatorial approach of Section 5.

Let us assume that the U_i's are chosen independently according to the Haar measure. If we define $X_i = U_i^* A_i U_i$ then the X_i's are asymptotically free (according to a theorem of Voiculescu [31]) and with fixed distribution μ uniquely defined by the distribution of the A_i's. We are interested in using our setup to compute limits of moments of these variables or in other word to compute the moments of free variables:

$$\mu(X_{i_1} \cdots X_{i_k}).$$

According to our interpretation this can be computed by looking at the maps above the star of type $X_{i_1} \cdots X_{i_k}$ without any other stars, in other words we have to focus on computations of $\mathbb{M}(q) = \mathbb{M}_0(q)$ which turns out to be equal to $\mu(q)$ where μ is the free state product (see Corollary 3.2).

We are interested in using this method to compute some non-commutative moments of free variables, in relation with Speicher's non-crossing cumulants theory, cf. [27].

Let $A_1, \dots, A_n$ be self-adjoint variables and U a unitary matrix, free from the A_i's. Then choosing k indices $i_1, \dots, i_k$ in $\{1, n\}$ one has

$$\mu(A_{i_1} \cdots A_{i_k}) = \mu\big(U^* U A_{i_1} \cdots U^* U A_{i_k}\big).$$

Let us apply the Schwinger–Dyson equation with respect to U to the above equality, and let us rearrange the sum according to the non-crossing partition of A_i's generated by the oriented edges. Obviously one obtains a formula of type

$$\mu(A_{i_1} \cdots A_{i_k}) = \sum_{\pi \in NC(k)} \tilde{K}_\pi(A_{i_1}, \dots, A_{i_k}) \tag{18}$$

where $NC(k)$ is the non-crossing partitions and $\tilde{K}_\pi$ is a k-linear form multiplicative along the blocks of π in the sense of Speicher: if $\pi = \{V_1, \dots, V_n\}$ with the block $V_i = \{a_1^i, \dots, a_{r_i}^i\}$

$$\tilde{K}_\pi(X_1 \cdots X_k) = \prod_i \tilde{K}_{(r_i)}(X_{a_1^i}, \dots, X_{a_{r_i}^i})$$

where (r_i) represents the partition on r_i elements with only one block.

The fact that such a formula holds true for any choice of non-commutative laws for A_i's proves via the moment-cumulant formula that $\tilde{K}_\pi$ has to be Speicher's non-crossing cumulants K_π. But it is also given as a sum on maps by our graphical model.

Let us recap this in the following proposition:

Proposition 6.1. *The nth non-crossing cumulant of the variables $A_1, \dots, A_p$ is the sum of the weights of all maps build over the star build by putting in the clockwise order a ring, a diamond of color i_1, a ring, a diamond of color $i_2, \dots,$ a ring, a diamond of color i_p.*

Many other properties of the cumulants can be read from our graphical representation; for example, the fact that $K_n(X_1, \dots, X_n) = 0$ as soon as there are occurrence of free elements or the non-crossing Moebius formula due to Speicher.

It is interesting to mention here that papers [24] and [25] have developed a calculus on annuli which seems to be related to our graphical model. However these approaches only deal with the asymptotics of second order cumulants whereas our approach via formal calculus, see Section 4, allows us to deal with arbitrary order cumulants. The actual relation can be found in [10], where convolution on partitioned permutations is showed to be the relevant algebraic tool to handle higher order freeness.

But the results in our paper give an explicit algorithmic description of the Moebius inversion formula and therefore of higher order cumulants. As in the one star case, cumulants are also obtained by inserting an outer U^*U between each variable of each star and by looking at generating function where U is linked to its neighboring U^*.

7. Application to the asymptotics of $I_N(V, A_i^N)$

In this section, we investigate the free energy by using the combinatorial interpretation of the previous section.

Let $(q_1, \dots, q_n)$ be fixed monomials in $\mathbb{C}\langle (U_i, U_i^*, A_i)_{1 \leqslant i \leqslant m}\rangle$, let $V = \sum t_i q_i$ be a self-adjoint polynomial and $I_N(V, A_i)$ be given by (1).

Theorem 7.1. *There exists $\varepsilon = \varepsilon(q_1, \ldots, q_n)$ so that for any $\mathbf{t} \in \mathbb{C}^n \cap B(0, \varepsilon)$ such that $V = V^*$ for any $\alpha \in [-1, 1]$,*

$$F_{V,\tau}(\alpha) := \lim_{N\to\infty} \frac{1}{N^2} \log I_N\big(\alpha V_{\mathbf{t}}, A_i^N\big) = \sum_{\mathbf{k}\in\mathbb{N}^n\setminus(0,\ldots,0)} \prod_{i=1}^{n} \frac{(\alpha t_i)^{k_i}}{k_i!} \mathbb{M}_{\mathbf{k}}(q_1, \ldots, q_n, \tau).$$

Moreover,

$$\mathbb{M}_{\mathbf{k}}(q_1, \ldots, q_n, \tau) = \sum_{m \text{ admissible maps with } k_i \text{ stars } q_i} M_m(\tau)$$

is the weighted sum of maps constructed above k_i stars of type q_i for all i, after choosing one of them as a root star (this is well defined according to Corollary 5.1*).*

Proof. Let us define

$$F_{\mathbf{t}}^N = \frac{1}{N^2} \log I_N\big(V_{\mathbf{t}}, A_i^N\big).$$

Then, if $\alpha \in \mathbb{R}$,

$$\partial_\alpha F_{\alpha\mathbf{t}}^N = \int \hat{\mu}^N(V_{\mathbf{t}})\, d\mu_{V_{\alpha\mathbf{t}}}^N.$$

Assume that $\mathbf{t}$ is small enough so that Corollary 3.1 holds and remark that $V_{\alpha\mathbf{t}}$ is self-adjoint and such that $|\alpha t_i| \leqslant \varepsilon$ for all i and all $0 \leqslant \alpha \leqslant 1$. Thus, for $\alpha \in [0, 1]$,

$$\lim_{N\to\infty} \partial_\alpha F_{\alpha\mathbf{t}}^N = \mu_{\alpha\mathbf{t}}(V_{\mathbf{t}})$$

with $\mu_{\alpha\mathbf{t}}$ the solution to $\mathbf{SD}[\alpha V_{\mathbf{t}}, \tau]$. By dominated convergence theorem (since $\partial_\alpha F_{\alpha\mathbf{t}}^N$ is uniformly bounded in N and $\alpha \in [0, 1]$), we deduce that

$$\lim_{N\to\infty} F_{\alpha\mathbf{t}}^N = \int_0^1 \mu_{\alpha\mathbf{t}}(V_{\mathbf{t}})\, d\alpha$$

where we used that $F_0^N = 0$. $\square$

Here also, we obtain the following important corollary, as a consequence of Corollary 4.1.

Corollary 7.1. *The following holds true:*

$$\lim_{N\to\infty} \frac{\partial^k}{\partial z^k} N^{-2} \log \int_{\mathcal{U}_N^m} e^{zNTr(V(U_i, U_i^*, A_i^N, 1\leqslant i\leqslant m))}\, dU_1 \cdots dU_m \bigg|_{z=0} = \frac{\partial^k}{\partial z^k} F_{V,\tau}(z)\bigg|_{z=0}.$$

In particular, this result allows us to give an expansion of the Harish-Chandra–Itzykson–Zuber integral as a generating function of the number of some maps. Let us recall the exact expression of this integral:

$$F_N^{A,B}(z) := \frac{1}{N^2} \log HCIZ(zA, B) = \frac{1}{N^2} \log \int_{\mathcal{U}_N} e^{zN \operatorname{Tr}(U^* AUB)} \, dU.$$

If A and B are Hermitian, the maps appearing in the expansion contain only stars of type U^*AUB (see the star in the middle of Fig. 1). Besides we can build these maps without considering the rings attached to variable U^* since we will always be able to choose the root element to be a U (a U^* always comes with a U for this potential).

Since the number of diagrams is growing quickly we compute only the first term of the expansion. Note that when gluing the arrow of the root element of the root star, we must always glue it to another incoming arrow of another star and hence we shall never see the case of a root star with no U_i's. Therefore, we do not see dotted edges between incoming arrows.

Besides, we consider only the case where the distribution is centered, that is when $\tau(A) = \tau(B) = 0$. The other cases can be deduced easily from this one since we have the relation

$$F_N^{a+A,b+B}(z) = F_N^{A,B}(z) + \frac{z}{N}(b \operatorname{Tr} A + a \operatorname{Tr} B) + zab.$$

In terms of diagrams, this means that we only need to consider diagrams such that no face contains only one diamond.

According to the previous theorem, $\lim_{N\to\infty} F_N^{A,B}(z)$ has, for small z, an expansion $\sum_n F_n z^n$. We now use this graphical representation to compute the first terms of this integral.

Since the distributions are centered, the first term F_1 is zero.

The second term F_2 consists of maps constructed with two stars of type U^*AUB. There is only one way to add edges between these two stars to construct a connected map without faces which contains only one diamond, this is represented by Fig. 6. We obtain a map with two faces. One has two diamonds associated to A and the other one two diamonds associated to B. Thus the weight of this map is $\tau(A^2)\tau(B^2)$. Since there is no gluing between the rings they are no other signs. There is only one way to distribute the labels on this picture (that is the second distribution leads to the same map) thus to obtain F_2 we only need to divide by 2!,

$$F_2 = \frac{1}{2}\tau\left(A^2\right)\tau\left(B^2\right).$$

We can continue this for the next terms in the expansion, the third term (see Fig. 7) is in the same spirit and leads to

$$F_3 = \frac{1}{3}\tau\left(A^3\right)\tau\left(B^3\right).$$

The fourth term is the first one where gluings between the rings appear. Thus weights with negative coefficients can occur. The sign of a map is easy to compute, it is -1 to the power the number of dotted lines in the map. Equivalently since in the case of HCIZ integral the number of oriented edges is equal to the number of stars, this number is also equal to the number of faces of the map and thus to the number of factor in the product of moments of the weight. In Fig. 8,

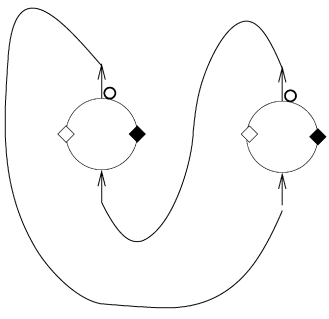

Fig. 6. Second term in the expansion of the HCIZ integral.

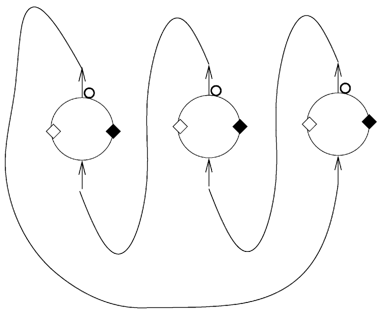

Fig. 7. Third term in the expansion of the HCIZ integral.

we have drawn all unlabeled planar maps one can construct with 4 stars. To compute the exact coefficient of each map one has to multiply it by the number of way to distribute the labels and divide by 4!.

This leads to,

$$\begin{aligned} F_4 = {} & \frac{1}{4}\tau\big(A^4\big)\tau\big(B^4\big) - \frac{1}{2}\tau\big(A^2\big)^2\tau\big(B^4\big) - \frac{1}{2}\tau\big(A^4\big)\tau\big(B^2\big)^2 \\ & + \frac{1}{2}\tau\big(A^2\big)^2\tau\big(B^2\big)^2 + \frac{1}{4}\tau\big(A^2\big)^2\tau\big(B^2\big)^2. \end{aligned}$$

Here the weight are given in the same order as the maps in the figure. Note a new and interesting feature that appears in the third map: two rings are linked by more than one dotted edge.

The other terms can be computed in the same way, for example Fig. 9 represents the fifth term and gives

$$\begin{aligned} F_5 = {} & \frac{1}{5}\tau\big(A^5\big)\tau\big(B^5\big) - \tau\big(A^2\big)\tau\big(A^3\big)\tau\big(B^5\big) - \tau\big(A^5\big)\tau\big(B^2\big)\tau\big(B^3\big) \\ & + 4\tau\big(A^2\big)\tau\big(A^3\big)\tau\big(B^2\big)\tau\big(B^3\big). \end{aligned}$$

Thus the first terms agree with the expansion given in [35] on page 23, besides this allows us to answer a question raised in this paper. Indeed, the authors ask if there is an explanation to the fact that the coefficient of F_n all seem to be integer multiple of $\frac{1}{n}$. This is easy to prove

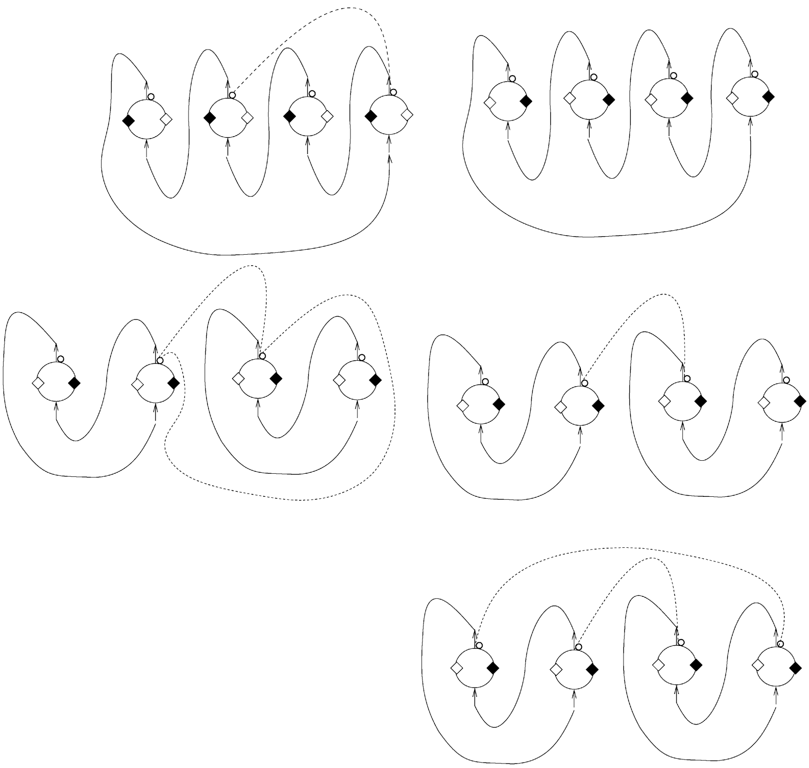

Fig. 8. Fourth term in the expansion of the IZ integral.

with this graphical interpretation. To compute the contribution of a given unlabeled map we must distribute the labels $\{1, \dots, n\}$ on its stars, count the number of different map that we obtain and divide by $n!$. But after choosing the star which received the label 1 we have $(n-1)!$ ways to distribute the remaining labels and they all lead to different maps (note that on the other hand, due to possible symmetry in the unlabeled map, different choices for the star with the label 1 may lead to the same maps). Thus the coefficient in front of this map is a multiple of $\frac{(n-1)!}{n!} = 1/n$. More precisely it is $1/n$ times the number of choices of the star which carry the label 1 that will lead to different maps, in particular it is always less than 1 (the coefficient 4 in the expression of F^5 comes from the contribution of 4 non-isomorphic non-labeled maps).

To conclude this section, we wish to point out that we can recover results in [8] and [15] about scalings of HCIZ integral. In these two papers, the scaling where A has small rank is studied, which amounts to considering only terms $\tau(A^k) \cdot P(B)$. Here the transformation depicted in Section 6 applies and we see that $P(B)$ has to be $k^{-1}K_k(B)$. In particular this means in the case that A is a rank 1 projection, that $N^{-1} \log HCIZ$ tends to the primitive of Voiculescu's R-transform.

Finally, Theorem 7.1 also includes the case where A and B are not Hermitian but U^*AUB self-adjoint. Then, if we denote $\Re e(C) = (C + C^*)/2$, $\Im m(C) = (C - C^*)/2i$ for $C = A$ and B, we can write $U^*AUB = U^*\Re e(A)U\Re e(B) + U^*\Im m(A)U\Im m(B)$ so that now there are two types of stars, namely those decorated with cubes corresponding to $\Re e(A)$ and $\Re e(B)$, or to $\Im m(A)$ and $\Im m(B)$. The edges will be the same than in the case of Hermitian matrices since

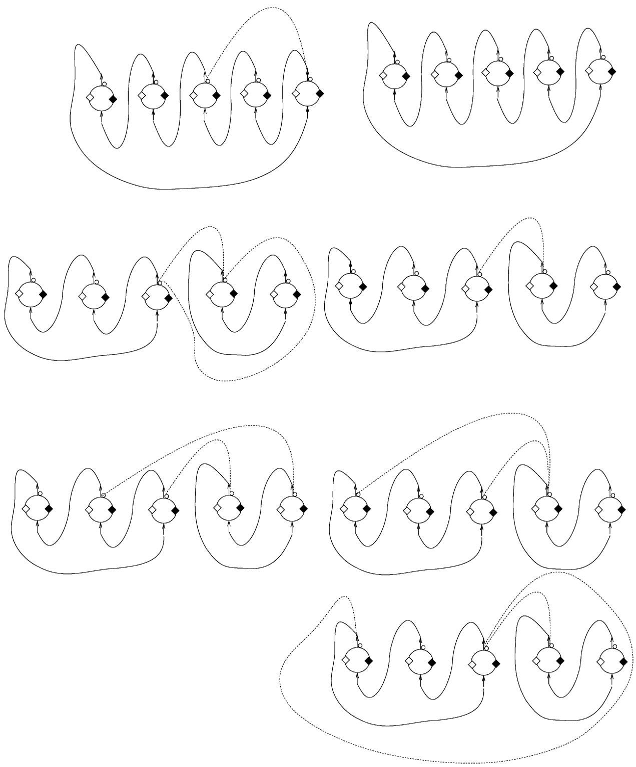

Fig. 9. Fifth term in the expansion of the IZ integral.

the system of arrows and rings is the same, but the weights will now be given in terms of joint moments of $(\Re e(A), \Im m(A))$ and $(\Re e(B), \Im m(B))$. This gives Theorem 0.3.

8. Application to Voiculescu free entropy

Voiculescu's microstates free entropy is defined as the asymptotic volume of matrices whose empirical distribution approximates sufficiently well a tracial state. Up to a Gaussian factor, it is given by

$$\chi(\mu) = \limsup_{\substack{\varepsilon\downarrow 0\\ k\uparrow\infty, R\uparrow\infty}} \limsup_{N\to\infty} \frac{1}{N^2} \log \mu_N^{\otimes m}\big(\Gamma_R(\mu,\varepsilon,k)\big)$$

with μ_N the Gaussian measure on $\mathcal{H}_N$ and $\Gamma_R(\mu,\varepsilon,k)$ the microstates

$$\Gamma_R(\mu,\varepsilon,k) = \left\{X_1,\dots,X_m \in \mathcal{H}_N\colon \left|\frac{1}{N}\operatorname{Tr}(X_{i_1}\cdots X_{i_p}) - \mu(X_{i_1}\cdots X_{i_p})\right| < \varepsilon,\right.$$
$$\left. p \leqslant k,\ i_\ell \in \{1,\dots,m\},\ \|X_i\|_\infty \leqslant R\right\}.$$

When $m = 1$, it is well known [30] that $\mu \in \mathcal{P}(\mathbb{R})$ and

$$\chi(\mu) = I(\mu) = \iint \log|x-y|\,d\mu(x)\,d\mu(y) - \frac{1}{2}\int x^2\,d\mu(x) + \text{const}.$$

Moreover, we can replace the lim sup by a lim inf in the definition of χ. Such answers (convergence and formula for χ) are still open in general when $m \geqslant 2$ (see [5] for bounds). However, if μ is the law of m free variables with respective laws μ_i, then these questions are settled and

$$\chi(\mu) = \sum_{i=1}^{m} I(\mu_i).$$

We here want to emphasize that our result provides a small step towards dependent variables by showing convergence and giving a formula for the type of laws μ solutions of the Schwinger–Dyson's equations $\mathbf{SD}[V,\tau]$. Indeed, we shall prove that

Theorem 8.1. *Let μ be the law of m self-adjoint variables X_i with marginal distribution $(\mu_1,\dots,\mu_m)$. Assume that X_i can be decomposed as $X_i = U_i D_i U_i^*$ with U_i unitary matrices such that the joint law ν of $(D_i, U_i, U_i^*)_{1\leqslant i\leqslant m}$ satisfy $\mathbf{SD}[V,\tau]$ with τ the law of m free variables with marginal distribution $\mu_1,\dots,\mu_m$ and some potential $V = \sum_{i=1}^{n} t_i q_i$. Assume that the t_i's are small enough so that Corollary* 3.1 *holds. Assume also that the hypotheses of Theorem* 7.1 *hold. Then,*

$$\chi(\mu) = \liminf_{\substack{\varepsilon\downarrow 0\\ k\uparrow\infty}} \liminf_{N\to\infty} \frac{1}{N^2}\log \mu_N^{\otimes m}\big(\Gamma_R(\mu,\varepsilon,k)\big)$$

and a formula of $\chi(\mu)$ can be given in terms of the $\mu^{\mathbf{k}}$'s of Theorem 4.2.

Proof. Indeed, let us consider $V = V(U_i A_i U_i^*, 1\leqslant i\leqslant m)$ with V a self-adjoint polynomial and μ the unique solution of $\mathbf{SD}[V,\tau]$ with τ the law of the A_i, $1\leqslant i\leqslant m$ which is now chosen to be the law of m free variables with marginals distribution μ_i, $1\leqslant i\leqslant m$. Under the law $\mu_N^{\otimes m}$, we can diagonalize the matrices $X_i = U_i D_i U_i^*$ with U_i following the Haar measure on $\mathcal{U}_N$, and if d is the Dudley metric, we find that for N sufficiently large

$$\begin{aligned}
\mathbb{L}_N &:= \mu_N^{\otimes m}\big(\Gamma_R(\mu,\varepsilon,k)\big)\\
&= \mu_N^{\otimes m}\big(d\big(\hat\mu^N_{D_i},\mu_i\big) < \varepsilon;\ \hat\mu^N_{U_i D_i U_i^*, 1\leqslant i\leqslant m} \in \Gamma_R(\mu,\varepsilon,k)\big)\\
&= \int_{\substack{d(\hat\mu^N_{D_i},\mu_i)<\varepsilon\\ \|D_i\|_\infty\leqslant R}} \Bigg(\int_{\hat\mu^N_{(U_i D_i U_i^*)_{1\leqslant i\leqslant m}}\in\Gamma_R(\mu,\varepsilon,k)} dU_1\cdots dU_m\Bigg)\prod_{1\leqslant i\leqslant m} d\sigma_N(\lambda_i)
\end{aligned}$$

where we denoted $d\sigma_N$ the probability measure on $\mathbb{R}^N$

$$d\sigma_N(\eta) := Z_N^{-1}\prod_{k\neq j}|\eta_k-\eta_j|^2 e^{-\frac{N}{2}\sum(\eta_j)^2}\prod_{1\leqslant j\leqslant N} d\eta_j.$$

In these notations, $D_i = \operatorname{diag}(\lambda_1^i, \dots, \lambda_N^i)$ and $\lambda = (\lambda_1, \dots, \lambda_N)$. Hereafter, $\hat{\mu}^N_{\{E_i\}_{1\leqslant i\leqslant n}}$ denotes the empirical distribution of $\{E_i\}_{1\leqslant i\leqslant n}$; $\hat{\mu}^N_{\{E_i\}_{1\leqslant i\leqslant n}}(P) = N^{-1}\operatorname{Tr}(P(E_i, 1\leqslant i\leqslant n))$. As a consequence, applying the large deviations result of [3] to the diagonal matrices D_i, we find that there exists $o(1)$ going to zero with ε such that

$$\begin{aligned}
\mathbb{L}_N &\leqslant e^{N^2\sum_{i=1}^m I(\mu_i)+N^2 o(1)} \sup_{\substack{d(\hat{\mu}^N_{D_i},\mu_i)<\varepsilon\\ \|D_i\|_\infty\leqslant R}} \int_{\hat{\mu}^N_{\{U_iD_iU_i^*\}_{1\leqslant i\leqslant m}}\in\Gamma_R(\mu,\varepsilon,k)} dU_1\cdots dU_m \\
&:= e^{N^2\sum_{i=1}^m I(\mu_i)+N^2 o(1)}\mathbb{L}_N^1
\end{aligned}$$

with for k greater than the degree of V,

$$\begin{aligned}
\mathbb{L}_N^1 &= \sup_{\substack{d(\hat{\mu}^N_{D_i},\mu_i)<\varepsilon\\ \|D_i\|_\infty\leqslant R}} \int_{\hat{\mu}^N_{\{U_iD_iU_i^*\}_{1\leqslant i\leqslant m}}\in\Gamma_R(\mu,\varepsilon,k)} e^{N\operatorname{Tr}(V)-N\operatorname{Tr}(V)}\, dU_1\cdots dU_m \\
&= e^{-N^2\mu(V)+N^2\varepsilon} \sup_{\substack{d(\hat{\mu}^N_{D_i},\mu_i)<\varepsilon\\ \|D_i\|_\infty\leqslant R}} \int_{\hat{\mu}^N_{\{U_iD_iU_i^*\}_{1\leqslant i\leqslant m}}\in\Gamma_R(\mu,\varepsilon,k)} e^{N\operatorname{Tr}(V)}\, dU_1\cdots dU_m \\
&\leqslant e^{-N^2\mu(V)+N^2\varepsilon} \sup_{\substack{d(\hat{\mu}^N_{D_i},\mu_i)<\varepsilon\\ \|D_i\|_\infty\leqslant R}} \int e^{N\operatorname{Tr}(V)}\, dU_1\cdots dU_m \\
&= e^{-N^2\mu(V)+N^2\varepsilon} \sup_{d(\hat{\mu}^N_{D_i},\mu_i)<\varepsilon,\, \|D_i\|_\infty\leqslant R} I_N(V, D_i).
\end{aligned}$$

Now, for fixed R, any D_i, D_i' in $d(\hat{\mu}^N_{D_i}, \mu_i) < \varepsilon$, $\|D_i\|_\infty \leqslant R$

$$\left|\frac{1}{N^2}\log I_N(V, D_i) - \frac{1}{N^2}\log I_N\bigl(V, D_i'\bigr)\right| \leqslant \eta(\varepsilon, R),$$

with $\eta(\varepsilon, R)$ going to zero as ε goes to zero for any fixed R. Hence,

$$\limsup_{N\to\infty}\frac{1}{N^2}\log I_N(V, D_i) \leqslant F(V, \mu_i) + \eta(\varepsilon, R)$$

with $F(V, \mu_i)$ the limit of $N^{-2}\log I_N(V, A_i)$ given in Theorem 7.1 when the distribution of the A_i converges to free variables with marginal distribution μ_i. We thus have proved, letting ε going to zero and then R, k to infinity, that

$$\chi(\mu) \leqslant \sum_{i=1}^m I(\mu_{A_i}) - \mu(V) + F(V, \mu_i).$$

Conversely, we have

$$\mathbb{L}_N \geqslant e^{N^2 \sum_{i=1}^m I(\mu_i)+N^2 o(\varepsilon)} \mathbb{L}_N^2$$

with

$$\begin{aligned}
\mathbb{L}_N^2 &:= \inf_{\substack{d(\hat{\mu}^N_{D_i},\mu_i)<\varepsilon \\ \|D_i\|_\infty \leqslant R}} \int_{\hat{\mu}^N_{(U_i D_i U_i^*)_{1\leqslant i\leqslant m}} \in \Gamma_R(\mu,\varepsilon,k)} dU_1 \cdots dU_m \\
&= e^{-N^2\mu(V)+N^2 o(\varepsilon)} \inf_{\substack{d(\hat{\mu}^N_{D_i},\mu_i)<\varepsilon \\ \|D_i\|_\infty \leqslant R}} \int_{\hat{\mu}^N_{\{U_i D_i U_i^*\}_{1\leqslant i\leqslant m}} \in \Gamma_R(\mu,\varepsilon,k)} e^{N\,\mathrm{Tr}(V)}\, dU_1 \cdots dU_m \\
&\geqslant e^{-N^2\mu(V)+N^2 o(\varepsilon)} \inf_{\substack{d(\hat{\mu}^N_{D_i},\mu_i)<\delta \\ \|D_i\|_\infty \leqslant R}} \int_{\hat{\mu}^N_{\{U_i D_i U_i^*\}_{1\leqslant i\leqslant m}} \in \Gamma_R(\mu,\varepsilon,k)} e^{N\,\mathrm{Tr}(V)}\, dU_1 \cdots dU_m
\end{aligned}$$

for any $\delta < \varepsilon$. Now, choosing δ and using the continuity of $\hat{\mu}^N_{\{U_i D_i U_i^*\}_{1\leqslant i\leqslant m}}$ in the distribution of the uniformly bounded variables D_i, we find by Corollary 3.1 and our hypothesis that

$$\liminf_{N\to\infty} \frac{\int_{\hat{\mu}^N_{U_i D_i U_i^*, 1\leqslant i\leqslant m} \in \Gamma_R(\mu,\varepsilon,k)} e^{N\,\mathrm{Tr}(V)} dU_1 \cdots dU_m}{\int e^{N\,\mathrm{Tr}(V)} dU_1 \cdots dU_m} = 1$$

which insures that

$$\chi(\mu) \geqslant \sum_{i=1}^m I(\mu_i) - \mu(V) + F(V,\mu_i).$$

Thus we have proved that

$$\chi(\mu) = \sum_{i=1}^m I(\mu_i) - \mu(V) + F(V,\mu_i).$$

Note that $\mu(V)$ and $F(V,\mu_i)$ can be written in terms of the $\mu^{\mathbf{k}}$ of Theorem 4.2 by Theorem 7.1. □

9. Generalization to integrals over the orthogonal group

In a recent article [34], Zuber shows that the large N asymptotics of two matrix integrals (the integral with external magnetic field and the Harish-Chandra–Itzykson–Zuber integral) enjoy a universality property in the sense that they are the same (up to a proper rescaling) if we integrate over the unitary or the orthogonal group. This property was also obtained (but not explicitly stated) in the case of Harish-Chandra–Itzykson–Zuber integral in [18] where the rate functions for the large deviation principle for the law of the spectral measure process of the Hermitian and the symmetric Brownian motion were shown to differ only by a factor two. The Harish-Chandra–Itzykson–Zuber integral is rather special in the family of angular integrals and we can compute many interesting related quantities, regardless of the group on which integration is taken (see [4,13]).

In this section, we generalize this universality property by relating the large N limit of any small parameter integrals over the orthogonal group with its complex analogue.

Let us define

$$I_N^1\bigl(V, A_i^N\bigr) := \int\limits_{\mathcal{O}_N^m} e^{N\,\mathrm{Tr}(V(O_i, O_i^*, A_i^N, 1\leqslant i\leqslant m))}\, dO_1\cdots dO_m \tag{19}$$

where $(A_i^N, 1\leqslant i\leqslant m)$ are $N\times N$ deterministic symmetric uniformly bounded matrices, dO denotes the Haar measure on the orthogonal group $\mathcal{O}_N$ (normalized so that $\int_{\mathcal{O}_N} dO = 1$). In this section we will assume that V is a non-commutative polynomial in the O_i, O_i^*, A_i^N with *real* coefficients. Here, $O^* = O^t$ is the standard involution $O_{ij}^* = O_{ji}$. Observe that if P is a polynomial, $P(O_i, O_i^*, A_i^N, 1\leqslant i\leqslant m)^t = P^*(O_i, O_i^*, A_i^N, 1\leqslant i\leqslant m)$ so that we keep also the notation P^*.

We then claim that we have the following analogue of Theorem 7.1, which shows that the first order of integrals over the orthogonal group is the same as on the unitary group (up to proper renormalizations);

Theorem 9.1. *There exists $\varepsilon = \varepsilon(q_1,\ldots,q_n)$ so that for any $\mathbf{t}\in\mathbb{R}^n\cap B(0,\varepsilon)$ such that $V = V^* = \sum t_i q_i$, if we define*

$$F_{V,\tau}^1 := \lim_{N\to\infty}\frac{1}{N^2}\log I_N^1\bigl(V_{\mathbf{t}}, A_i^N\bigr)$$

then $F_{V,\tau}^1$ exists and

$$F^1_{\frac{1}{2}V,\tau} = \frac{1}{2}\sum_{\mathbf{k}\in\mathbb{N}^n\setminus(0,\ldots,0)}\prod_{1\leqslant i\leqslant n}\frac{t_i^{k_i}}{k_i!}\mathbb{M}_{\mathbf{k}}(q_1,\ldots,q_n,\tau) = \frac{1}{2}F_{V,\tau}.$$

Moreover,

$$\mathbb{M}_{\mathbf{k}}(q_1,\ldots,q_n,\tau) = \sum_{m \text{ admissible maps with } k_i \text{ stars } q_i} M_m(\tau)$$

is the weighted sum of maps constructed above k_i stars of type q_i for all i, after choosing one of them as a root star.

The proof is based on the fact that if $\mu_V^{N,1}$ denotes the law on $\mathcal{O}_N^m$ given by

$$\mu_{\frac{1}{2}V}^{N,1}(dO_1,\ldots,dO_m) := \frac{1}{I_N^1(\frac{1}{2}V, A_i^N)}e^{\frac{N}{2}\mathrm{Tr}(V(O_i,O_i^*,A_i^N,1\leqslant i\leqslant m))}\,dO_1\cdots dO_m$$

and $\hat{\mu}^N$ is the empirical distribution of $(O_i, O_i^*, A_i, 1\leqslant i\leqslant m)$, then we have the analogue of Corollary 3.1.

Theorem 9.2. *Assume that $V = \sum t_i q_i$ is self-adjoint. Let D be an integer and τ a tracial state in $\mathcal{M}|_{(A_i)_{1\leqslant i\leqslant m}}$. There exists $\varepsilon = \varepsilon(D, m) > 0$ such that if $|t_i| \leqslant \varepsilon$, $\hat{\mu}^N$ converges almost surely under $\mu^{N,1}_{\frac{1}{2}V}$ to the unique solution $\mu_{\mathbf{t}}$ of the Schwinger–Dyson equation* $\mathbf{SD}[V, \tau]$. *Moreover, $\bar{\mu}^{N,1}_{\frac{1}{2}V} = \mu^{N,1}_{\frac{1}{2}V}(\hat{\mu}^N)$ converges as well to this solution as N goes to infinity.*

In fact, since then we know that $\mu_{\mathbf{t}}(P)$ expands as a generating function of the $\mathbb{M}_{\mathbf{k}}(q_1, \ldots, q_n, \tau)$'s, Theorem 9.1 follows readily since for any $\alpha \in [0, 1]$,

$$\partial_\alpha \frac{1}{N^2} \log I^1_N\left(\frac{\alpha}{2} V_{\mathbf{t}}, A^N_i\right) = \frac{1}{2} \bar{\mu}^{N,1}_{\frac{1}{2}V}(V)$$

converges towards $\frac{1}{2}\mu_{\mathbf{t}}(V)$.

Proof of Theorem 9.2. The proof follows the same lines as the proof of Theorem 2.1; we make the change of variables $\mathbf{O} = (O_1, \ldots, O_m) \in \mathcal{O}^m_N \to \Psi(\mathbf{O}) = (\Psi_1(\mathbf{O}), \ldots, \Psi_m(\mathbf{O})) \in \mathcal{O}^m_N$ with

$$\Psi_j(\mathbf{O}) = O_j e^{\frac{\lambda}{N} P_j(\mathbf{O})}$$

where the P_j are antisymmetric polynomials (i.e. $P_j^* = -P_j$). The only change is that now $P_j(\mathbf{O})$ are matrices with real coefficients and the differentials hold in the direction of $\mathcal{A}^1_N$ which are the antisymmetric matrices with real coefficients. For N large enough, Ψ is a diffeomorphism; it is as in the complex case a local diffeomorphism which is injective. As such, its image is open and compact. $\mathcal{O}^m_N$ is not connected but the union of copies of $SO^\varepsilon(N) = \{O \in \mathcal{O}_N; \det(O) = +\varepsilon\}$, $\varepsilon = +1$ or -1. Since $\det(\Psi_j(\mathbf{O})) = \det(O_j)\det(e^{\frac{\lambda}{N} P_j(\mathbf{O})}) = \det(O_j)$, Ψ maps $SO^{\varepsilon_1}(N) \times SO^{\varepsilon_2}(N) \times \cdots \times SO^{\varepsilon_m}(N)$ into itself for each choice of $\varepsilon_i \in \{1, -1\}$. Therefore, by connectedness of this set, $\Psi(SO^{\varepsilon_1}(N) \times \cdots \times SO^{\varepsilon_m}(N))$ is open and closed and therefore equals $SO^{\varepsilon_1}(N) \times SO^{\varepsilon_2}(N) \times \cdots \times SO^{\varepsilon_m}(N)$. Thus, Ψ is a diffeomorphism of $\mathcal{O}^m_N$. Like in the proof of Lemma 2.1, we need to compute the Jacobian of this change of variable. The same arguments apply to show that

$$\left|\det J_\Psi(\mathbf{O})\right| = \exp\left(\frac{\lambda}{N} \operatorname{Tr} \tilde{\Phi} + O(1)\right)$$

with $\tilde{\Phi}$ the linear operator defined on antisymmetric matrices by

$$\tilde{\Phi}.A = \sum_i \partial_i P_i \sharp A.$$

A basis of $\mathcal{A}^1_N$ is given, for $k < l$, by

$$E^1(kl)_{rj} = \frac{1_{r=k,j=l} - 1_{r=l,j=k}}{\sqrt{2}}.$$

Therefore, the trace of any linear endomorphism φ on $\mathcal{A}^1_N$ defined by $\varphi(X) = \sum_\ell A_\ell X B_\ell$, for uniformly bounded matrices A_ℓ, B_ℓ, is now given by

$$\begin{aligned}
\mathrm{Tr}(\varphi) &= \sum_{k<l} \mathrm{Tr}\big(E^1(kl)^* \varphi\big(E^1(kl)\big)\big) = \frac{1}{2} \sum_{\ell} \bigg(\sum_{k \neq l} A^{\ell}_{ll} B^{\ell}_{kk} - \sum_{k \neq l} A^{\ell}_{lk} B^{\ell}_{lk} \bigg) \\
&= \frac{1}{2} \sum_{\ell} \mathrm{Tr}\big(A^{\ell}\big) \mathrm{Tr}\big(B^{\ell}\big) + \mathrm{Tr}\big(A_{\ell} B^{t}_{\ell}\big) \\
&= \frac{1}{2} \sum_{\ell} \mathrm{Tr}\big(A^{\ell}\big) \mathrm{Tr}\big(B^{\ell}\big) + N O(1)
\end{aligned}$$

since the operator norm of A_ℓ and B_ℓ is uniformly bounded, $O(1)$ is uniformly bounded in N.

We can apply this bound to our case where A_ℓ and B_ℓ are given by $\partial_i P_i =: \sum_\ell A_\ell \otimes B_\ell$. The A_ℓ and B_ℓ's are uniformly bounded since the O_j's and the A_j's are and non-zero for a finite number of ℓ's, thus we deduce that

$$\big|\det J_\Psi(\mathbf{O})\big| = \exp\Bigg(\frac{\lambda}{2N} \sum_{i=1}^{m} \mathrm{Tr} \otimes \mathrm{Tr}(\partial_i P_i) + O(1) \Bigg)$$

with $O(1)$ bounded uniformly in N. Since $O(1)$ is uniformly bounded, we can now proceed exactly as in the proof of Theorem 2.1 to show that for any $r \in \{1, \ldots, m\}$,

$$\lim_{N\to\infty} \bigg\{ \frac{1}{2} \hat{\mu}^N \otimes \hat{\mu}^N(\partial_r P) + \frac{1}{2N} \hat{\mu}^N(D_r V P) \bigg\} = 0, \quad \mu^{N,1}_{\frac{1}{2}V} \text{ a.s.}$$

As a consequence, for any limit point τ of $\hat{\mu}^N$, any antisymmetric polynomial P,

$$\tau \otimes \tau(\partial_r P) + \tau(D_r V P) = 0. \tag{20}$$

If P is symmetric, we claim that for any $r \in \{1, \ldots, m\}$,

$$\tau \otimes \tau(\partial_r P) = \tau(D_r V P) = 0 \tag{21}$$

so that (20) still holds. Indeed, if Q is a word in the $(O_i, O_i^*, A_i, 1 \leqslant i \leqslant m)$,

$$\begin{aligned}
\partial_r Q &= \sum_{Q = Q_1 O_r Q_2} Q_1 O_r \otimes Q_2 - \sum_{Q = Q_1 O_r^* Q_2} Q_1 \otimes O_r^* Q_2, \\
\partial_r Q^* &= \sum_{Q^* = Q_1 O_r Q_2} Q_1 O_r \otimes Q_2 - \sum_{Q^* = Q_1 O_r^* Q_2} Q_1 \otimes O_r^* Q_2 \\
&= \sum_{Q = Q_2^* O_r^* Q_1^*} Q_1 O_r \otimes Q_2 - \sum_{Q = Q_2^* O_r Q_1^*} Q_1 \otimes O_r^* Q_2 \\
&= \sum_{Q = Q_1 O_r^* Q_2} (O_r^* Q_2)^* \otimes Q_1^* - \sum_{Q = Q_1 O_r Q_2} Q_2^* \otimes (Q_1 O_r)^*.
\end{aligned}$$

Since the trace is invariant under transposition, we deduce that for all P, $\hat{\mu}^N(P^*) = \hat{\mu}^N(P)$ and thus,

$$\hat{\mu}^N \otimes \hat{\mu}^N\big(\partial_r Q + \partial_r Q^*\big) = 0. \tag{22}$$

With the same method, we can deal with the cyclic derivative term. Indeed, since $D_r(Q^*) = -(D_r Q)^*$, if we write $V = Q + Q^*$, we obtain:

$$\begin{aligned}\hat{\mu}^N\big(D_r V\big(P+P^*\big)\big) &= \hat{\mu}^N\big(D_r\big(Q+Q^*\big)\big(P+P^*\big)\big)\\ &= \hat{\mu}^N\big(D_r Q\big(P+P^*\big)\big) - \hat{\mu}^N\big((D_r Q)^*\big(P+P^*\big)\big)\\ &= \hat{\mu}^N\big(D_r Q\big(P+P^*\big)\big) - \hat{\mu}^N\big(\big(P+P^*\big)D_r Q\big) = 0.\end{aligned}$$

To sum up,

$$\hat{\mu}^N \otimes \hat{\mu}^N(\partial_r P) = \hat{\mu}^N(D_r V P) = 0$$

from which we get (21) by going to the limit. Since any polynomial P can be decomposed as the sum of a symmetric polynomial $(P + P^*/2)$ and an antisymmetric polynomial $(P - P^*/2)$, we conclude by linearity that (20) holds for any polynomial P. By uniqueness of the solutions to this equation for sufficiently small parameters t_i proved in Theorem 3.1, the proof is complete. □

Acknowledgments

The three authors would like to express gratitude to an anonymous referee for his careful reading of the preliminary version of our manuscript and for his very useful suggestions of improvements. We thank J.-B. Zuber for very fruitful discussions, which in particular led us to insert the last section of this article. During the Spring 2007 (during which the chapter 5 was worked out) Alice Guionnet visited the UC Berkeley Department of Mathematics. Her visit was supported in part by funds from NSF Grants DMS-0405778, DMS-0605166 and DMS-0079945. Benoît Collins' research was partly supported by an NSERC grant. Édouard Maurel-Segala visited Stanford University during 2006-007 and was supported by funds from NSF grant DMS-0244323.

References

[1] S. Albeverio, L. Pastur, M. Shcherbina, On the $1/n$ expansion for some unitary invariant ensembles of random matrices, Comm. Math. Phys. 224 (2001) 271–305.
[2] J. Ambjorn, L. Chekhov, C.F. Kristjansen, Yu. Makeenko, Matrix model calculations beyond the spherical limit, Nuclear Phys. B 404 (1993) 127–172.
[3] G. Ben Arous, A. Guionnet, Large deviations for Wigner's law and Voiculescu's non-commutative entropy, Probab. Theory Related Fields 108 (1997) 517–542.
[4] M. Bergère, B. Eynard, Some properties of angular integrals, arXiv:0805.4482, 2008.
[5] P. Biane, M. Capitaine, A. Guionnet, Large deviations bounds for the law of the trajectories of the Hermitian Brownian motion, Invent. Math. 152 (2003) 433–459.
[6] P. Biane, Logarithmic Sobolev inequalities, matrix models and free entropy, Acta Math. Sin. (Engl. Ser.) 19 (3) (2003) 497–506.
[7] E. Brézin, C. Itzykson, G. Parisi, J.B. Zuber, Planar diagrams, Comm. Math. Phys. 59 (1978) 35–51.
[8] B. Collins, Moments and cumulants of polynomial random variables on unitary groups, the Itzykson–Zuber integral, and free probability, Int. Math. Res. Not. 17 (2003) 953–982.
[9] B. Collins, K. Dykema, A linearization of Connes' embedding problem, New York J. Math. 14 (2008) 617–641.
[10] B. Collins, J. Mingo, P. Śniady, R. Speicher, Second order freeness and fluctuations of random matrices III. Higher order freeness and free cumulants, Doc. Math. 12 (2007) 1–70.
[11] B. Collins, P. Śniady, Integration with respect to the Haar measure on unitary, orthogonal and symplectic groups, Comm. Math. Phys. 264 (2006) 773–795.

[12] N.M. Ercolani, 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. 14 (2003) 755–820.
[13] A. Ferrer, B. Eynard, P. Di Francesco, J.-B. Zuber, Correlation functions of Harish-Chandra integrals over the orthogonal and the symplectic groups, J. Stat. Phys. 129 (2007) 885–935.
[14] A. Guionnet, First order asymptotics of matrix integrals; a rigorous approach towards the understanding of matrix models, Comm. Math. Phys. 244 (2004) 527–569.
[15] A. Guionnet, M. Maïda, A Fourier view on the R-transform and related asymptotics of spherical integrals, J. Funct. Anal. 222 (2005) 435–490.
[16] A. Guionnet, E. Maurel-Segala, Combinatorial aspects of matrix models, ALEA Lat. Am. J. Probab. Math. Stat. (2006) 241–279, electronic.
[17] A. Guionnet, E. Maurel-Segala, Second order asymptotics for matrix models, Ann. Probab. 35 (2007) 2160–2212.
[18] A. Guionnet, O. Zeitouni, Large deviations asymptotics for spherical integrals, J. Funct. Anal. 188 (2002) 461–515.
[19] X. Harish-Chandra, Differential operators on a semisimple Lie algebra, Amer. J. Math. 79 (1957) 87–120.
[20] G. 't Hooft, A planar diagram theory for strong interactions, Nuclear Phys. B 72 (1974) 461–473.
[21] F. Hiai, D. Petz, The Semicircle Law, Free Random Variables and Entropy, Math. Surveys Monogr., vol. 77, American Mathematical Society, 2000.
[22] I. Klep, M. Schweighofer, Connes' embedding conjecture and sums of hermitian squares, Adv. Math. 217 (4) (2008) 1816–1837.
[23] E. Maurel-Segala, High order expansion for matrix models and enumeration of maps, arXiv: http://front.math.ucdavis.edu/0608.5192, 2006.
[24] J. Mingo, P. Śniady, R. Speicher, Second order freeness and fluctuations of random matrices: II. Unitary random matrices, Adv. Math. 209 (2007) 212–240.
[25] J. Mingo, P. Śniady, R. Speicher, E. Tan, Second order cumulants of products, preprint, arXiv:0708.0586, 2007.
[26] Alexandru Nica, Roland Speicher, Lectures on the Combinatorics of Free Probability, London Math. Soc. Lecture Note Ser., vol. 335, 2006.
[27] R. Speicher, Multiplicative functions on the lattice of noncrossing partitions and free convolution, Math. Ann. 298 (1994) 611–628.
[28] A. Zvonkin, Matrix integrals and map enumeration: An accessible introduction, Math. Comput. Modelling 26 (1997) 281–304.
[29] W.T. Tutte, On the enumeration of planar maps, Bull. Amer. Math. Soc. 74 (1968) 64–74.
[30] D.V. Voiculescu, Lectures on free probability theory, in: Lecture Notes in Math., vol. 1738, 2000, pp. 279–349.
[31] D. Voiculescu, A strengthened asymptotic freeness result for random matrices with applications to free entropy, Int. Math. Res. Not. 1 (1998) 41–63.
[32] D.V. 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) 101–166.
[33] D.V. Voiculescu, K.J. Dykema, A. Nica, Free Random Variables, American Mathematical Society, Providence, RI, 1992.
[34] J.B. Zuber, On the large N limit of matrix integrals over the orthogonal group, arXiv:0805.0315, 2008.
[35] P. Zinn Justin, J.B. Zuber, On some integrals over the $U(N)$ unitary group and their large N limit, J. Phys. A 36 (2003) 3173–3194.