Japan. J. Math. 11, 33–68 (2016)
DOI: 10.1007/s11537-016-1489-1

JAPANESE JOURNAL OF MATHEMATICS
SINCE 1924

Free analysis and random matrices*

Alice Guionnet**

Received: 2 December 2015 / Revised: 12 February 2016 / Accepted: 11 March 2016
Published online: 16 May 2016

Communicated by: Yasuyuki Kawahigashi

Abstract. We describe the Schwinger–Dyson equation related with the free difference quotient. Such an equation appears in different fields such as combinatorics (via the problem of the enumeration of planar maps), operator algebra (via the definition of a natural integration by parts in free probability), in classical probability (via random matrices or particles in repulsive interaction). In these lecture notes, we shall discuss when this equation uniquely defines the system and in such a case how it leads to deep properties of the solution. This analysis can be extended to systems which approximately satisfy these equations, such as random matrices or Coulomb gas interacting particle systems.

Keywords and phrases: random matrices, non-commutative measure, Schwinger–Dyson equation

Mathematics Subject Classification (2010): 15A52 (primary), 46L50, 46L54 (secondary)

1. Introduction

Dan-Virgil Voiculescu gave the Takagi Lectures [49] in 2007 on some aspects of free analysis. He emphasizes the role of the free difference quotient that he introduced in [47]. It is the natural derivation for non-commutative variables

* This article is based on the 14th Takagi Lectures that the author delivered at University of Tokyo on November 15 and 16, 2014.

** This work was partially supported by the NSF and Simons foundation.

A. Guionnet
Department of Mathematics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge MA, 02139-4307, USA
and
Centre national de la recherche scientifique, 3, rue Michel-Ange, 75794 Paris cedex 16, France
(e-mail: guionnet@math.mit.edu)

with the highest degree of non-commutativity, as introduced in free probability theory. The free difference quotient is as well the central object of these lectures notes, as we discuss the analysis of equations, called Schwinger–Dyson (or loop) equations, that it governs. These equations appear in different domains in combinatorics, operator algebras, or probability theory. In free probability, they play the role of an integration by parts for non-commutative laws. A classical analogue of such equations is given by

$$\int x f(x)\, d\gamma(x) = \int f'(x)\, d\gamma(x) \tag{1}$$

for a large set of test functions, for instance polynomial functions. If we add the condition $\int 1\, d\gamma(x) = 1$, we have a unique solution. For instance we can compute recursively the moments and see that they are given by the number of pair partitions. In fact this equation characterizes the Gaussian law and it is often used to show convergence to this law, for instance by Stein's method. We shall consider similar equations but in a non-commutative setting. They are often derived by integration by parts or, more generally, by using Gauge invariance and perturbations. As in the classical setting (1), the Schwinger–Dyson equations provide a bridge between combinatorics and analysis. Integration by parts can ultimately be used to uniquely define the density of the measure in classical probability theory. In non-commutative probability theory, there is no notion of density, and integration by parts is a legitimate replacement for that. However, Schwinger–Dyson equation does not in general specify the non-commutative law uniquely. Also, unlike in classical probability theory, it is not easy to say much about the properties of a law satisfying such a Schwinger–Dyson equation. In fact, in such a generality, the only properties of such a law that are known were proved by Y. Dabrowski [16]. He considers the von Neumann algebra associated with the solution by the Gelfand–Naimark–Segal construction (this is a non-commutative analogue of the space of bounded measurable functions). He showed the non-Gamma property (there are no non-trivial asymptotically central sequences of elements) and factoriality (the associated von Neumann algebra has a trivial center). Under a further technical assumption, Y. Dabrowski and A. Ioana [19] proved that there are no Cartan subalgebras. However, one could expect much more. For instance, in the classical case, if the density of a probability measure is smooth and strictly positive, one can build it as a smooth transport of the Lebesgue measure, hence showing that the associated C^*- and W^*-algebras are isomorphic. One of the goal of these notes is to develop techniques to study Schwinger–Dyson equation and its solutions, and eventually to investigate such questions. To be more precise, let ∂_{X_i} be the free difference quotient given, for any choice of indices $j_1, \ldots, j_k \in \{1, \ldots, d\}$, by

$$\partial_{X_i} X_{j_1} \cdots X_{j_k} = \sum_{j_\ell = i} X_{j_1} \cdots X_{j_{\ell-1}} \otimes X_{j_{\ell+1}} \cdots X_{j_k}.$$

Let μ be a linear functional on the set of polynomial functions in d non-commutative indeterminates. We say that μ satisfies the Schwinger–Dyson (or loop) equation if and only if

$$\mu \otimes \mu(\partial_{X_i} P) = \mu(PJ_i) \tag{2}$$

for any polynomial P and some given polynomial (or more general) functions $(J_i, 1 \leq i \leq d)$. This can be understood as an integration by parts, with the notable difference from the classical case, that the difference quotient takes values in the tensor product. In the classical case where ∂_{X_i} is the classical derivative, such an equation has a unique solution and $J_i = -\partial_{x_i} \log \frac{d\mu}{dx_i}$. Knowing $J_i, 1 \leq i \leq d$, defines uniquely μ and in fact gives all properties you might wish to know about μ. Even uniqueness of solutions to Schwinger–Dyson equation is not true in general in the non-commutative setting, and we shall first investigate this question. Then we shall study properties of these solutions and mainly prove that in the domain where we can prove uniqueness, we can also build transport maps between solutions. Namely assume that $J_i = D_{X_i} V$ with D_{X_i} the cyclic derivative:

$$D_{X_i} X_{j_1} \cdots X_{j_k} = \sum_{j_\ell = i} X_{j_{\ell+1}} \cdots X_{j_k} X_{j_1} \cdots X_{j_{\ell-1}}.$$

If $V = \sum X_i^2$, it is easy to see that there is a unique solution to the corresponding Schwinger–Dyson equation, called the law of d semi-circle laws, and denoted by σ^d. If V is a small perturbation of the quadratic case or even strictly convex, we can prove uniqueness of the solutions, denoted by μ_V. Then we can prove that there are smooth functions (namely absolutely converging series) $T_1, \ldots, T_d$ ant $T'_1, \ldots, T'_d$ such that for all polynomial P

$$\mu_V(P) = \sigma^d(P(T_1(X_1, \ldots, X_d), \ldots, T_d(X_1, \ldots, X_d))),$$
$$\sigma^d(P) = \mu_V(P(T'_1(X_1, \ldots, X_d), \ldots, T'_d(X_1, \ldots, X_d))).$$

Such a property allows to show that the associated C^*- and von Neumann algebras are isomorphic. It turns out that generating functions for planar maps also satisfy such Schwinger–Dyson equations. This is reminiscent to the characterization of the Gaussian law as the law whose moments count pair partitions. Ultimately, this allows to identify the moments of solutions with generating functions of planar maps, and eventually transport maps could be used to compute critical exponents of the latter. Moreover, this property has several other consequences: the associated C^*-algebras are projectionless, and in particular self-adjoint polynomials have connected support [28]. Furthermore, the associated von Neumann algebras and C^*-algebras have the so-called Haagerup property [28].

There are also domains in probability theory where the Schwinger–Dyson equation appears, but only at a large parameter (a dimension) limit. For instance, we can construct random models such that the Schwinger–Dyson equation is true, but only in average, or/and up to a small error. This is the case for matrix models which are laws on $N \times N$ random matrices where typically the linear functional given by the normalized trace $\frac{1}{N}\mathrm{Tr}$ satisfies

$$\mathbb{E}\Big[\frac{1}{N}\mathrm{Tr} \otimes \frac{1}{N}\mathrm{Tr}(\partial_{X_i} P)\Big] = \mathbb{E}\Big[\frac{1}{N}\mathrm{Tr}(P D_{X_i} V)\Big] \tag{3}$$

for all polynomial P and $i \in \{1, \ldots, d\}$. It turns out that one can prove in some cases that $\frac{1}{N}\mathrm{Tr}(P)$ self-averages (i.e, is close to its expectation) and therefore any limit point satisfies the Schwinger–Dyson equation (2). When uniqueness of the solutions to the latter holds, we deduce that it converges towards this limit. But in fact Schwinger–Dyson equation can be used even further to find a complete expansion for $\mathbb{E}[\mathrm{Tr}(P)]$ in terms of N^{-1}. Indeed, the large dimension behavior of such systems can be analyzed thanks to (3), by expanding the observables around their limit, in the spirit of Stein method or, more simply, perturbation theory. These are called topological expansion as in the case of random matrices, each coefficient of the series corresponds to a generating function for the enumeration of maps of a given genus. Moreover, the ideas related to transport maps can be developed in this setting as well, yielding approximate transport maps (that is functions which map one probability to the other, up to a small error). The latter allows to derive fine properties of the eigenvalues such as the universality of the fluctuations of the eigenvalues. To be more precise let us consider the so-called β-models given by the probability on $\mathbb{R}^N$:

$$dP^V_{N,\beta}(\lambda_1, \ldots, \lambda_N) = \frac{1}{Z_N} \prod_{i<j} |\lambda_i - \lambda_j|^\beta e^{-N \sum V(\lambda_i)} \prod d\lambda_i \,.$$

The case $\beta = 2$ corresponds to the random matrix model introduced above with $d = 1$: Equation (3) can be checked with $\frac{1}{N}\mathrm{Tr}$ replaced by the empirical measure of the λ_i's. The fluctuations of the variables $(\lambda_i)_{1 \le i \le N}$ in the case $V(x) = \beta x^2/2$ were studied first in the case $\beta = 2$ by Tracy and Widom [43, 44, 20] based on the integrable structure of the law, and much more recently for all β's by Ramírez, Rider, Virág [40]. It is shown that if we order the eigenvalues $\lambda_1 < \cdots < \lambda_N$, the largest eigenvalue fluctuates around its limit 2 like $N^{2/3}$ and the limiting law is denoted TW_β. Inside the bulk, the correlation functions converge vaguely towards the Sine_β distribution. This entails the convergence in average of the spacings distribution $N(\lambda_i - \lambda_{i-1})$ to the Gaudin distribution when $\beta = 2$ [42]. We can construct approximate transport maps from $P_N^{\beta x^2}$ to P_N^V for a large family of nice potentials V. Controlling the large dimension behavior of this transport map, we can prove that the fluctuations of the eigenvalues are like in the case $V = \beta x^2$ for any of these nice potentials. This kind of arguments can be generalized to several matrix models.

Hence, Schwinger–Dyson equations seem to be a common theme in several problems, where it can be used to analyze fine properties such as isomorphisms and universality. Uniqueness of solutions to Schwinger–Dyson equation allow to show that apparently different quantities such as generating functions for the enumeration of planar maps and asymptotics of matrix integrals are related, as both quantities satisfy these equations. In fact, topological expansions can also be retrieved from Schwinger–Dyson equations through the so-called topological recursion which are based on the free difference quotient, see e.g. the discussion in Sect. 4.1. These topological recursions seem to appear in many other domains, see e.g. the work of B. Eynard et al, see e.g. [21,7].

We will first describe domains where Schwinger–Dyson equation shows up. We start by relating it with the problem of enumerating planar maps. Subsequently, it represents integration by parts in the context of non-commutative variables. This can be extended to other settings such as planar algebras and/or type III factors [Sect. 2.8]. The fact that Schwinger–Dyson equation appears in seemingly different domains allows to connect them, moments of non-commutative variables being often described by combinatorial objects, namely planar maps. There are other domains in probability where such an equation appears, at least asymptotically. This is the case of random matrices, or particles systems in repulsive interaction. This allows to derive topological expansion (that is large dimension expansions of the observables) [Sect. 4.1] and universality (that is criteria to show that fluctuations of some observables do not depend much on the model) [Sect. 4.3 and 4.3.2] .

The idea developed in these lecture notes are inspired by Voiculescu's program to develop free probability as a generalization of classical probability for non-commutative variables, encapsulating the central notion of freeness which replaces independence. The analogy between these two domains allows to bring ideas from one to the other. In the intermediate level, random matrices are random objects which converge towards non-commutative variables: as such, understanding their limit allows to analyze them. Reciprocally, random matrices can be used to find out how to adapt concepts from classical probability to the free set-up. We here illustrate this point with the example of transport maps, which are developed in free probability and random matrix theory based on ideas going back to Monge and Ampère.

2. The free difference quotient and the Schwinger–Dyson equations

2.1. The free difference quotient

The free difference quotient ∂_X was defined by D.-V. Voiculescu as a natural derivation in the context of non-commutative variables. Let $X_1, \dots, X_d$ be non-commutative indeterminates. We will assume that the space of polynomials $\mathbb{C}\langle X_1, \dots, X_d\rangle$ is equipped with an involution $*$ and that the indeterminates $X_1, \dots, X_d$ are self-adjoint with respect to this involution; $X_i = X_i^*$, $1 \le i \le d$. Hence, for any complex number z, any $j_1, \dots, j_k \in \{1, \dots, d\}$

$$(zX_{j_1} \cdots X_{j_k})^* = \bar{z} X_{j_k} \cdots X_{j_1}.$$

The free difference quotient with respect to the variable X_i is then given, for any choice of indices $j_1, \dots, j_k \in \{1, \dots, d\}$ by

$$\partial_{X_i} X_{j_1} \cdots X_{j_k} = \sum_{j_\ell = i} X_{j_1} \cdots X_{j_{\ell-1}} \otimes X_{j_{\ell+1}} \cdots X_{j_k}.$$

In other words, ∂_{X_i} satisfies the Leibniz rule

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

as well as

$$\partial_{X_i} X_j = 1_{j=i} 1 \otimes 1.$$

The fact that the difference quotient takes values in the tensor product allows to keep track of the position of the variable which was differentiated. Moreover, such a definition is a natural extension of classical derivatives applied to matrices. Indeed, if $X_1^N, \dots, X_d^N$ are $N \times N$ Hermitian matrices, a straightforward computation shows that for all $i \in \{1, \dots, d\}, \ell, k, r, s \in \{1, \dots, N\}$,

$$\partial_{(X_i^N)_{k\ell}} (P(X_1^N, \dots, X_d^N))_{rs} = ((\partial_{X_i} P)(X_1^N, \dots, X_d^N))_{r\ell, ks}, \tag{5}$$

where $(A \otimes B)_{r\ell,ks} := A_{r\ell} B_{ks}$. Note that the free difference quotient is not the standard derivative even in the case where one considers only one variable as it is given by

$$\partial_X X^k = \sum_{\ell=0}^{k-1} X^\ell \otimes X^{k-\ell-1}.$$

Identifying the tensor product space with the space of functions of two variables x and y we have

$$\partial_X X^k(x, y) = \sum_{\ell=0}^{k-1} x^\ell y^{k-\ell-1} = \frac{x^k - y^k}{x - y}$$

and therefore for a general C^1 function we get

$$\partial_X f(x, y) = \frac{f(x) - f(y)}{x - y}. \tag{6}$$

Another derivative of interest is the cyclic derivative D_{X_i} given by

$$D_{X_i} P = (m \circ \partial_{X_i}) P, \quad m(A \otimes B) = BA.$$

Again this is a natural notion when applied to matrices as for all $i \in \{1, \ldots, d\}$, $k, \ell \in \{1, \ldots, N\}$

$$\partial_{(X_i^N)_{k\ell}} \mathrm{Tr}(P(X_1^N, \ldots, X_d^N)) = (D_{X_i} P)(X_1^N, \ldots, X_d^N)_{k\ell}.$$

2.2. Non-commutative laws

A non-commutative law is a linear form μ on $\mathbb{C}\langle X_1, \ldots, X_d\rangle$ with values in $\mathbb{C}$ which satisfies

1. The positivity property: for all polynomial $P \in \mathbb{C}\langle X_1, \ldots, X_d\rangle$
$$\mu(P^*) = \overline{\mu(P)}, \quad \mu(PP^*) \geq 0.$$
2. The mass condition
$$\mu(1) = 1.$$
3. The trace property: for all polynomials $P, Q \in \mathbb{C}\langle X_1, \ldots, X_d\rangle$
$$\mu(PQ) = \mu(QP).$$

Typical examples are again provided by matrices: if $X^N = (X_1^N, \ldots, X_d^N)$ are d $N \times N$ Hermitian matrices,

$$\tau_{X^N}(P) := \frac{1}{N} \mathrm{Tr}(P(X_1^N, \ldots, X_d^N))$$

is a non-commutative law, with $\mathrm{Tr}(A) = \sum_{i=1}^N A_{ii}$. We will call it the empirical non-commutative law of X^N. If X^N is random, $\mathbb{E}[\tau_{X^N}]$ is also a non-commutative law. The non-commutative laws that we shall consider will be bounded, that is that there exists $R < \infty$ such that for all $i_k \in \{1, \ldots, d\}$, all $\ell \in \mathbb{N}$,

$$|\mu(X_{i_1} \cdots X_{i_\ell})| \leq R^\ell.$$

By the Gelfand–Naimark–Segal construction, we can associate to μ a Hilbert space H, $\Omega \in H$, and $a_1, \ldots, a_d$ bounded linear operators on H so that for all polynomial P

$$\mu(P(X_1, \ldots, X_d)) = \langle \Omega, P(a_1, \ldots, a_d)\Omega\rangle_H.$$

The C^*-algebra associated to μ is also described by the C^*-algebra generated by $a_1, \ldots, a_d$ inside $B(H)$, the set of bounded linear operators on H. We can also associate to μ the von Neumann algebra generated by $a_1, \ldots, a_d$ as the weak closure of its C^*-algebra in $B(H)$.

2.3. Schwinger–Dyson equation

A linear form μ on $\mathbb{C}\langle X_1, \ldots, X_d\rangle$ satisfies the Schwinger–Dyson equation with conjugate variables $J = (J_1, \ldots, J_d)$ if and only if for any polynomial P, any $i \in \{1, \ldots, d\}$,

$$\mu \otimes \mu(\partial_{X_i} P) = \mu(J_i P)$$

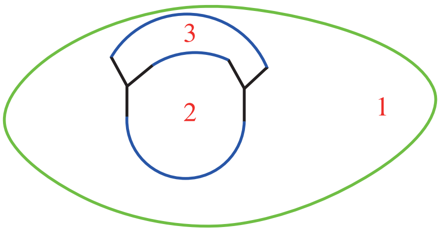
and $\mu(1) = 1$. This equation makes sense when the J_i are polynomials, but also in the C^*- or the von Neumann algebras associated with μ. We shall however mainly consider this equation when the conjugate variables are polynomials, or absolutely converging series.

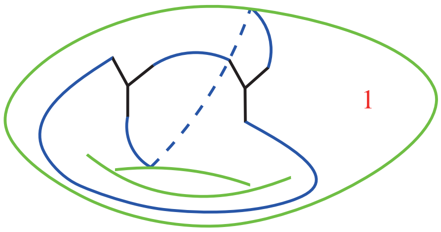
We next discuss settings where this equation appears, and start with the enumeration of planar maps.

2.4. Generating function for planar maps

We shall see here that generating functions for planar maps satisfy Schwinger–Dyson equation. We summarize below the enumeration problem; more details and discussions can be found for instance in [50]. A map is a connected graph which is properly embedded into a surface, that is so that its edges do not cross and the faces (obtained by cutting the surface along the edges of the graph) are homeomorphic to disks. The genus g of a map is the genus of this surface. It can be computed thanks to Euler formula

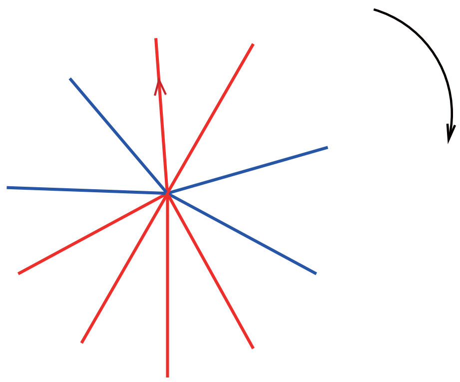
$$2 - 2g = \sharp\{vertices\} - \sharp\{edges\} + \sharp\{faces\}.$$



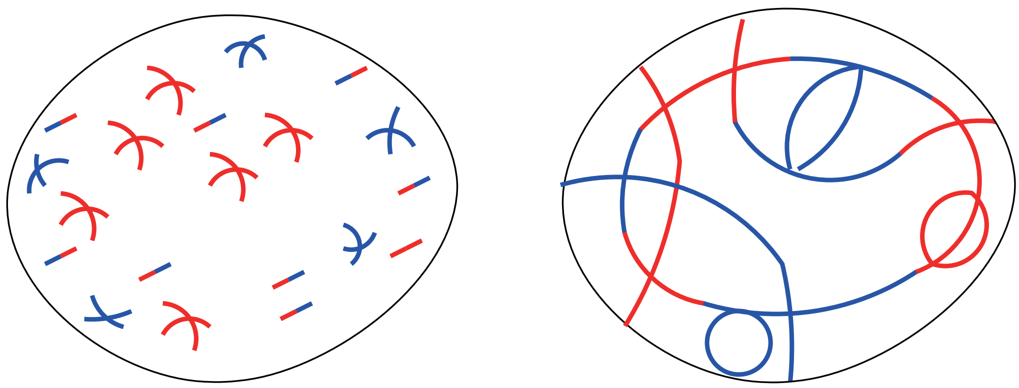


$2 - 2g = 2 - 3 + 3$ $\qquad$ $2 - 2g = 2 - 3 + 1$

We are interested in enumerating maps with a given genus and a given set of vertices. To do that, we shall consider rooted maps, that is maps equipped with a distinguished edge, called the root. This in particular means that we will count maps regardless of some symmetries. In fact, we will even count maps built over sets of vertices which are already drawn on the (orientable) surface and with prescribed degree. A nice way to do that is to start by drawing the vertices on the surface as follows: a vertex of degree p is drawn on an orientable surface together with p "half-edges" around it as well as a half-edge root. In other words, a vertex comes with labelled half-edges corresponding with the way it should be drawn on the surface. The question of enumerating maps then becomes the question of matching the end points of the half-edges. More generally, we will be interested by counting maps with colored edges. For instance, we would like to count maps with n red vertices of degree 4, m blue vertices of degree 4 and p edges between the blue and red vertices: this is called the Ising model on random graphs. To formulate this enumeration problem, we associate (bijectively) colored vertices to monomials in d indeterminates (if we are dealing with enumerating maps with d different colors) as follows. Let $q(X_1, \ldots, X_d) = X_{i_1} X_{i_2} \cdots X_{i_p}$ be a monomial. A "star of type q" is the vertex with first (the root) half-edge of color i_1, the second color i_2 etc until the last half edge which has color i_p.



We then let $M((q_i, k_i)_{1 \le i \le m}; g)$ denote the number of maps with genus g build on k_i stars of type q_i, $1 \le i \le m$.



Ising model on random graphs

Given monomials $q, q_1, \ldots, q_k$, complex variables $t_i \in \mathbb{C}, 1 \le i \le d$, and monomial q, we set

$$M_t(q) = \sum_{n_1, \ldots, n_k \ge 0} \prod \frac{t_i^{n_i}}{n_i!} M((q,1), (q_i, n_i)_{1 \le i \le k}; 0)$$

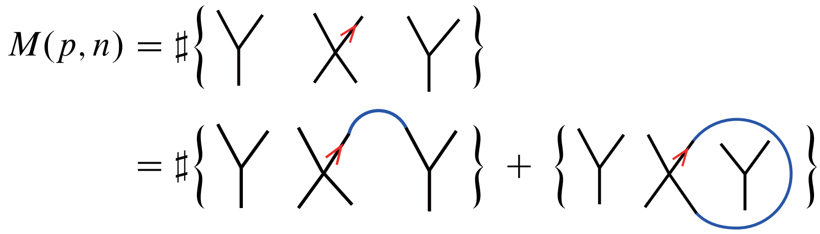
be the generating function for the enumeration of planar maps with one star of type q (which varies and play in combinatorics the role of an external face) and stars of type q_i, $1 \le i \le k$, coming from the model. This formal series can be seen to be absolutely convergent for $|t_i| \leq \epsilon$, $1 \leq i \leq d$, for ϵ small enough. For polynomials $P = \sum \alpha_i q_i$, $M_t(P)$ is defined by linearity by $M_t(P) = \sum \alpha_i M_t(q_i)$.

Theorem 1 (Tutte [46]). *For $t_1, \ldots, t_k$ small enough, M_t satisfies the Schwinger–Dyson equation: for all monomial q, all $i \in \{1, \ldots, d\}$,*

$$M_t \otimes M_t(\partial_{X_i} q) = M_t(q D_i V_t)$$

with $V_t = \frac{1}{2} \sum X_i^2 - \sum t_i q_i$.

The proof of this identity follows from the techniques of "surgery" introduced by Tutte. Let us illustrate it in the case of the enumeration of planar maps with vertices of degree 3. Let $M(p,n) = M((x^n, 1), (x^3, p); 0)$ be the number of planar maps with p vertices of degree 3 and one of degree n (again vertices should be thought as stars). Then, we can get an induction relation on $M(p,n)$ by considering what can happen to the root edge of the vertex of degree n (sometimes called the external face). Either it should be matched with an edge of another star (of degree 3) or to another edge of the vertex of degree n.



In both cases, erasing the matched edge, we arrive either at the situation where the external face has one edge extra, but one star of degree 3 disappeared, or the planar map is cut into two disjoint planar maps (where one contains a number $\ell \in [0, p]$ of vertices of degree three, and the other $p - \ell$). This yields the relation

$$M(p,n) = 3pM(p-1, n+1) + \sum_{k=0}^{n-2} \sum_{\ell=0}^{p} \binom{p}{\ell} M(\ell, k) M(p-\ell, n-k-2).$$

Here the combinatorial factors $3p$ and $\binom{p}{\ell}$ are due to the fact that we consider labelled edges; no symmetries are taken into account. Putting $M_t(x^n) := \sum_{p \geq 0} \frac{t^p}{p!} M(p,n)$, we deduce from the above that it satisfies the Schwinger–Dyson equation with potential $V_t = \frac{1}{2}x^2 - tx^3$:

$$M_t \otimes M_t(\partial_x x^n) = M_t(x^n(x - 3tx^2)).$$

2.5. *Non-commutative variables*

In the framework of free probability, a non-commutative law μ is said to have conjugate variables $J_i, 1 \leq i \leq d$, if and only if it satisfies the Schwinger–Dyson equation: for any polynomial P, any $i \in \{1, \ldots, d\}$

$$\mu \otimes \mu(\partial_{X_i} P) = \mu(P J_i).$$

Schwinger–Dyson equation can be seen as an integration by part formula: in fact we shall see that it can be retrieved from classical integration by parts when applied to random matrices. In the classical case where the difference quotient would be replaced by the classical derivative and μ simply be a probability measure, the J_i would simply be the derivative of the log-density. The case where J_i is a cyclic gradient of a polynomial corresponds to the Schwinger–Dyson equation satisfied by the generating functions for the enumeration of planar maps. In fact, it was shown by D.-V. Voiculescu [48] that if the J_i are polynomial functions they have to be cyclic gradient of a polynomial. Hence, in this case, the Schwinger–Dyson equations which are relevant are of the same type as those encountered in the enumeration of planar maps. However, in general, potentials need not be small perturbations of the quadratic potential.

2.6. *Approximate Schwinger–Dyson equations and large random matrices*

In this section, we show that large random matrices ensembles satisfy the Schwinger–Dyson equation in average. The first example is given by the Gaussian Unitary Ensemble (GUE) which is described by the $N \times N$ Hermitian matrices X^N with independent entries which are complex centered Gaussian variables with covariance N^{-1} above the diagonal and real centered Gaussian variables with covariance N^{-1} on the diagonal. In other words, the law of X^N is given by the law on $N \times N$ Hermitian matrices given by

$$d\mathbb{P}_N(X^N) = \frac{1}{Z^N} \exp\Big\{ -\frac{N}{2} \mathrm{Tr}((X^N)^2) \Big\} dX^N$$

where $dX_N = \prod_{i<j} d\Re(X_N(ij)) \prod_{i \leq j} d\Im(X_N(ij))$ is the Lebesgue measure on the entries and Z_N is the normalization constant. Take $X_1^N, \ldots, X_d^N$ be d independent GUE matrices. According to the Gaussian integration by parts

$$\begin{aligned} &\int (x+iy) f(x+iy, x-iy) e^{-N(x^2+y^2)}\, dx\, dy \\ =&\frac{1}{N} \int \partial_2 f(x+iy, x-iy) e^{-N(x^2+y^2)}\, dx\, dy, \end{aligned} \tag{7}$$

and by definition (5), one readily finds that

$$\begin{aligned}&\mathbb{E}_{\mathbb{P}_N^{\otimes d}}\Big(\frac{1}{N}\mathrm{Tr}(X_i^N P(X_1^N,\dots,X_d^N))\Big)\\ =&\mathbb{E}_{\mathbb{P}_N^{\otimes d}}\Big(\frac{1}{N}\mathrm{Tr}\otimes\frac{1}{N}\mathrm{Tr}((\partial_{X_i}P)(X_1^N,\dots,X_d^N))\Big).\end{aligned} \tag{8}$$

Here and after, $\mathbb{E}_P$ denotes the expectation under a probability measure P. Hence, independent GUE matrices satisfy the Schwinger–Dyson equation in average, for the quadratic potential $V = \frac{1}{2}\sum X_i^2$. We shall see later that as $\frac{1}{N}\mathrm{Tr}(P(X_1^N,\dots,X_d^N))$ self-averages, it converges as N goes to infinity and its limit satisfies the Schwinger–Dyson equation. This can be generalized to the case of random matrices in interaction. To this end take a polynomial V so that

$$d\mathbb{P}_N^V(X_1^N,\dots,X_d^N) = \frac{1}{Z_N^V}\exp\{-N\mathrm{Tr}(V(X_1^N,\dots,X_d^N))\}\,dX_1^N\cdots dX_d^N$$

makes sense (that is Z_N^V is well defined and finite). Then, again by integration by parts, one easily sees that the following averaged Schwinger–Dyson equation holds:

$$\mathbb{E}_{\mathbb{P}_N^V}\Big(\frac{1}{N}\mathrm{Tr}\otimes\frac{1}{N}\mathrm{Tr}(\partial_{X_i}P)\Big) = \mathbb{E}_{\mathbb{P}_N^V}\Big(\frac{1}{N}\mathrm{Tr}(D_{X_i}VP)\Big).$$

In the case where Z_N^V is infinite, one can add a cutoff so that the matrices keep an operator norm bounded (this however creates boundary terms in the Schwinger–Dyson equation).

2.7. Approximate Schwinger–Dyson equation and Coulomb gas interacting particle systems

Let us consider the following β-models:

$$dP_N^{V,\beta}(\lambda_1,\dots,\lambda_N) = \frac{1}{Z_N^V}\prod_{i<j}|\lambda_i-\lambda_j|^\beta e^{-N\sum V(\lambda_i)}\prod d\lambda_i.$$

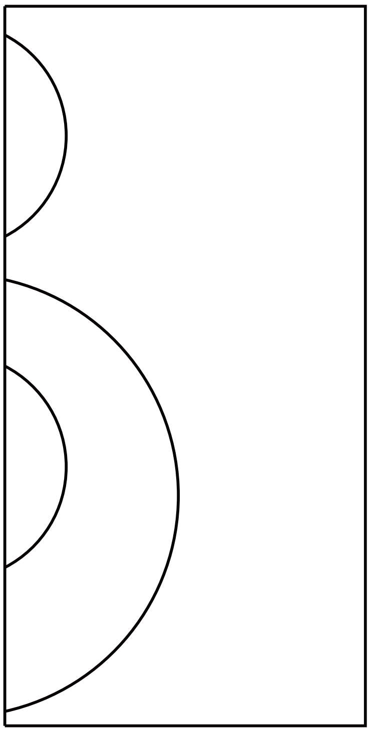
When $\beta = 2$, this corresponds to the joint law of the eigenvalues of the matrix models we just considered but with only one matrix. Then, the empirical measure $L_N = \frac{1}{N}\sum_{i=1}^N \delta_{\lambda_i}$ satisfies the approximate Schwinger–Dyson equation:

$$\begin{aligned}&\int\Big(\frac{\beta}{2}\iint \partial_X f(x,y)\,dL_N(x)\,dL_N(y) - \int V'(x)f(x)\,dL_N(x)\Big)\,dP_N^{V,\beta}\\ =&\frac{1}{N}\Big(\frac{\beta}{2}-1\Big)\iint f'(x)\,dL_N(x)\,dP_N^{V,\beta},\end{aligned} \tag{9}$$

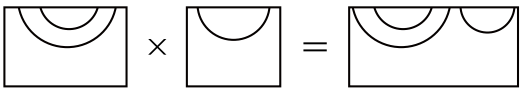
for any C^1 function f. Here, ∂_X was defined in (6). In this case, we see that an additional error term appears when we deal with the case $\beta \neq 2$: it will yield an expansion of the observable $\iint x^k\,dL_N(x)\,dP_N^{V,\beta}$ in $1/N$ instead of $1/N^2$.

2.8. Generalization of Schwinger–Dyson equations

It is possible to define analogous equations in other settings, for instance in models build on a structure of planar maps, such as planar algebras [24]. A simple example of planar algebra is given by Temperly–Lieb algebra. Temperley–Lieb elements are boxes containing an even number of boundary points, a starting point, and non-intersecting strings between these points:



These elements could as well be endowed with a shading: we do not discuss this point here. We can endow this set with the multiplication:

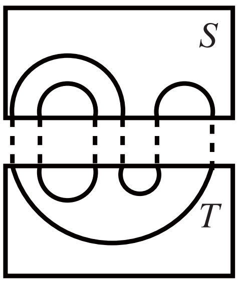


and the trace given by

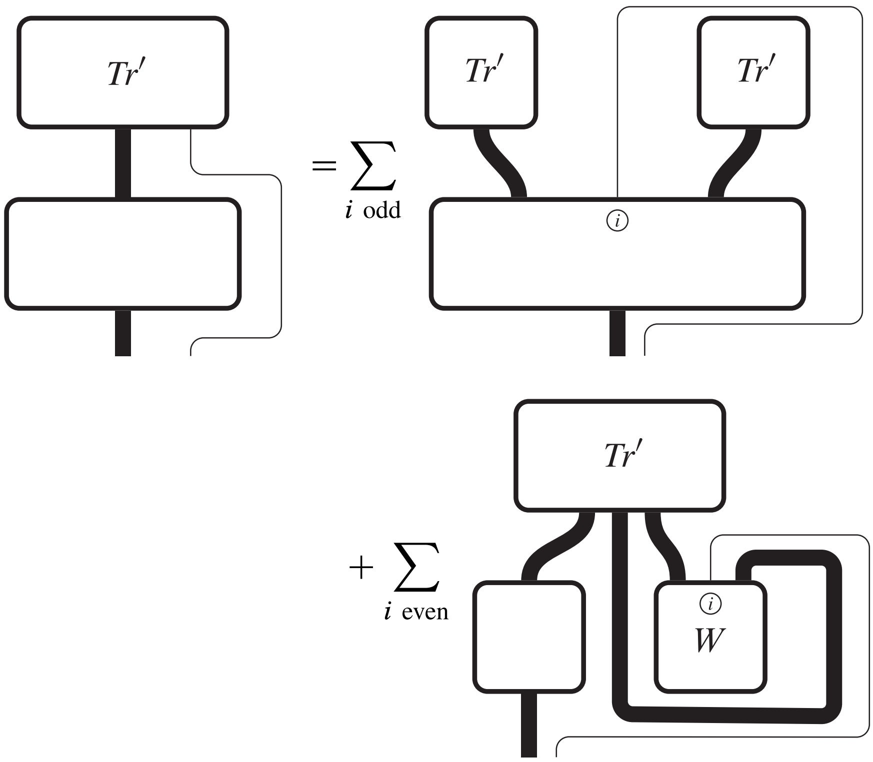
$$\tau(S) = \sum_{R \in \mathrm{TL}} \delta^{\sharp \text{loops in } S.R}$$

where the sum holds over Temperley diagrams with the same number of boundary points as S and

$T.S =$



In the above drawing, $T.S$ has two loops. For $\delta \in I := \{2\cos(\frac{\pi}{n})\}_{n\geq 4} \cup]2,\infty[$, it was shown [23] that τ is a tracial state, as a limit of matrix models. This can also be shown by combinatorial arguments [31]. Moreover, the von Neumann algebra associated to τ was shown to be a factor (and a tower with index δ^2 can be built). τ can also be constructed by matricial approximation. In [24], perturbations of such a law were introduced, allowing to replace the sum over Temperley–Lieb diagrams by a sum over planar diagrams constructed by matching the end points of copies of some Temperley–Lieb elements. This in particular includes the Potts model on random graphs. In these cases, the Schwinger–Dyson equation has a similar graphical interpretation that we gave for planar maps, and can be seen to be equivalent to the following diagrammatic equation:



(10)

Namely, expected values can be seen to amount to either split the diagrams into two disjoint pieces or to glue it with elements from the potential: this is very similar to what we saw for the generating functions of planar maps.

Schwinger–Dyson equation can also be derived for more general matrices, for instance unitary matrices following the Haar measure [15,27]. Let us for instance consider

$$\mathscr{P}_N^V(dU_1,\ldots,dU_m) = \frac{1}{I_N(V,A_i)} e^{N\mathrm{Tr}(V(U_i,U_i^*,A_i,1\leq i\leq m))}\, dU_1\cdots dU_m$$

for some polynomial V in $(U_i,U_i^*,A_i)_{1\leq i\leq m}$ and dU the Haar measure on the unitary group. Traces of polynomials also satisfy an averaged Schwinger–Dyson equation but the free difference quotient has to be modified. Let us first define the analogue to the free difference quotient given by the linear maps ∂_j such that for all $i, j \in \{1, \ldots, m\}$

$$\partial_j A_i = 0, \quad \partial_j U_i = 1_{i=j} U_j \otimes 1, \quad \partial_j U_i^* = -1_{i=j} 1 \otimes U_j^* \qquad (11)$$

and satisfying the Leibnitz rule (4). We let $D_j = m \circ \partial_j$ be the associated cyclic gradient (recall that $m(A \otimes B) = BA$). Using the invariance by multiplication of the Haar measure (which is analogous to using the invariance by translation of Lebesgue measure which led us to the integration by parts we used for Hermitian matrices) one can prove the Schwinger–Dyson equation:

$$\mathbb{E}_{\mathscr{P}_N^V}\Big(\frac{1}{N}\mathrm{Tr} \otimes \frac{1}{N}\mathrm{Tr}(\partial_j P)\Big) + \mathbb{E}_{\mathscr{P}_N^V}\Big(\frac{1}{N}\mathrm{Tr}(P D_j V)\Big) = 0\,. \qquad (12)$$

This is proved by noticing that if we set $U_j(t) = U_j e^{itB}$ and leave the other $U_k(t) = U_k$ unchanged for a Hermitian matrix B then for all $k, l \in \{1, \ldots, N\}$

$$\partial_t \int P(U_p(t), 1 \le p \le m, A_i)(k,l) e^{N\mathrm{Tr}(V(U_p(t), U_p(t)^*, A_i))} dU_1 \cdots dU_m = 0.$$

Taking B to be the matrix $B_{st} = 1_{st=kl} + 1_{st=lk}$ or $B_{st} = i1_{st=kl} - i1_{st=lk}$ shows that we can choose by linearity B to be the matrix $B_{st} = 1_{st=kl}$ even though this is not self-adjoint. This yields the result after summation over k and l.

It is also possible to define Schwinger–Dyson equations in case of non-tracial states [36], that is type III factors. Again the appropriate differential calculus has to be developed. Another extension concerns discrete analogue of the β-ensembles consisting in summing over λ_i taking values in a lattice instead of on the real line. Integration by parts give equations which are not easy to handle, but Nekrasov [35,34,33] defined non-perturbative Schwinger–Dyson equations which can be analyzed [6].

To sum up, the Schwinger–Dyson equations show up in many non-commutative or ordered systems. The free difference quotient appears as the natural derivation in this setting. We shall now analyze these equations.

3. Properties of solutions to Schwinger–Dyson equations

Even though the Schwinger–Dyson equation can be seen as an integration by part formula in free probability, it is not clear at all that it has a unique solution. In fact it does not in general. Indeed, even if $d = 1$, there may be several solutions to the Schwinger–Dyson equation. Indeed, we can choose V so that any solution will have the same disconnected support, for instance by taking V with two global minima and with large enough values in between these minima. Then, the Schwinger–Dyson equation will have a solution for any choice of mass given to these connected components. We start by investigating this question of uniqueness and then we study properties of the solutions when they are unique.

3.1. Uniqueness

Let μ be a linear form on $\mathbb{C}\langle X_1, \cdots, X_d\rangle$ so that

$$\mu \otimes \mu(\partial_{X_i} P) = \mu(D_{X_i} V P), \quad \forall P \in \mathbb{C}\langle X_1, \ldots, X_d\rangle, \quad \forall i \in \{1, \ldots, d\}. \tag{13}$$

Assume $\mu(1) = 1$. We investigate in the next subsection the cases where we know that a unique solution exists.

3.1.1. Quadratic case Note that when $V = \frac{1}{2}\sum X_i^2$, uniqueness is clear since then

$$\mu(X_i P) = \mu \otimes \mu(\partial_{X_i} P)$$

allows to define $\mu(X_{i_1} \cdots X_{i_k})$ for all choices of $i_1, \ldots, i_k \in \{1, \ldots, d\}$ by induction over k, starting from $\mu(I) = 1$. This solution is denoted by σ^d. It is the solution for the enumeration of maps with one star and therefore we see that $\sigma^d(X_{i_1} \cdots X_{i_k})$ corresponds to the number of planar maps that one can build over a star of type $X_{i_1} \cdots X_{i_k}$. Note that we can equivalently see a star as a circle decorated with colored dots one colored dot (with color i_1) and with colored dots drawn in the clockwise order of color $i_2, i_3, \ldots, i_k$, and that in case of a single star this is also equivalent to a line with first point with color i_1, second $i_2, \ldots$. In the later case, the planar map can be identified with a non-crossing pair partition blocks of points of the same color. The non-crossing property is simply that if (a, b) and (c, d) are two pairings then $a < b < c < d$ or $a < c < d < b$ and $i_a = i_b, i_c = i_d$.This interpretation of $\sigma^d(q)$ can be compared with that of Gaussian moments which are given by the enumeration of pair partitions. In fact, σ^d plays the role of the law of independent Gaussian variables in free probability. We will see that it also gives the limit of moments of GUE variables. Moreover, it encapsulates the notion of freeness which is central in free probability. In fact, σ^d can be defined as follows:

- if we restrict σ^d to polynomials in one variable, it is given by the semi-circle distribution, that is

$$\sigma^d(X_i^k) = \frac{1}{2\pi}\int_{-2}^{2} x^k \sqrt{4 - x^2}\, dx\,.$$

- whereas more general moments can be computed from the property that $X_1, \ldots, X_d$ are free under σ^d : we say that $X_1, \ldots, X_d$ are free under μ if and only if

$$\mu(q_1(X_{i_1})q_2(X_{i_2})\cdots q_\ell(X_{i_\ell})) = 0 \tag{14}$$

for any polynomials $q_k, 1 \le k \le \ell$, so that $\mu(q_k(X_{i_k}))$ vanishes and any indices $i_j \in \{1, \dots, d\}$ so that $i_k \neq i_{k+1}, 1 \le k \le \ell - 1$. It is not hard to see that (14) defines uniquely the law once the marginal distributions are known.

The equivalence of the two definitions of σ^d above, by planar maps or freeness, amounts to showing that any planar map built on a star of type q must leave at least one subset of connected one-color half-edges (corresponding to each of the monomials q_i) to having only self-matchings (this contribution then vanishes by centering). This can be checked by induction over the degree of q.

3.1.2. Perturbative case

Theorem 2 ([25]). *Let $R \in (2, +\infty)$ and $V = \frac{1}{2}\sum X_i^2 + \epsilon W$ for a self-adjoint polynomial $W = \sum_{i=1}^p t_i q_i$, for some monomials q_i and complex numbers t_i. Then for ϵ small enough (depending on R) there exists a unique solution μ_V to* (13) *so that $\mu(1) = 1$ and for any $k \in \mathbb{N}$, any $i_1, \dots, i_k \in \{1, \dots, d\}^k$*

$$|\mu(X_{i_1} \cdots X_{i_k})| \le R^k . \tag{15}$$

Moreover it is a tracial state: for any polynomials P, Q,

$$\mu(P^*) = \overline{\mu(P)}, \quad \mu(PP^*) \ge 0, \qquad \mu(PQ) = \mu(QP) .$$

The idea of the proof is perturbative, the situation $t_i \equiv 0$ being clear according to the previous section. Indeed, take two solutions $\tau, \tilde{\tau}$ and denote by

$$\Delta_k := \sup_{q:\deg(q) \le k} |\tau(q) - \tilde{\tau}(q)|$$

where the supremum is taken on monomials (or words) of degree smaller or equal to k. We have

$$\Delta_{k+1} = \max_i \sup_{q:\deg(q)=k} |\tau(X_i q) - \tilde{\tau}(X_i q)|$$

whereas by using (13), (15) and $\Delta_0 = 0$, if $D + 1 = \max_{1 \le i \le p} \deg(q_i)$ we get for any monomial q with degree k,

$$\begin{aligned} |\tau(X_i q) - \tilde{\tau}(X_i q)| &\le |\tau \otimes \tau(\partial_{X_i} q) - \tilde{\tau} \otimes \tilde{\tau}(\partial_{X_i} q)| + D\epsilon \sum_{j=1}^p |t_j| \Delta_{k+D-1} \\ &\le 2 \sum_{l=1}^{k-1} \Delta_l R^{k-1-l} + D\epsilon \sum_{j=1}^p |t_j| \Delta_{k+D-1}. \end{aligned}$$

Hence,

$$\Delta_{k+1} \le 2\sum_{l=1}^{k-1} \Delta_l R^{k-1-l} + D\epsilon \sum |t_j| \Delta_{k+D-1}, \qquad \Delta_k \le 2R^k$$

so that for $\gamma < R^{-1} \wedge 1$, we deduce that

$$\Delta_\gamma := \sum_{k\ge 1} \gamma^k \Delta_k \le \frac{2\gamma^2}{1-R\gamma}\Delta_\gamma + \frac{\epsilon D \sum |t_j|}{\gamma^{D-2}}\Delta_\gamma\,. \tag{16}$$

Hence, if ϵ is small enough so that there exists $\gamma \in (0, 1/R \wedge 1)$ such that

$$\frac{2\gamma^2}{1-R\gamma} + \frac{\epsilon D \sum |t_j|}{\gamma^{D-2}} < 1$$

we deduce that $\Delta_\gamma = 0$, which implies $\tau = \tilde{\tau}$.

3.1.3. Strictly convex case We need to find the right definition of strict convexity in the non-commutative setup. It should as in the classical case be related with the idea that a second order differential of the function is bounded below by a multiple of the identity. The difficulty here is that derivatives have values in the tensor product, and there are many possible topologies that can be implemented on the space of tensor product valued functions. We will use classes of differentiable functions with values in the Haagerup tensor product $M^{\otimes_{ehcD}2}$ of some von Neumann algebra M. We do not detail more precisely this space here (see e.g. [17]) but emphasize that it is nice, for instance to make sure that composition of smooth functions are smooth, or to get smoothness of natural operations that show up in the analysis.

Proposition 3. *The following are equivalent.*

1. *$A = A^* \in M_d(M^{\otimes_{ehcD}2})$ has a semigroup of contraction e^{-At} for all $t \ge 0$.*
2. *$A = A^* \in M_d(M^{\otimes_{ehcD}2})$ has a resolvent family for all $\alpha > 0$, $\alpha + A$ is invertible in $M_d(M^{\otimes_{ehcD}2})$ and $||\frac{\alpha}{\alpha+A}|| \le 1$.*

In this case we say $A \ge 0$.

Definition 4. *Let $c, R \in \mathbb{R}^{+,*}$. Let $V = V^*$ be a smooth function on non-commutative variables.*

V is said to be generalized (c, R)-convex if $(\partial_{X_i} D_{X_j} V)(X_1, \ldots, X_d) - c\,Id \ge 0$ for some $c > 0$ and all $X_1, \ldots, X_d$ self-adjoint variables bounded by R in the von Neumann algebra M.

We then can prove that

Theorem 5. *Let $R > 0$. Then there exists $c(R) > 0$ so that if V is (c, R)-convex for some $c \geq c(R)$, there exists a unique solution to the Schwinger–Dyson equation* (13).

The arguments to prove this theorem are more sophisticated than the previous ones as they appeal to dynamics. In fact, it is shown that the solutions to Schwinger–Dyson equation are stationary solutions of the free stochastic differential equation given by

$$X_t^i = X^i + S_t^i - \frac{1}{2}\int_0^t D_{X_i} V(X_s)\, ds \tag{17}$$

where $(S^1, S^2, \ldots, S^d)$ are d free Brownian motion. Such equations are well defined and discussed in [4,5]. Let us shortly describe them. The free Brownian motion S is constructed similarly to the Brownian motion by the requirements that

1. $S_0 = 0$,
2. For any $t_1 \leq t_2 \leq \cdots \leq t_r$,

   $$(t_r - t_{r-1})^{-1/2}(S_{t_r} - S_{t_{r-1}}), \ldots, (t_2 - t_1)^{-1/2}(S_{t_2} - S_{t_1}), t_1^{-1/2} S_{t_1}$$

   are free semi-circular variables with distribution σ^r.

$(S^1, S^2, \ldots, S^d)$ are free in the sense that their marginals satisfy (14). Existence and uniqueness of solutions to (17) can be proved by using standard Picard arguments (here we assume that V is a polynomial, but any non-commutative function satisfying a Lipschitz property would work) and the fact that solutions stay bounded uniformly in time by R if their initial condition has sufficiently small norm when the potential is (c, R)-convex, see [4]. Let τ be the non-commutative law of $X_t = (X_t^1, \ldots, X_t^d), t \geq 0$, with initial condition X. The free analogue of Itô's calculus implies that the distribution of X_t satisfies the differential equation

$$\begin{aligned}\tau(P(X_t)) =& \tau(P(X)) + \int_0^t \tau \otimes \tau\Big(\frac{1}{2}\sum_{i=1}^d \partial_{X_i} D_{X_i} P(X_s)\Big)\, ds \\ &- \frac{1}{2}\int_0^t \tau\Big(\sum_{i=1}^d D_{X_i} P(X_s) D_{X_i} V(X_s)\Big)\, ds\,.\end{aligned}$$

Hence, we see that the right hand side vanishes when the non-commutative distribution of X_s satisfies the Schwinger–Dyson equation. Moreover, the strict convexity of V allows to show, as in the classical case, that the law of X_s does not depend much on the initial condition X when s goes to infinity, hence implying uniqueness of the stationary measure. This allows to conclude.

Remark 6. Uniqueness of solutions fails in general, even in the case $d = 1$ as we already noticed.

3.2. Free transport and isomorphisms

In this section, we discuss the construction of transport maps between non-commutative laws. Let τ, μ be two non-commutative laws of d (resp. m) variables $X = (X_1, \dots, X_d)$ (resp. $Y = (Y_1, \dots, Y_m)$). We seek for "transport maps" $T = (T_1, \dots, T_m)$ and $T' = (T'_1, \dots, T'_d)$ of d (resp. m) variables so that for all polynomials P, Q

$$\tau(P(X)) = \mu(P(T_1(Y), \dots, T_d(Y))),$$
$$\mu(Q(Y)) = \tau(Q(T'_1(X), \dots, T'_m(X))).$$

In this case, we denote $\tau = T\sharp\mu$ and $\mu = T'\sharp\tau$. At this point, transport map could either be polynomials, or in the C^* algebra, or in the von Neumann algebra associated with the non-commutative laws by the GNS construction. In fact, the existence of such transport maps gives isomorphisms between these algebras (depending on the regularity of the transport maps). The classical analogue of such a question is, for given two probability measures μ and ν on $\mathbb{R}^d$ and $\mathbb{R}^m$ respectively, whether we can build a transport map from μ to ν, that is a measurable function $T : \mathbb{R}^d \to \mathbb{R}^m$ so that for all bounded continuous function f

$$\int f(T(x))\, d\mu(x) = \int f(x)\, d\nu(x)\,.$$

We denote $T\sharp\mu = \nu$. We can similarly seek for T' so that $T'\sharp\nu = \mu$. Von Neumann proved that, except for very degenerate cases (such as μ being a Dirac mass at a point), one can always build a transport map between two probability measures. For instance, one can always build transport maps from the Lebesgue measure on [0, 1] to any probability measure on $\mathbb{R}^m$ for any $m \geq 1$. In fact, this transport map may be quite rough, and no more than measurable.

In the non-commutative setting, this type of question is widely open. In fact, a result of Ozawa [38] shows that there is no non-commutative law analogous to Lebesgue measure in the sense that any non-commutative measure could be seen as its image by some transport map.

In this section, we want to discuss the existence of such transport map in very nice cases, in fact for solutions to Schwinger–Dyson equation. We will restrict ourselves to when this equation has a unique solution as seen in the previous section. So let V be a self-adjoint polynomial and consider the solution μ_V to the Schwinger–Dyson equation

$$\mu_V \otimes \mu_V\, (\partial_{X_i} P) = \mu_V(D_{X_i} VP)\,. \tag{18}$$

When we restrict ourselves to pertubative setting, a natural topology on the space of functions is given by the Banach norm

$$\|P\|_R = \sum |\langle q, P\rangle| R^{\deg(q)} \tag{19}$$

where the sum is over monomials q entering in the decomposition of P by $P = \sum\langle q, P\rangle q$ and $\deg(q)$ denotes the degree of the monomial q. We let $\mathscr{A}^R$ be the completion of the set of polynomials for $\|\cdot\|_R$.

Theorem 7 (G.–Shlyakhtenko [29]). *Let $A > A' > 4$ and $V = \frac{1}{2}\sum X_i^2 + \beta W$ self-adjoint with $\|W\|_A < \infty$. Then, for β small enough, there exist F^V, T^V transport maps in $\mathscr{A}^{A'}$ between μ_V and $\sigma^d = \mu_{\sum X_i^2/2}$ so that*

$$\mu_V = T^V \sharp \sigma^d, \quad \sigma^d = F^V \sharp \mu_V .$$

In particular the related C^* algebras and von Neumann algebras are isomorphic.

In a work in progress with Y. Dabrowski, we extend this result to the strictly convex case:

Theorem 8 (Dabrowski–G.-Shlyakhtenko [18]). *Let $R > 4$. Then, there exists $c(R) > 0$ finite so that for any (c, R)-convex function V, $c \geq c(R)$, the C^*- and W^*-algebras associated with the unique solution μ_V to* (18) *are isomorphic to that of d free semicircular variables.*

These results allow to prove new isomorphisms. Let us for instance consider q-Gaussian laws introduced by Bozejko and Speicher [4] as an interpolation between classical Gaussian variables and free semi-circular laws. The moments of d q-Gaussian variables are described by

$$\tau_{q,d}(X_{i_1} \cdots X_{i_p}) = \sum_{\pi} q^{i(\pi)}, \quad \forall i_k \in \{1, \cdots, d\}$$

where the sum is over pair partitions and $i(\pi)$ counts the number of crossings. Hence, $q = 1$ corresponds to commutative Gaussian variables whereas the case $q = 0$ corresponds to free semi-circular variables. It was shown by Y. Dabrowski that for qd small enough, $\tau_{q,d}$ satisfies the Schwinger–Dyson equation with a potential V which is a small perturbation of the quadratic potential in some $\mathscr{A}^R$, as in Theorem 7. Hence, we can apply our result to prove the following corollary.

Corollary 9 ([29]). *Let $d \in \mathbb{N}^*$. For qd small enough, there exist $R > 0$ finite and transport maps in $\mathscr{A}^R$ between $\tau_{q,d}$ and $\tau_{0,d} = \sigma^d$. In particular the C^*-algebra and von Neumann algebras of q-Gaussian laws, q small, are isomorphic to that of the free semicircle law σ^d.*

The ideas to construct such transport maps follow very classical ideas going back to Monge–Ampère, that we now describe in the classical probability setting. Let μ, ν be probability measures on $\mathbb{R}^d$ that have smooth densities

$$\mu(dx) = e^{-V(x)}\, dx, \quad \nu(dx) = e^{-W(x)}\, dx .$$

Then $T\sharp\mu = \nu$ is equivalent to

$$\int f(T(x))e^{-V(x)}\,dx = \int f(x)e^{-W(x)}\,dx$$
$$= \int f(T(y))e^{-W(T(y))}JT(y)\,dy$$

with JT the Jacobian of T. Hence, it is equivalent to the transport equation

$$V(x) = W(T(x)) - \log JT(x)\,.$$

Monge–Ampère equation amounts to taking T to be a gradient. Such an equation can not be solved in general, except in the case where $V - W$ is small. Indeed, then we can look for a solution T close to identity, in which case JT is close to one and therefore the equation does not touch the singularity of the logarithm. Hence, it can be solved by the implicit function theorem.

This type of arguments can be generalized to the non-commutative setting by showing that the Schwinger–Dyson equation is equivalent to a free Monge–Ampère equation. The latter takes the form

$$(1\otimes\tau+\tau\otimes 1)\mathrm{Tr}\log\mathscr{J}F = \mathscr{S}\Big[\Big\{\frac{1}{2}\sum F(X)_j^2 + W(F(X))\Big\} - \frac{1}{2}\sum X_j^2\Big]$$

where $\mathscr{S}$ is a certain symmetrization operator and $\mathscr{J}$ is a natural generalization of the notion of Jacobian based on the free difference quotient. Based on this formula, one can develop in the non-commutative setting arguments similar to those of the implicit function theorem.

B. Nelson showed that it is also possible to extend these ideas to more general setups; to construct a transport map in the case of type III factors [36] and in the case of finite depth sub factor planar algebras [37].

Such arguments do not work a priori to prove non-perturbative results and with Dabrowski and Shlyakhtenko, we used a follow up to this idea based on interpolating potentials. Let us first outline this idea in the classical set up. Then, again consider for potentials V, W properly renormalized, the probability measures

$$\mu_V(dx) = e^{-V(x)}\,dx\,,\quad \mu_W(dx) = e^{-W(x)}\,dx.$$

Consider for $V_t = (1-t)W + tV + \log\int e^{-(1-t)W(y)-tV(y)}\,dy$, the transport map $T_{0,t}$ from μ_W to μ_{V_t}. Let

$$\varphi_t = \partial_t T_{0,t}\circ T_{0,t}^{-1}. \tag{20}$$

Some algebra reveals that Monge–Ampère equation becomes

$$L_t\psi_t = W - V \tag{21}$$

if we assume that $\varphi_t = \nabla \psi_t$. Moreover L_t is the infinitesimal generator:

$$L_t = \Delta - \nabla V_t \cdot \nabla.$$

Hence, to construct a transport map it is enough to solve the Poisson equation (21) and then the transport equation (20) driven by $\nabla \psi_t$ to find $T_{0,t}$. Taking $t = 1$ provides a transport map from μ_W to μ_V. To solve the Poisson equation (21), one needs to invert L_t, that is find the Green function. Alternatively, one can consider the semi-group $P_s^t = e^{sL_t}$ and in the case when it converges fast enough to a unique invariant measure μ_t (for instance in the case when V_t is strictly convex), use that (21) is satisfied if

$$\psi_t(x) = -\int_0^\infty [P_s^t - \mu_{V_t}](W - V)(x)\, ds\,.$$

This scheme can be generalized to the free setting by using again free stochastic differential equations. Indeed, for a strictly convex interaction, if $X_{\cdot}^{Y,t}$ is the solution of the free SDE with potential V_t and initial condition Y, we can define

$$\psi_t(Y) = -\int_0^\infty (\tau[(V - W)(X_s^{Y,t})|Y] - \tau[(V - W)(X_\infty^{Y,t})])\, ds,$$

where above we have used conditional expectation with respect to the initial data. The above integral converges absolutely provided that we have a strictly convex interaction. The main point to finally construct the transport by the transport equation is to make sure that ψ_t described above is a sufficiently smooth function. This in fact requires to understand better the set up of free analysis and its natural topologies.

4. Models asymptotically driven by ∂_X

In this section we investigate models which are only asymptotically driven by the free difference quotient, such as random matrices or particles in Coulomb gas interaction. We show that the approximate Schwinger–Dyson equation then allows to get a topological expansion, that is an expansion of observables in terms of the dimension. The first order is given by the solution to the Schwinger–Dyson equation, and the corrections are computed recursively by linearizing the approximate Schwinger–Dyson equation around their limit.

This expansion is said to be topological because if the dimension is seen as a parameter, then the observables can be seen as generating functions for maps of all genus. This can be proved again because the equations are driven by the free difference quotient, yielding a combinatorial interpretation: each correction can be interpreted as some number of maps with a given genus. We also investigate transport maps in this setting and show that we can construct approximate transport maps to prove this time the universality of the local fluctuations of the eigenvalues. Hence, this section follows roughly the scheme we have developed in the previous section, but pushed to get corrections of the solutions of the approximate Schwinger–Dyson equation with respect to the solution of the limiting equation.

4.1. *Matrix models*

Let V be a self-adjoint polynomial in d non-commutative indeterminates and consider the matrix integral

$$d\mathbb{P}_N^V(X_1^N,\dots,X_d^N) = \frac{1}{Z_N^V} e^{-N\mathrm{Tr}(V(X_1^N,\dots,X_d^N))} \prod_i 1_{\|X_i^N\|\le R}\, dX_i^N, \quad (22)$$

where $\|X\|$ denotes the spectral radius of the matrix X. dX_i^N denotes the Lebesgue measure. We denote in short $\mathrm{Tr}(q)$ for $\mathrm{Tr}(q(X_1^N,\dots,X_d^N))$ and $d\mathbb{P}_N^V$ for $d\mathbb{P}_N^V(X_1^N,\dots,X_d^N)$. Then we first state the topological expansion for this matrix model.

Theorem 10 (G.–Maurel–Segala [25]). *Take* $R \in (2,+\infty)$ *and fix* $K \in \mathbb{N}$. *Let* q_i, $1 \le i \le p$, *be monomials. Assume* $V_t = V_t^* = \frac{1}{2}\sum X_i^2 - \sum_{i=1}^p t_i q_i$ *with* t_i *small enough. Then*

1.
$$\mathbb{E}_{\mathbb{P}_N^{V_t}}\Big[\frac{1}{N}\mathrm{Tr}(q)\Big] = \sum_{g=0}^K \frac{1}{N^{2g}}\tau_g^t(q) + o\Big(\frac{1}{N^{2K}}\Big)$$

with $\tau_g^t(q) = \sum_{k_i\ge 0}\prod \frac{t_i^{k_i}}{k_i!} M((q,1),(q_i,k_i)_{1\le i\le p};g)$, *where* $M((q,1),(q_i,k_i)_{1\le i\le p};g)$ *is the number of maps with genus* g *build over a star of type* q *and* k_i *stars of type* q_i, $1 \le i \le p$.

2.
$$\frac{1}{N^2}\log\frac{Z_{V_t}^N}{Z_{V_0}^N} = \sum_{g=0}^K \frac{1}{N^{2g}}F_g(t) + o\Big(\frac{1}{N^{2K}}\Big),$$

where $F_g(t) = \sum\prod \frac{t_i^{k_i}}{k_i!} M((q_i,k_i)_{1\le i\le p};g)$.

Note that the cutoff R is mainly required to make the partition function Z_V^N finite. In the case when V is strictly convex, it can be removed. The main tool to find these asymptotic expansions is again the Schwinger–Dyson equation which is a consequence of integration by parts and given by

$$\mathbb{E}_{\mathbb{P}_N^V}\Big(\frac{1}{N}\mathrm{Tr}(D_{X_i}(V)P)\Big) = \mathbb{E}_{\mathbb{P}_N^V}\Big(\frac{1}{N}\mathrm{Tr}\otimes\frac{1}{N}\mathrm{Tr}(\partial_{X_i}P)\Big). \quad (23)$$

Moreover, to study these equations we will need tools such as concentration of measure and Brascamp–Lieb inequality (see [1] and [22]). They imply that there exists a finite constant C (which only depends on c) so that for any monomial q of degree less than $\sqrt{N}$

$$\int \left|\frac{1}{N}\mathrm{Tr}(q)\right| d\mathbb{P}_N^V \le C^{\deg q}\,, \tag{24}$$

and

$$\int \left|\frac{1}{N}\mathrm{Tr}(q) - \int \frac{1}{N}\mathrm{Tr}(q)\, d\mathbb{P}_N^V\right|^2 d\mathbb{P}_N^V \le \frac{C^{\deg(q)}}{N^2}\,. \tag{25}$$

As a consequence of (24), the family $\{\int \frac{1}{N}\mathrm{Tr}(q)\, d\mathbb{P}_N^V, q\}$ indexed by monomials in non-commutative variables is tight. Any limit point $\{\tau(q), q\}$ can be extended by linearity to polynomials and then satisfies the Schwinger–Dyson equation

$$\mu_V(PD_{X_i}V) = \mu_V \otimes \mu_V(\partial_{X_i}P) \tag{26}$$

with $\mu_V(I) = 1$. Moreover, for any monomial q, we deduce from (24) that

$$|\mu_V(q)| \le C^{\deg(q)}\,. \tag{27}$$

But we know by Theorem 2 that such an equation has a unique solution provided that the t_i's are small enough. We then deduce readily that $\{\int \frac{1}{N}\mathrm{Tr}(q)\, d\mathbb{P}_N^V, q\}$ converges towards $\{\mu_V(q), q\}$ with μ_V the unique solution to the Schwinger–Dyson equation (26).

To derive the corrections to this limit, the first point is to prove an a priori rough estimate by showing that there exists a finite constant $C > 0$ so that for all t_i's small enough, all monomials q of degree less than $N^{1/2-\varepsilon}$ for $\varepsilon > 0$, we have

$$\left|\mathbb{E}_{\mathbb{P}_N^V}\left[\frac{1}{N}\mathrm{Tr}(q)\right] - \mu_V(q)\right| \le \frac{C^{\deg(q)}}{N^2}\,. \tag{28}$$

The proof elaborates on the ideas developed around (16) to prove uniqueness of the solution to Schwinger–Dyson equation and the concentration inequalities (25), see [26]. We next turn to the precise estimate of the asymptotics of $\overline{\delta}^N(P) = \mathbb{E}_{\mathbb{P}_N^V}[\mathrm{Tr}(P)] - N\mu_V(P)$. To this end, we shall introduce the following cumulants:

$$\begin{aligned} W_2^V(P,Q) &= \mathbb{E}_{\mathbb{P}_N^V}[(\mathrm{Tr}\,P - \mathbb{E}_{\mathbb{P}_N^V}\mathrm{Tr}\,P)(\mathrm{Tr}\,Q - \mathbb{E}_{\mathbb{P}_N^V}[\mathrm{Tr}\,Q])] \\ &= \partial_\varepsilon \mathbb{P}^{V-\epsilon N^{-1}Q}(\mathrm{Tr}\,P)|_{\varepsilon=0}, \\ W_3^V(P,Q,R) &= \partial_\varepsilon W_2^{V-\epsilon N^{-1}R}(P,Q)\,. \end{aligned}$$

Rewriting (23) reveals after some algebra:

$$\mathbb{E}_{\mathbb{P}_N^V}[\mathrm{Tr}(\Xi_i P)] = \frac{1}{N} W_2^V(\partial_{X_i} P) + \frac{1}{N} \overline{\delta}^N \otimes \overline{\delta}^N (\partial_{X_i} P)\,, \tag{29}$$

where

$$\Xi_i P = \partial_{X_i} V \sharp P - (\mu_V \otimes I + I \otimes \mu_V) \partial_{X_i} P\,.$$

By our a priori estimate (28) on $\overline{\delta}^N$ the last term in the right hand side of (29) is at most of order N^{-3}. Hence, to estimate the first order correction, we would like to estimate the asymptotics of W_2^V as well as "invert" Ξ_i. It turns out that even though Ξ_i is hardly invertible, a combination of the Ξ_i's is. Because Ξ can be modified by cyclic rearrangement since it is taken under the trace, we let Ξ be the operator on $\mathscr{A}_0 = \{P \in \mathbb{C}\langle X_1, \ldots, X_d\rangle : \mu_V(P) = 0\}$ given by

$$\Xi P = \sum_i (\partial_{X_i} P \sharp D_{X_i} V - (\mu_V \otimes I + I \otimes \mu_V) \partial_{X_i} D_{X_i} P)\,.$$

Then the image of Ξ lies in $\mathscr{A}_0$ by the Schwinger–Dyson equations. We see that in the case where $V = \sum X_i^2/2$, Ξ is the sum of the degree operator and an operator which lowers the degree. Indeed, we have for any monomial P

$$\Xi_0 P := \sum_i \partial_{X_i} P \# D_{X_i} V = \deg(P) P.$$

It can be inverted on the space of polynomials with no constant terms: indeed for any non-constant monomial P we have

$$\Xi_0^{-1}(P) = \frac{1}{\deg(P)} P.$$

Therefore, Ξ is invertible. When V is a small perturbation of the quadratic potential, it can be seen that Ξ is invertible on the space given by the closure of polynomials for the norm $\|\cdot\|_C$, see (19).

To estimate W_2^V, we obtain a new equation by replacing V by $V + \varepsilon W$ and identifying the linear terms in ε in (23). we find

$$\begin{aligned} W_2^V(P, W) =& \mathbb{E}_{\mathbb{P}_N^V}\Big[\frac{1}{N}\mathrm{Tr}\Big(\sum_i D_{X_i} \Xi^{-1} P D_{X_i} W\Big)\Big] \\ &+ N^{-1}\Big(W_3^V\Big(\sum_i \partial_{X_i} D_{X_i} \Xi^{-1} P, W\Big) \\ &+ (W_2^V \otimes \overline{\delta}^N + \overline{\delta}^N \otimes W_2^V)\Big(\sum_i \partial_{X_i} D_{X_i} \Xi^{-1} P, W\Big)\Big)\,. \end{aligned}$$

It turns out that the terms in W_3^V and W_2^V are bounded by concentration inequalities (25) whereas $\bar{\delta}^N$ is bounded by our previous rough estimate (28). Hence we conclude that

$$\lim_{N\to\infty} W_2^V(P,W) = \mu_V\Big(\sum_i D_{X_i}\,\Xi^{-1}P \times D_{X_i}W\Big) =: w_2(P,W)$$

and therefore plugging this back into (29) we deduce the first order correction is given by

$$\mathbb{E}_{\mathbb{P}_N^V}\Big[\frac{1}{N}\mathrm{Tr}(P)\Big] = \mu_V(P) + \frac{1}{N^2}w_2\Big[\sum_i \partial_{X_i} D_{X_i}(\Xi^{-1}P)\Big] + o(N^{-2})\,.$$

The next orders of the asymptotic expansion can be derived similarly, by considering a family of Schwinger–Dyson equations which are obtained by perturbating the first one with respect to small additional potentials. We refer the interested reader to [26,32] for full details.

Hence, we see that deriving the topological expansion boils down to

- Derive Schwinger–Dyson equations,
- Use concentration arguments to be able to separate elements on different scales,
- Invert a linear differential operator, which can be interpreted as the limit of the infinitesimal generators related with the Gibbs measures.

It turns out that Schwinger–Dyson equations can be established for many other models which are not directly related with Gaussian random matrices. It seems that a large family of Schwinger–Dyson equations give rise to topological expansions. We describe below the case of the β-ensembles and the integration over the unitary group.

4.2. *Topological expansion for β-matrix models*

The law of the eigenvalues of the GUE follows the distribution on $\mathbb{R}^N$

$$dP_N(\lambda) = \frac{1}{Z_N}\prod_{i<j}|\lambda_i-\lambda_j|^2 e^{-N\sum\lambda_i^2}\prod d\lambda_i$$

as can be checked by doing the change of variables associating to X its ordered eigenvalues and a parametrization of its eigenvectors. β-ensembles are the following generalization of this distribution:

$$dP_N^{V,\beta}(\lambda) = \frac{1}{Z_{N,\beta}^V}\prod_{i<j}|\lambda_i-\lambda_j|^\beta e^{-N\sum V(\lambda_i)}\prod d\lambda_i\,.$$

It is related with invariant matrix ensembles only in the cases $\beta = 1, 2, 4$ and *a priori* has no relations with Gaussian entries otherwise. However, it was proved in [8], see [13] for a formal proof, that β-ensembles have a large-N (also called topological) expansion which generalizes the expansion we discussed for GUE matrices in the previous section (restricted to $d = 1$). More precisely, assume that

Assumption 11. • $\lim_{|x|\mapsto\infty} \frac{V(x)}{2\beta \log|x|} > 1$,

- *V is analytic in a neighborhood of the real line,*
- *Consider*

$$I(\mu) = \iint \Big(\frac{V(x) + V(y)}{2} - \frac{\beta}{2} \log|x-y|\Big)\, d\mu(x)\, d\mu(y). \tag{30}$$

It is strictly convex and therefore has a unique minimizer μ_V which is characterized by the fact that the effective potential $V_{\text{eff}}(x) = V(x) - 2\beta \int \log |x-y|\, d\mu_V(y)$ is equal to a constant C on the support of μ_V and greater than C outside of the support. We assume that the support of μ_V is connected.

- *Assume also that the effective potential V_{eff} achieves its minimal value only on the support of μ_V.*
- *The density of μ_V vanishes like a square root at the boundary of the support.*

Then for any $z \in \mathbb{C}\backslash\mathbb{R}$, and $K \geq 0$

$$\int \frac{1}{N} \sum_{i=1}^{N} \frac{1}{z - \lambda_i}\, dP_N^{V,\beta}(\lambda) = \sum_{k=0}^{K} N^{-k}\, W^{V,k}(z) + o(N^{-K}), \tag{31}$$

where $o(N^{-K})$ is uniform on compacts. Moreover, we have

$$W^{V,k}(z) = \sum_{g=0}^{\lfloor k/2 \rfloor} \Big(\frac{\beta}{2}\Big)^{-g} \Big(1 - \frac{2}{\beta}\Big)^{k-2g+1} \mathscr{W}^{V;(g;k-2g+1)(z)}.$$

Note that the hypothesis that the support is connected is important since otherwise the result is not true in general. The proof of this expansion relies as well on the Schwinger–Dyson equations (9). As a consequence, one sees that the equilibrium measure μ_V satisfies the limiting Schwinger–Dyson equation

$$\frac{\beta}{2} \iint \frac{f(x) - f(y)}{x - y}\, d\mu_V(x)\, d\mu_V(y) = \int f(x) V'(x)\, d\mu_V(x)\,. \tag{32}$$

If V is a small perturbation of the quadratic potential one can develop arguments similar to those of the previous section to check that moments under μ_V are generating functions for planar maps.

In fact, the limiting equation (32) does not always have a unique solution as it is a weak characterization of the minimizers of (30), but it does as soon as the support of the limiting equation has a connected support. Indeed in this case we can use the theory of integral equations, and in particular that of the airfoil equation [45], to get uniqueness of the solution of Schwinger–Dyson equation and invertibility of the relevant linear operators. In any case, μ_V governs the first order of the expansion. To get the higher order terms in the expansion the idea is, as in the previous section, to write equations for all the cumulants

$$W_n^V(x_1,\ldots,x_n) = \partial_{\epsilon_1}\cdots\partial_{\epsilon_n}(\ln Z_{N,\beta}^{V-\frac{2}{\beta N}\sum_i \frac{\epsilon_i}{x_i-\bullet}})|_{\epsilon_i=0}$$

by differentiating the Schwinger–Dyson equation (9) with respect to the potential. As the linear differential operator Ξ appearing when one linearizes the Schwinger–Dyson equation around its limit is invertible (in the connected support case), the strategy developed in the previous section can be generalized.

In the case where the support of the limiting equilibrium measure has r connected components S_j, $1 \le j \le r$, $r \ge 2$, it turns out that the Schwinger–Dyson equation (32) has a unique solution given the mass of each connected components S_j, $1 \le j \le r$. Moreover, in this case the operator Ξ is as well invertible. Hence, we can use the previous scheme to expand in N the partition function

$$Z_{n_1,\ldots,n_r}^V = \int \prod_{s=1}^{r} \prod_{j=n_1+\cdots+n_{j-1}+1}^{n_1+\cdots+n_j} 1_{\lambda_j\in S_j^\delta} \prod_{i<j} |\lambda_i-\lambda_j|^\beta e^{-N\sum V(\lambda_i)} \prod d\lambda_i,$$

where $S_j^\delta = \{x : d(x, S_j) \le \delta\}$ for some $\delta > 0$. The expansion does not depend on δ because the eigenvalues will stay inside $\bigcup S_j$ with very large probability. This allows to retrieve [9] an expansion in N for the partition function where the number of eigenvalues in each S_j is randomly chosen:

$$Z_N^V = \sum_{n_1+\cdots+n_r=N} Z_{n_1,\ldots,n_r}^V$$

and for the correlators $W_n^V(x_1,\ldots,x_n)$. As the sum above is dominated by one set of filling fractions $n_1,\ldots,n_r$, these expansions contain a theta-function, representing the fact that the eigenvalues at the boundary of the support will tunnel from S_j to S_{j+1} for some $j \in \{1,\ldots,r-1\}$ following a discrete gaussian law (which depends on the values of $N\mu_V(S_j) - \lfloor N\mu_V(S_j)\rfloor$, $1 \le j \le r$, and therefore may only converge under subsequences). In particular, central limit theorem for linear statistics do not hold in general, see [9, Sect. 8.3] and also [39].

Finally, the previous considerations also generalize [10] to more general potentials of the form $N^{-k+1}\sum_{i_1,\ldots,i_k=1}^N V(\lambda_{i_1},\ldots,\lambda_{i_k})$. A more complicated generalization is to deal with models which interact via a potential which vanishes at the origin, but which is not homogenous as the Coulomb gas interaction. In a recent work with Borot and Kozlowski, we consider the case where the interaction is given by

$$\prod_{i<j} \sinh(\omega_1(\lambda_i - \lambda_j)) \sinh(\omega_2(\lambda_i - \lambda_j)) .$$

4.2.1. Topological expansion for the Haar measure on the unitary group In this section we shall consider the Haar measure dU on the unitary group, that is the unique measure on $U(N)$ which is invariant under left multiplication by unitary matrices and with mass one. We consider matrix integrals given by

$$I_N(V, A_i) = \int e^{N \mathrm{Tr}(V(A_i, U_i, U_i^*, 1 \le i \le m))} \, dU_1 \cdots dU_m ,$$

where $(A_i, 1 \le i \le m)$ are $N \times N$ deterministic uniformly bounded matrices, and V is a polynomial function in the non-commutative variables $(U_i, U_i^*, A_i, 1 \le i \le m)$. A well-known example is the Harich-Chandra–Itzykson–Zuber integral

$$HCIZ(A_1, A_2) = \int e^{N \mathrm{Tr}(A_1 U A_2 U^*)} \, dU .$$

For technical reasons, we assume that the polynomial V satisfies $\mathrm{Tr}(V(U_i, U_i^*, A_i, 1 \le i \le m)) \in \mathbb{R}$ for all $U_i \in U(N)$ and all Hermitian matrices A_i, $1 \le i \le m$ and $N \in \mathbb{N}$. Under those very general assumptions, the formal convergence of the integrals could already be deduced from [14]. The following theorem is a precise description of the results from [15,27] which gives an asymptotic expansion:

Theorem 12 ([15,27]). *Let $K \in \mathbb{N}$. Under the above hypotheses and if we further assume that the spectral radius of the $N \times N$ matrices $(A_i, 1 \le i \le m)$ is uniformly bounded (by say M), there exists $\epsilon = \epsilon(M, V) > 0$ so that for $z \in [-\epsilon, \epsilon]$, the expansion*

$$\frac{1}{N^2} \log I_N(zV, A_i) = \sum_{k=0}^{K} \frac{1}{N^{2g}} F_g^{zV} + o\Big(\frac{1}{N^{2K}}\Big)$$

holds. Moreover, $F_{V,\tau}(z)$ is an analytic function of $z \in \{z \in \mathbb{C} : |z| \le \epsilon\}$ which only depends on the empirical non-commutative law of the $(A_i)_{1 \le i \le m}$. Furthermore, if we let

$$\mathscr{P}_N^{zV}(dU_1, \ldots, dU_m) = \frac{1}{I_N(zV, A_i)} e^{zN \mathrm{Tr}(V(U_i, U_i^*, A_i, 1 \le i \le m))} \, dU_1 \cdots dU_m$$

for every polynomial P in $(U_i, U_i^*, A_i)_{1\le i\le m}$ we have the expansion

$$\int \frac{1}{N}\mathrm{Tr}(P((U_i, U_i^*, A_i)_{1\le i\le m}))\, d\mathscr{P}_N^{zV} = \sum_{k=0}^{K} \frac{1}{N^{2g}} \tau_g^{zV}(P) + o\Big(\frac{1}{N^{2K}}\Big).$$

A combinatorial interpretation of the limits F_0^{zV} and τ_0^{zV} in terms of planar maps was provided in [15], but is still open for higher orders $F_g^{zV}, \tau_g^{zV}, g \ge 1$.

The strategy of the proof of Theorem 12 is again to study the Schwinger–Dyson equations under the associated Gibbs measure $\mathcal{P}_N^V$, that is (12). By using concentration of measure, we know that for every polynomial P, $N^{-1}\mathrm{Tr}(P(U_i, U_i^*, A_i))$ is not far from its expectation and therefore we deduce that the limit points of these (bounded) quantities $\tau(P)$ satisfy the Schwinger–Dyson equation

$$\begin{aligned} &\tau \otimes \tau(\partial_j P) + z\tau(D_j VP) = 0\,, \\ &\tau(Q(A_1, \dots, A_m)) = \frac{1}{N}\mathrm{Tr}(Q(A_1, \dots, A_m)) =: \tau_A(Q), \end{aligned} \tag{33}$$

where ∂_j is the modified free difference quotient as given in (11) and Q is a polynomial in the deterministic matrices $A_1, \dots, A_m$ whereas P is a polynomial in $U_i, U_i^*, A_i, 1 \le i \le m$. Uniqueness of the solution to such an equation in the perturbative regime is done as in the Gaussian case; when $z = 0$ it is clear as it defines all moments recursively from the knowledge of the empirical non-commutative law τ_A and a perturbation argument shows this is still true for small parameters. Uniqueness of the solutions to (33) provides the convergence whereas the study of this solution shows that it expands as a generating series in the enumeration of some planar maps. Next orders can be derived by arguments similar to those developed for the GUE.

4.3. *Approximate transport maps and universality*

In this section, we discuss the generalization of transport maps to models which are approximately driven by the free difference quotient and show that this entails universality properties for the fluctuations of the eigenvalues.

4.3.1. β-models Let us again consider the β-models given by

$$dP_{N,\beta}^V(\lambda_1, \dots, \lambda_N) = \frac{1}{Z_N} \prod_{i<j} |\lambda_i - \lambda_j|^\beta e^{-N\sum V(\lambda_i)} \prod d\lambda_i$$

and recall that there is an equilibrium measure μ_V such that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} f(\lambda_i) = \int f(x)\, d\mu_V(x) \quad a.s.$$

Theorem 13 ([2]). *Assume V, W are C^{31} and satisfy Assumption 11. Then there exists $T_0 \in C^{19}(\mathbb{R}, \mathbb{R})$, $T_1^N = (T_1^{N,1}, \ldots, T_N^{N,N}) \in C^1(\mathbb{R}^N, \mathbb{R}^N)$ so that*

$$\left\| \left(T_0^{\otimes N} + \frac{T_1^N}{N} \right) \sharp P_N^V - P_N^W \right\|_{TV} \leq const. \sqrt{\frac{\log N}{N}} ,$$

where $\| \cdot \|_{TV}$ is the total variation norm. Moreover, there exists a positive finite constant C such that $\sup_{1 \leq k \leq N} \|T_1^{N,k}\|_{L^1(P_N^V)} \leq C \log N$ *and*

$$P_{N,\beta}^V \left(\sup_{k,k'} \frac{|T_1^{N,k}(\lambda) - T_1^{N,k'}(\lambda)|}{\sqrt{N} |\lambda_k - \lambda_{k'}|} \leq C \ \log N \right) \geq 1 - N^{N/C} . \qquad (34)$$

In this theorem T_0 is simply the optimal transport map from the equilibrium measures μ_V to μ_W; $T_0 \sharp \mu_V = \mu_W$. In particular T_0 is increasing. The proof of such a result follows the approach by Poisson equation developed in Sect. 3.2. The idea is to write these equations for β-models and find an approximate solution by linearizing these equations around the limit, and showing that the leftovers only produce a small error in the density, hence producing a small error in total variation. Theorem 13 entails universality of the fluctuations, namely that the local fluctuations of the eigenvalues under $P_{N,\beta}^V$ and $P_{N,\beta}^W$ are the same. Taking $V = \frac{1}{2} x^2$, it yields for instance the following theorem:

Theorem 14. *Let W be C^{31} and satisfying Assumption 11. Assume that the eigenvalues are ordered $\lambda_1 \leq \cdots \leq \lambda_N$. Then,*

1. *if the support of μ_V is given by $[a, b]$, there is a constant ρ so that $N^{2/3}$ $\rho(\lambda_N - a)$ converges in distribution to the Tracy–Widom law* TW_β. *A similar result holds at the left edge b.*
2. *Take $\varepsilon > 0$ and $i \in [\varepsilon N, (1 - \varepsilon) N]$ so that $i/N \simeq E$. Then, there exists a constant c_E so that $N c_E (\lambda_i - \lambda_{i-1})$ fluctuates as in the case where $W = \frac{x^2}{2}$, in particular following the Gaudin distribution if $\beta = 2$.*

Such a result was also proved in [11, 12] under weaker assumptions on V (but for $\beta \geq 1$) and by Scherbina [41] (in the bulk but including cases where the support of μ_V is not connected, an extension that was also considered in [3] by using approximate transport maps).

It is quite clear that approximate transport maps entail universality, as in the case where T_1^N is null, the eigenvalues λ_i^W under $P_N^{W,\beta}$ are just push forward of those under $P_N^{x^2}$ so that

$$N^{2/3} (\lambda_i^W - a) = N^{2/3} (T_0(\lambda_i^{x^2}) - T_0(2)) \simeq N^{2/3} T_0'(2) (\lambda_i^{x^2} - 2)$$

and a similar result holds inside the bulk. (34) allows indeed to neglect T_1^N.

4.3.2. Matrix models One would like to understand as well the fluctuations of the eigenvalues of polynomials in several independent matrices from the GUE or of matrices randomly chosen according to (22). Let us consider d independent matrices $X_1^N, \ldots, X_d^N$ taken from the GUE and P a polynomial in d noncommutative indeterminates. Then, it was shown (in fact also based on the analysis of Schwinger–Dyson equation) by Haagerup and Thorbjørnsen [30] that the operator norm of $P(X_1^N, \ldots, X_d^N)$ converges almost surely towards its free limit:

$$\|P\| = \lim_{n\to\infty} \sigma^d((PP^*)^n)^{1/2n} .$$

One could wonder about the fluctuations around this limit. Of course, it should depend on properties of the polynomial P since even in the case $d = 1$, the largest eigenvalue of $P(X_1^N)$ could either be the image of the largest eigenvalue of X_1^N or of an eigenvalue inside the bulk. It would be interesting to understand what are the optimal conditions so that the fluctuations are still given by the Tracy–Widom law. With Figalli, we could prove that this is the case when P is a small perturbation of the identity.

Theorem 15. *Let P be a self-adjoint polynomial. For ϵ small enough (independent of N), the eigenvalues of*

$$Y^N = X_1^N + \epsilon P(X_1^N, \ldots, X_d^N)$$

fluctuates locally as when $\epsilon = 0$, that is following the sine-kernel in the bulk and the Tracy–Widom law at the edge.

In fact, for ϵ small enough, a change of variable shows that the law of Y^N is of the type described in (22) (but with a slightly more complicated potential), hence we now focus on proving universality in the latter case. Take $V_a = \sum_{i=1}^d W_i(X_i) + aW(X_1, \ldots, X_d)$ and consider

$$\mathbb{P}_N^{V_a}(dX_1^N, \ldots, dX_d^N) = \frac{1}{Z_{V_a}^N} e^{-N\mathrm{Tr}(V_a(X_1^N,\ldots,X_d^N))} \prod_i dX_i^N$$

where integration holds over Hermitian or symmetric matrices. Denote $P_N^{V_a,i}$ the law of the ordered eigenvalues of X_i^N under $P_N^{V_a}$. The main result is that we can again construct approximate transport maps:

Theorem 16 (Figalli–G. 2014). *Assume V, W_i smooth enough, $W_i'' \geq c > 0$. Then, there exists $a(c) > 0$ so that for $a \in [-a(c), a(c)]$, there exist $T_0 : \mathbb{R} \mapsto \mathbb{R}$, $T_1^N : \mathbb{R}^N \mapsto \mathbb{R}^N$ smooth functions so that*

$$\left\| \left(T_0^{\otimes N} + \frac{T_1^N}{N}\right) \sharp P_N^{\sum X_j^2, i} - P_N^{V_a,i} \right\|_{TV} \leq const. \sqrt{\frac{\log N}{N}} ,$$

with $\sup_{\substack{1\le k\le N\\1\le i\le d}} \mathbb{E}[|T_{1,i}^{N,k}|] \le C\log N$ *and*

$$P_N^{\sum X_j^2,i}\Big(\sup_{\substack{1\le k,k'\le N\\1\le i\le d}} \frac{|T_{1,i}^{N,k}-T_{1.,i}^{N,k'}|}{\sqrt{N}|\lambda_k^i-\lambda_{k'}^i|}\ge C\log N\Big)\le N^{-cN}.$$

As a consequence, under $\mathbb{P}_N^{V_a}$, *the law of the spacing distribution of* X_i^N, $Nc_j^i(\lambda_j^i-\lambda_{j+1}^i)$, *converges to the Gaudin distribution, and that of* $N^{2/3}c_i(\max_j \lambda_j^i - C_i)$ *to the Tracy–Widom law, for some appropriate constants* c_i, C_i *and* c_j^i.

The idea to prove this theorem follows the arguments used in the previous section together with the fact that we can control integrals over the unitary groups by [27], hence rewrite, the law of the eigenvalues $P_N^{V_a}$ of $X_N^1,\dots,X_N^d$ under $\mathbb{P}_{V_a}^N$ is

$$P_N^{V_a}(d\lambda_j^i)=\frac{1}{\tilde{Z}_N^V}I_N^{aV}(\lambda_j^i)\prod_{i=1}^d\prod_{j<k}|\lambda_j^i-\lambda_k^i|^\beta e^{-N\sum W_i(\lambda_j^i)}\,d\lambda_j^i,$$

where

$$I_N^{aV}(\lambda_j^i)=\int e^{-aN\mathrm{Tr}(V(U_1^N D(\lambda^1)(U_1^N)^*,\dots,U_d^N D(\lambda^1)(U_d^N)^*))}\,dU_1^N\cdots dU_d^N$$

and integration holds under the Haar measure over the unitary group or the orthogonal group. By Theorem 12, we get

$$\begin{aligned}&I_N^{aV}(\lambda_j^i)\\&=\Big(1+O\Big(\frac{1}{N}\Big)\Big)\exp\Big\{(N^2F_2^a+NF_1^a+F_0^a)\Big(\frac{1}{N}\sum\delta_{\lambda_j^i},1\le i\le d\Big)\Big\}.\end{aligned}$$

We are then in position to construct approximate transport maps as for β-models.

References

[1] G.W. Anderson, A. Guionnet and O. Zeitouni, An Introduction to Random Matrices, Cambridge Stud. Adv. Math., **118**, Cambridge Univ. Press, Cambridge, 2010.

[2] F. Bekerman, A. Figalli and A. Guionnet, Transport maps for β-matrix models and universality, Comm. Math. Phys., **338** (2015), 589–619.

[3] F. Bekerman, Transport maps for β-matrix models in the multi-cut regime, preprint, arXiv:1512.00302.

[4] P. Biane and R. Speicher, Stochastic calculus with respect to free Brownian motion and analysis on Wigner space, Probab. Theory Related Fields, **112** (1998), 373–409.

[5] P. Biane and R. Speicher, Free diffusions, free entropy and free Fisher information, Ann. Inst. H. Poincaré Probab. Statist., **37** (2001), 581–606.

[6] A. Borodin, V. Gorin and A. Guionnet, Gaussian asymptotics of discrete β-ensembles, preprint, arXiv:1505.03760.

[7] G. Borot and B. Eynard, All order asymptotics of hyperbolic knot invariants from non-perturbative topological recursion of A-polynomials, preprint, arXiv:1205.2261.

[8] G. Borot and A. Guionnet, Asymptotic expansion of β matrix models in the one-cut regime, Comm. Math. Phys., **317** (2013), 447–483.

[9] G. Borot and A. Guionnet, Asymptotic expansion of β matrix models in the multi-cut regime, preprint, arXiv:1303.1045.

[10] G. Borot, A. Guionnet and K.K. Kozlowski, Large-N asymptotic expansion for mean field models with coulomb gas interaction, Int. Math. Res. Not. IMRN, **2015**, 10451–10524.

[11] P. Bourgade, L. Erdős and H.-T. Yau, Edge universality of beta ensembles, Comm. Math. Phys., **332** (2014), 261–353.

[12] P. Bourgade, L. Erdős and H.-T. Yau, Universality of general β-ensembles, Duke Math. J., **163** (2014), 1127–1190.

[13] L. Chekhov and B. Eynard, Matrix eigenvalue model: Feynman graph technique for all genera, J. High Energy Phys., **2006**, no. 12, 026.

[14] B. Collins, Moments and cumulants of polynomial random variables on unitary groups, the Itzykson–Zuber integral, and free probability, Int. Math. Res. Not., **2003**, 953–982.

[15] B. Collins, A. Guionnet and E. Maurel-Segala, Asymptotics of unitary and orthogonal matrix integrals, Adv. Math., **222** (2009), 172–215.

[16] Y. Dabrowski, A note about proving non-Γ under a finite non-microstates free Fisher information assumption, J. Funct. Anal., **258** (2010), 3662–3674.

[17] Y. Dabrowski, Analytic functions relative to a covariance map η: I. Generalized Haagerup products and analytic relations, preprint, arXiv:1503.05515.

[18] Y. Dabrowski, A. Guionnet and D. Shlyakhtenko, Free transport for convex potentials, preprint, 2016.

[19] Y. Dabrowski and A. Ioana, Unbounded derivations, free dilations, and indecomposability results for II_1 factors, Trans. Amer. Math. Soc., **368** (2016), 4525–4560.

[20] P. Deift and D. Gioev, Random Matrix Theory: Invariant Ensembles and Universality, Courant Lect. Notes Math., **18**, Courant Inst. Math. Sci., New York, NY, 2009.

[21] B. Eynard and N. Orantin, Invariants of algebraic curves and topological expansion, Commun. Number Theory Phys., **1** (2007), 347–452.

[22] A. Guionnet, Large Random Matrices: Lectures on Macroscopic Asymptotics. Lectures from the 36th Probability Summer School held in Saint-Flour, 2006, Lecture Notes in Math., **1957**, Springer-Verlag, 2009.

[23] A. Guionnet, V.F.R. Jones and D. Shlyakhtenko, Random matrices, free probability, planar algebras and subfactors, In: Quanta of Maths, Clay Math. Proc., **11**, Amer. Math. Soc., Providence, RI, 2010, pp. 201–239.

[24] A. Guionnet, V.F.R. Jones, D. Shlyakhtenko and P. Zinn-Justin, Loop models, random matrices and planar algebras, Comm. Math. Phys., **316** (2012), 45–97.

[25] A. Guionnet and E. Maurel-Segala, Combinatorial aspects of matrix models, ALEA Lat. Am. J. Probab. Math. Stat., **1** (2006), 241–279.

[26] A. Guionnet and E. Maurel-Segala, Second order asymptotics for matrix models, Ann. Probab., **35** (2007), 2160–2212.

[27] A. Guionnet and J. Novak, Asymptotics of unitary multimatrix models: The Schwinger–Dyson lattice and topological recursion, preprint, arXiv:1401.2703.

[28] A. Guionnet and D. Shlyakhtenko, Free diffusions and matrix models with strictly convex interaction, Geom. Funct. Anal., **18** (2009), 1875–1916.

[29] A. Guionnet and D. Shlyakhtenko, Free monotone transport, Invent. Math., **197** (2014), 613–661.

[30] U. Haagerup and S. Thorbjørnsen, A new application of random matrices: $\mathrm{Ext}(C^*_{\mathrm{red}}(F_2))$ is not a group, Ann. of Math. (2), **162** (2005), 711–775.
[31] V.F.R. Jones, D. Shlyakhtenko and K. Walker, An orthogonal approach to the subfactor of a planar algebra, Pacific J. Math., **246** (2010), 187–197.
[32] E. Maurel-Segala, High order expansion of matrix models and enumeration of maps, preprint, arXiv:math/0608192.
[33] N.A. Nekrasov, Non-perturbative Dyson–Schwinger equations and BPS/CFT correspondence, in preparation (2015).
[34] N.A. Nekrasov and V. Pestun, Seiberg–Witten geometry of four dimensional $N = 2$ quiver gauge theories, preprint, arXiv:1211.2240.
[35] N.A. Nekrasov, V. Pestun and S. Shatashvili, Quantum geometry and quiver gauge theories, preprint, arXiv:1312.6689 [hep-th].
[36] B. Nelson, Free monotone transport without a trace, Comm. Math. Phys., **334** (2015), 1245–1298.
[37] B. Nelson, Free transport for finite depth subfactor planar algebras, J. Funct. Anal., **268** (2015), 2586–2620.
[38] N. Ozawa, There is no separable universal II_1-factor, Proc. Amer. Math. Soc., **132** (2004), 487–490 (electronic).
[39] L. Pastur, Limiting laws of linear eigenvalue statistics for Hermitian matrix models, J. Math. Phys., **47** (2006), 103303.
[40] J.A. Ramírez, B. Rider and B. Virág, Beta ensembles, stochastic airy spectrum, and a diffusion, J. Amer. Math. Soc., **24** (2011), 919–944.
[41] M. Shcherbina, Change of variables as a method to study general β-models: bulk universality, preprint, arXiv:1310.7835.
[42] T. Tao, The asymptotic distribution of a single eigenvalue gap of a Wigner matrix, Probab. Theory Related Fields, **157** (2013), 81–106.
[43] C.A. Tracy and H. Widom, Level-spacing distributions and the Airy kernel, Comm. Math. Phys., **159** (1994), 151–174.
[44] C.A. Tracy and H. Widom, Level spacing distributions and the Bessel kernel, Comm. Math. Phys., **161** (1994), 289–309.
[45] F.G. Tricomi, Integral Equations, Pure and Applied Mathematics. Vol. V, Interscience Publishers, Inc., New York, NY, 1957.
[46] W.T. Tutte, A census of planar maps, Canad. J. Math., **15** (1963), 249–271.
[47] D.-V. Voiculescu, The coalgebra of the free difference quotient and free probability, Internat. Math. Res. Notices, **2000**, 79–106.
[48] D.-V. Voiculescu, A note on cyclic gradients, Indiana Univ. Math. J., **49** (2000), 837–841.
[49] D.-V. Voiculescu, Aspects of free analysis, Jpn. J. Math., **3** (2008), 163–183.
[50] A. Zvonkin, Matrix integrals and map enumeration: an accessible introduction, In: Combinatorics and Physics, Marseilles, 1995, Math. Comput. Modelling, **26**, Elsevier, Amsterdam, 1997, pp. 281–304.