Asymptotic expansion of smooth functions in polynomials in deterministic matrices and iid GUE matrices

Félix Parraud

Asymptotic expansion of smooth functions in polynomials in deterministic matrices and iid GUE matrices

Félix Parraud[1,2]

[1]Université de Lyon, ENSL, UMPA, 46 allée d'Italie, 69007 Lyon.
[2]Department of Mathematics, Graduate School of Science, Kyoto University, Kyoto 606-8502, Japan.

E-mail of the corresponding author: felix.parraud@ens-lyon.fr

Abstract

Let X^N be a family of $N \times N$ independent GUE random matrices, Z^N a family of deterministic matrices, P a self-adjoint non-commutative polynomial, that is for any N, $P(X^N)$ is self-adjoint, f a smooth function. We prove that for any k, if f is smooth enough, there exist deterministic constants $\alpha_i^P(f, Z^N)$ such that

$$\mathbb{E}\left[\frac{1}{N}\mathrm{Tr}\left(f(P(X^N, Z^N))\right)\right] = \sum_{i=0}^{k} \frac{\alpha_i^P(f, Z^N)}{N^{2i}} + \mathcal{O}(N^{-2k-2}).$$

Besides the constants $\alpha_i^P(f, Z^N)$ are built explicitly with the help of free probability. In particular, if x is a free semicircular system, then when the support of f and the spectrum of $P(x, Z^N)$ are disjoint, for any i, $\alpha_i^P(f, Z^N) = 0$. As a corollary, we prove that given $\alpha < 1/2$, for N large enough, every eigenvalue of $P(X^N, Z^N)$ is $N^{-\alpha}$-close from the spectrum of $P(x, Z^N)$.

1 Introduction

Asymptotic expansions in Random Matrix Theory created bridges between different worlds, including topology, statistical mechanics, and quantum field theory. In mathematics, a breakthrough was made in 1986 by Harer and Zagier who used the large dimension expansion of the moments of Gaussian matrices to compute the Euler characteristic of the moduli space of curves. A good introduction to this topic is given in the survey [32] by Zvonkin. In physics, the seminal works of t'Hooft [28] and Brézin, Parisi, Itzykson and Zuber [20] related matrix models with the enumeration of maps of any genus, hence providing a purely analytical tool to solve these hard combinatorial problems. Considering matrices in interaction via a potential, the so-called matrix models, indeed allows to consider the enumeration of maps with several vertices, including a possible coloring of the edges when the matrix model contains several matrices. This relation allowed to associate matrix models to statistical models on random graphs [38, 11, 23, 24, 26], as well as in [19] and [25] for the unitary case. This was also extended to the so-called β-ensembles in [27, 17, 12, 13, 15, 16]. Among other objects, these works study correlation functions and the so-called free energy and show that they expand as power series in the inverse of the dimension, and the coefficients of these expansions enumerate maps sorted by their genus. To compute asymptotic expansions, often referred to in the literature as topological expansions, one of the most successful methods is the loop equations method, see [35] and [36]. Depending on the model of random matrix, those are Tutte's equations, Schwinger-Dyson equations, Ward identities, Virasoro constraints, W-algebra or simply integration by part. This method was refined and used repeatedly in physics, see for example the work of Eynard and his collaborators, [21, 22, 18, 14]. At first those equations were only solved for the first few orders, however in 2004, in [22] and later [33] and [34], this method was refined to push the expansion to any orders recursively [37].

In this paper we want to generalize Harer-Zagier expansion for the moments of Gaussian matrices to more general smooth functions. Instead of a single GUE matrix, we will consider several independent matrices and deterministic matrices. We repeatedly use Schwinger-Dyson equations associated to GUE matrices to carry out our estimates. While we do not use the link between the coefficients of our expansion and map enumeration, as a corollary we get a new expression of these combinatorial objects. We show that the number of colored maps of genus g with a single specific vertex can be expressed as an integral, see remark 3.8 for a precise statement.

All papers quoted above have in common that they deal with polynomials or exponentials of polynomial evaluated in random matrices. With the exception of the work of Haagerup and Thorbjørnsen [10], smooth functions have not been considered. However being able to work with such functions is important for the applications. In particular we need to be able to work with functions with compact support to prove strong convergence results, that is proving the convergence of the spectrum for the Hausdorff distance. In this paper we establish a finite expansion of any orders around the dimension of the random matrix for the trace of smooth functions evaluated in polynomials in independent GUE random matrices. We refer to Definition 2.16 for a definition of those objects. The link between maps and topological expansion is a good motivation to prove such kind of theorem. Another motivation is to study the spectrum of polynomials in these random matrices: because we consider general smooth functions, our expansion will for instance allow to study the spectrum outside of the limiting bulk. In the case of a single GUE matrix, we have an explicit formula for the distribution of the eigenvalues of those random matrices, see Theorem 2.5.2 of [1]. However, if we consider polynomials in independent GUE matrices, we have no such result. The first result in this direction dates back to 1991 when Voiculescu proved in [7] that the renormalized trace of such polynomials converges towards a deterministic limit $\alpha(P)$. In particular given $X_1^N, \dots, X_d^N$ independent GUE matrices, the following holds true almost surely:

$$\lim_{N\to\infty} \frac{1}{N} \operatorname{Tr}_N \left(P(X_1^N, \dots, X_d^N) \right) = \alpha(P) \ . \tag{1}$$

Voiculescu computed the limit $\alpha(P)$ with the help of free probability. Besides if A_N is a self-adjoint matrix of size N, then one can define the empirical measure of its (real) eigenvalues by

$$\mu_{A_N} = \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i} \ ,$$

where δ_λ is the Dirac mass in λ and $\lambda_1, \dots, \lambda_N$ are the eigenvalue of A_N. In particular, if P is a self-adjoint polynomial, that is such that for any self adjoint matrices $A_1, \dots, A_d$, $P(A_1, \dots, A_d)$ is a self-adjoint matrix, then one can define the random measure $\mu_{P(X_1^N,\dots,X_d^N)}$. In this case, Voiculescu's result (1) implies that there exists a measure μ_P with compact support such that almost surely $\mu_{P(X_1^N,\dots,X_d^N)}$ converges weakly towards μ_P : it is given by $\mu_P(x^k) = \alpha(P^k)$ for all integer numbers k. Consequently, assuming we can apply Portmanteau theorem, the proportion of eigenvalues of $A_N = P(X_1^N, \dots, X_d^N)$ in the interval $[a, b]$, that is $\mu_{A_N}([a, b])$, converges towards $\mu_P([a, b])$.

Therefore in order to study the eigenvalues of a random matrix one has to study the renormalized trace of its moments. However if instead of studying the renormalized trace of polynomials in A_N, we study the non-renormalized trace of smooth function in A_N, then we can get precise information on the location of the eigenvalues. It all comes from the following remark, let f be a non-negative function such that f is equal to 1 on the interval $[a, b]$, then if $\sigma(A_N)$ is the spectrum of A_N,

$$\mathbb{P}\Big(\sigma(A_N) \cap [a, b] \neq \emptyset\Big) \leq \mathbb{P}\Big(\operatorname{Tr}_N \left(f(A_N) \right) \geq 1 \Big) \leq \mathbb{E}\Big[\operatorname{Tr}_N \left(f(A_N) \right) \Big].$$

Thus if one can show that the right-hand side of this inequality converges towards zero when N goes to infinity, then asymptotically there is no eigenvalue in the segment $[a, b]$. In the case of the random matrices that we study in this paper, that is polynomials in independent GUE matrices, a breakthrough was made in 2005 by Haagerup and Thorbjørnsen in [2]. They proved the almost sure convergence of the norm of those matrices. More precisely, they proved that for P a self-adjoint polynomial, almost surely, for any $\varepsilon > 0$ and N large enough,

$$\sigma\left(P(X_1^N, \dots, X_d^N) \right) \subset \operatorname{Supp} \mu_P + (-\varepsilon, \varepsilon) \ , \tag{2}$$

where $\operatorname{Supp} \mu_P$ is the support of the measure μ_P. In order to do so, they showed that given a smooth function f, there is a constant $\alpha_0^P(f)$, which can be computed explicitly with the help of free probability, such that

$$\mathbb{E}\Big[\frac{1}{N}\operatorname{Tr}_N\left(f(A_N)\right)\Big] = \alpha_0^P(f) + \mathcal{O}(N^{-2}).$$

A similar equality was proved in [8] with a better estimation of the dependency in the parameters such as f and Z^N in the $\mathcal{O}(N^{-2})$. Given the important consequences that studying the first two orders had, one can wonder what happens at the next order. More precisely, could we write this expectation as a finite order Taylor expansion, and what consequences would it have on the eigenvalues? That is, can we prove that for any k, if f is smooth enough, there exist deterministic constants $\alpha_i^P(f)$ such that

$$\mathbb{E}\left[\frac{1}{N}\operatorname{Tr}\left(f(P(X_1^N,\dots,X_d^N))\right)\right] \;=\; \sum_{i=0}^{k}\frac{\alpha_i^P(f)}{N^{2i}} \;+\; \mathcal{O}(N^{-2k-2})?$$

Haagerup and Thorbjørnsen gave a positive answer in 2010 (see [10]) for the specific case of a single GUE matrices, that is $d=1$. However the method of the proof relied heavily on the explicit formula of the law of the eigenvalues of a GUE matrix and since there is no equivalent for polynomials in GUE matrices we cannot adapt this proof. Instead, we developed a proof whose main tool is free probability. The main idea of the proof is to interpolate independent GUE matrices and free semicirculars with free Ornstein-Uhlenbeck processes. It is similar to the method used in [8]. The main result is the following Theorem.

Theorem 1.1. *We define,*

- $X^N = (X_1^N,\dots,X_d^N)$ *independent GUE matrices of size* N*,*
- $Z^N = (Z_1^N,\dots,Z_r^N,Z_1^{N^*},\dots,Z_r^{N^*})$ *deterministic matrices whose norm is uniformly bounded over* N*,*
- P *a self-adjoint polynomial which can be written as a linear combination of* $\mathbf{m}$ *monomials of degree at most* n *and coefficients at most* $c_{\max}$*,*
- $f:\mathbb{R}\mapsto\mathbb{R}$ *a function of class* $\mathcal{C}^{4(k+1)+2}$*. We define* $\|f\|_{\mathcal{C}^i}$ *the sum of the supremum on* $\mathbb{R}$ *of the first* i*-th derivatives of* f*.*

Then there exist deterministic coefficients $(\alpha_i^P(f,Z^N))_{1\le i\le k}$ *and constants* C,K *and* c *independent of* P*, such that with* $K_N=\max\{\left\|Z_1^N\right\|,\dots,\left\|Z_q^N\right\|,K\}$*,* $C_{\max}(P)=\max\{1,c_{\max}\}$*, for any* N*, if* $k\le cNn^{-1}$*,*

$$\begin{aligned}&\left|\mathbb{E}\left[\frac{1}{N}\operatorname{Tr}_N\Big(f(P(X^N,Z^N))\Big)\right] - \sum_{0\le i\le k}\frac{1}{N^{2i}}\alpha_i^P(f,Z^N)\right| \\ &\le \frac{1}{N^{2(k+1)}}\|f\|_{\mathcal{C}^{4(k+1)+2}}\times\Big(C\times n^2K_N^nC_{\max}\mathbf{m}\Big)^{4(k+1)+1}\times k^{12k}.\end{aligned} \tag{3}$$

Besides if we define $\widehat{K}_N$ *like* K_N *but with* 2 *instead of* K*, then we have that for any* i*,*

$$\left|\alpha_i^P(f,Z^N)\right| \le \|f\|_{\mathcal{C}^{4i+2}}\times\Big(C\times n^2\widehat{K}_N^nC_{\max}\mathbf{m}\Big)^{4i+1}\times i^{12i}. \tag{4}$$

Finally if f *and* g *are functions of class* $\mathcal{C}^{4(k+1)}$ *equal on a neighborhood of the spectrum of* $P(x,Z^N)$*, where* x *is a free semicircular system free from* $\mathbb{M}_N(\mathbb{C})$*, then for any* $i\le k$*,* $\alpha_i^P(f,Z^N)=\alpha_i^P(g,Z^N)$*. In particular if the support of* f *and the spectrum of* $P(x,Z^N)$ *are disjoint, then for any* i*,* $\alpha_i^P(f,Z^N)=0$*.*

This theorem is a consequence of the slightly sharper, but less explicit, Theorem 3.4. It is essentially the same statement, but instead of having the norm C^k of f, we make the moment of the Fourier transform of f appears. We also give an explicit expression for the coefficients α_i^P. The above Theorem calls for a few remarks.

- We assumed that the matrices Z^N are deterministic, but thanks to Fubini's theorem we can assume that they are random matrices as long as they are independent from X^N. In this situation though, K_N^n in the right side of the inequality is a random variable (and thus we need some additional assumptions if we want its expectation to be finite for instance).

- We assumed that the matrices Z^N were uniformly bounded over N. This is a technical assumption which is necessary to make sure that the coefficients α_i^P are well-defined. However as we can see in Theorem 3.4, one can relax this assumption. That being said, in order for equation (3) to be meaningful one has to be careful that the term ${K_N^n}^4$ is not compensating the term N^{-2}.
- The exponent 12 in the term k^{12k} is very suboptimal and could easily be optimized a bit more in the proof of Theorem 1.1. For a better bound we refer to Theorem 3.4, where the term k^{12k} is replaced by k^{3k}. However in order to work with the norm C^k instead of the moments of the Fourier transform, we were forced to increase this term.
- Although we cannot take $k = \infty$, hence only getting a finite Taylor expansion, we can still take k which depends on N. However to keep the last term under control we need to estimate the k-th derivative of f.
- Since the probability that there is an eigenvalue of $P(X^N, Z^N)$ outside of a neighborhood of $P(x, Z^N)$ is exponentially small as N goes to infinity. The hypothesis of smoothness on f only need to be verified on a neighborhood of $P(X^N, Z^N)$ for an asymptotic expansion to exist.

As we said earlier in the introduction, by studying the trace of a smooth function evaluated in $P(X_1^N, \dots, X_d^N)$, Haagerup and Thorbjørnsen were able to show in [2] that the spectrum of $P(X_1^N, \dots, X_d^N)$ converges for the Hausdorff distance towards an explicit subset of $\mathbb{R}$. We summarized this result in equation (2). With the full finite order Taylor expansion, by taking $f : x \to g(N^\alpha x)$ where g is a well-chosen smooth function, one can show the following proposition.

Corollary 1.2. *Let X^N be independent GUE matrices of size N, A^N a family of deterministic matrices whose norm is uniformly bounded over N, x be a free semicircular system and P a self-adjoint polynomial. Given $\alpha < 1/2$, almost surely for N large enough,*

$$\sigma\left(P(X^N, A^N)\right) \subset \sigma\left(P(x, A^N)\right) + N^{-\alpha},$$

where $\sigma(X)$ is the spectrum of X, and x is free from $\mathbb{M}_N(\mathbb{C})$.

In the case of a single GUE matrix, much more precise results were obtained by Tracy and Widom in [6]. They proved the existence of a continuous decreasing function F_2 from $\mathbb{R}$ to $[0, 1]$ such that if $\lambda_1(X^N)$ denotes the largest eigenvalue of X^N,

$$\lim_{N\to\infty} P\big(N^{2/3}(\lambda_1(X^N) - 2) \geq s\big) = F_2(s) \ .$$

This was generalized to β-matrix model in [29] and to polynomials in independent GUE matrices which are close to the identity in [30]. But there is no such result for general polynomials in independent GUE matrices. However with Theorem 1.1 we managed to get an estimate on the tail of the distribution of $\sqrt{N}\, \|P(X^N, A^N)\|$.

Corollary 1.3. *Let X^N be a vector of independent GUE matrices of size N, A^N a family of deterministic matrices whose norm is uniformly bounded over N, x be a free semicircular system and P a polynomial. Then there exists a constant C such that for N large enough,*

$$\mathbb{P}\left(\frac{\sqrt{N}}{\ln^4 N}\left(\|P(X^N, A^N)\| - \|P(x, A^N)\|\right) \geq C\,(\delta + 1)\right) \leq e^{-N} + e^{-\delta^2 \ln^8 N}.$$

This corollary is similar to Theorem 1.5 obtained in [8], but with a substantial improvement on the exponent since in this theorem, instead of 1/2, we only had 1/4. Theorem 1.5 of [8] also gave a similar bound on the probability that $\|P(X^N)\|$ be smaller than its deterministic limit, but Theorem 3.4 does not yield any improvement on this inequality. The proof of this corollary can be summarized in two steps: first use measure concentration to get an estimate on the probability that $\|P(X^N, A^N)\|$ is far from its expectation, and secondly use Theorem 1.1 to estimate the difference between the expectation and the deterministic limit. Finally it is worth noting that the exponent 1/2 comes from the fact that for every N^2 that we gain in equation (3), we also have to differentiate our function f four more times. Thus if we take $f : x \to g(N^\alpha x)$ where g is smooth, then in order for N^{-2} to compensate the differential, we have to take $\alpha = 1/2$. If we only had to differentiate our function three more times, then we could take $\alpha = 2/3$ which is the same exponent as in Tracy-Widom.

2 Framework and standard properties

2.1 Usual definitions in free probability

In order to be self-contained, we begin by recalling the following definitions from free probability.

Definition 2.1.
- *A $\mathcal{C}^*$-probability space $(\mathcal{A}, *, \tau, \|.\|)$ is a unital $\mathcal{C}^*$-algebra $(\mathcal{A}, *, \|.\|)$ endowed with a state τ, i.e. a linear map $\tau : \mathcal{A} \to \mathbb{C}$ satisfying $\tau(1_{\mathcal{A}}) = 1$ and $\tau(a^*a) \geq 0$ for all $a \in \mathcal{A}$. In this paper we always assume that τ is a trace, i.e. that it satisfies $\tau(ab) = \tau(ba)$ for any $a, b \in \mathcal{A}$. An element of $\mathcal{A}$ is called a (non commutative) random variable. We will always work with a faithful trace, namely, for $a \in \mathcal{A}$, $\tau(a^*a) = 0$ if and only if $a = 0$.*

- *Let $\mathcal{A}_1, \dots, \mathcal{A}_n$ be $*$-subalgebras of $\mathcal{A}$, having the same unit as $\mathcal{A}$. They are said to be free if for all k, for all $a_i \in \mathcal{A}_{j_i}$ such that $j_1 \neq j_2$, $j_2 \neq j_3$, $\dots$, $j_{k-1} \neq j_k$:*

$$\tau\Big((a_1 - \tau(a_1))(a_2 - \tau(a_2)) \dots (a_k - \tau(a_k))\Big) = 0.$$

  Families of non-commutative random variables are said to be free if the $*$-subalgebras they generate are free.

- *Let $A = (a_1, \dots, a_k)$ be a k-tuple of random variables. The joint distribution of the family A is the linear form $\mu_A : P \mapsto \tau\big[P(A, A^*)\big]$ on the set of polynomials in $2k$ non-commutative variables.*

- *A family of non commutative random variables $x = (x_1, \dots, x_d)$ is called a free semicircular system when the non commutative random variables are free, self-adjoint ($x_i = x_i^*$), and for all k in $\mathbb{N}$ and i, one has*

$$\tau(x_i^k) = \int t^k d\sigma(t),$$

  with $d\sigma(t) = \frac{1}{2\pi}\sqrt{4 - t^2}\ \mathbf{1}_{|t| \leq 2}\ dt$ the semicircle distribution.

It is important to note that thanks to [5, Theorem 7.9], that we recall below, one can consider free copies of any non-commutative random variable.

Theorem 2.2. *Let $(\mathcal{A}_i, \phi_i)_{i \in I}$ be a family of $\mathcal{C}^*$-probability spaces such that the functionals $\phi_i : \mathcal{A}_i \to \mathbb{C}$, $i \in I$, are faithful traces. Then there exist a $\mathcal{C}^*$-probability space $(\mathcal{A}, \phi)$ with ϕ a faithful trace, and a family of norm-preserving unital $*$-homomorphism $W_i : \mathcal{A}_i \to \mathcal{A}$, $i \in I$, such that:*

- *$\phi \circ W_i = \phi_i$, $\forall i \in I$.*
- *The unital $\mathcal{C}^*$-subalgebras form a free family in $(\mathcal{A}, \phi)$.*

Let us finally fix a few notations concerning the spaces and traces that we use in this paper.

Definition 2.3.
- *$(\mathcal{A}_N, \tau_N)$ is the free sum of $\mathbb{M}_N(\mathbb{C})$ with a system of d free semicircular variable, this is the $\mathcal{C}^*$-probability space built in Theorem 2.2. Note that when restricted to $\mathbb{M}_N(\mathbb{C})$, τ_N is just the regular renormalized trace on matrices. The restriction of τ_N to the $\mathcal{C}^*$-algebra generated by the free semicircular system x is denoted as τ. Note that one can view this space as the limit of a matrix space, we refer to Proposition 3.5 from [8].*

- *Tr_N is the non-renormalized trace on $\mathbb{M}_N(\mathbb{C})$.*

- *We denote $E_{r,s}$ the matrix with coefficients equal to 0 except in (r, s) where it is equal to one.*

- *We regularly identify $\mathbb{M}_N(\mathbb{C}) \otimes \mathbb{M}_k(\mathbb{C})$ with $\mathbb{M}_{kN}(\mathbb{C})$ through the isomorphism $E_{i,j} \otimes E_{r,s} \mapsto E_{i+rN, j+sN}$, similarly we identify $\mathrm{Tr}_N \otimes \mathrm{Tr}_k$ with Tr_{kN}.*

- *If $A^N = (A_1^N, \dots, A_d^N)$ and $B^k = (B_1^k, \dots, B_d^k)$ are two vectors of random matrices, then we denote $A^N \otimes B^k = (A_1^N \otimes B_1^k, \dots, A_d^N \otimes B_d^k)$. We typically use the notation $X^N \otimes I_k$ for the vector $(X_1^N \otimes I_k, \dots, X_1^N \otimes I_k)$.*

2.2 Non-commutative polynomials and derivatives

Let $\mathcal{A}_{d,2r} = \mathbb{C}\langle X_1, \ldots, X_d, Y_1, \ldots, Y_{2r}\rangle$ be the set of non-commutative polynomial in $p+2r$ variables. We set $q = 2r$ to simplify notations. We endow this vector space with the norm

$$\|P\|_A = \sum_{M \text{monomial}} |c_M(P)| A^{\deg M} , \tag{5}$$

where $c_M(P)$ is the coefficient of P for the monomial M and $\deg M$ the total degree of M (that is the sum of its degree in each letter $X_1, \ldots, X_d, Y_1, \ldots, Y_{2r}$). Let us define several maps which we use frequently in the sequel. First, for $A, B, C \in \mathcal{A}_{d,q}$, let

$$A \otimes B \# C = ACB, \ A \otimes B \tilde{\#} C = BCA, \ m(A \otimes B) = BA. \tag{6}$$

We define an involution $*$ on $\mathcal{A}_{d,q}$ by $X_i^* = X_i$, $Y_i^* = Y_{i+r}$ if $i \leq d+r$, $Y_i^* = Y_{i-r}$ else, and then we extend it to $\mathcal{A}_{d,q}$ by linearity and the formula $(\alpha PQ)^* = \overline{\alpha} Q^* P^*$. $P \in \mathcal{A}_{d,q}$ is said to be self-adjoint if $P^* = P$. Self-adjoint polynomials have the property that if $x_1, \ldots, x_d, z_1, \ldots, z_r$ are elements of a $\mathcal{C}^*$-algebra such as $x_1, \ldots, x_d$ are self-adjoint, then so is $P(x_1, \ldots, x_d, z_1, \ldots, z_r, z_1^*, \ldots, z_r^*)$.

Definition 2.4. *If $1 \leq i \leq d$, one defines the non-commutative derivative $\partial_i : \mathcal{A}_{d,q} \longrightarrow \mathcal{A}_{d,q} \otimes \mathcal{A}_{d,q}$ by its value on a monomial $M \in \mathcal{A}_{d,q}$ given by*

$$\partial_i M = \sum_{M = AX_iB} A \otimes B ,$$

and then extend it by linearity to all polynomials. We can also define ∂_i by induction with the formulas,

$$\begin{aligned} \forall P, Q \in \mathcal{A}_{d,q}, \quad & \partial_i(PQ) = \partial_i P \times 1 \otimes Q + P \otimes 1 \times \partial_i Q, \\ \forall i, j, \quad & \partial_i X_i = \partial_{i,j} 1 \otimes 1. \end{aligned} \tag{7}$$

Similarly, with m as in (6)*, one defines the cyclic derivative $D_i : \mathcal{A}_{d,q} \longrightarrow \mathcal{A}_{d,q}$ for $P \in \mathcal{A}_{d,q}$ by*

$$D_i P = m \circ \partial_i P .$$

Definition 2.5. *We define $\mathcal{F}_{d,q}$ to be the $*$-algebra generated by $\mathcal{A}_{d,q}$ and the family $\left\{ e^{\mathbf{i}Q} \mid Q \in \mathcal{A}_{d,q} \text{ self-adjoint} \right\}$.*

Then, as we will see in the next proposition, a natural way to extend the definition of ∂_i (and D_i) to $\mathcal{F}_{d,q}$ is by setting

$$\partial_i e^{\mathbf{i}Q} = \mathbf{i} \int_0^1 e^{\mathbf{i}\alpha Q} \otimes 1 \ \partial_i Q \ 1 \otimes e^{\mathbf{i}(1-\alpha)Q} d\alpha. \tag{8}$$

However we cannot define the integral properly on $\mathcal{F}_{d,q} \otimes \mathcal{F}_{d,q}$. After evaluating our polynomials in $\mathcal{C}^*$-algebras, the integral will be well-defined as we will see. Firstly, we need to define properly the operator norm of tensor of $\mathcal{C}^*$-algebras. We work with the minimal tensor product also named the spatial tensor product. For more information we refer to chapter 6 of [4].

Definition 2.6. *Let $\mathcal{A}$ and $\mathcal{B}$ be $\mathcal{C}^*$-algebra with faithful representations $(H_{\mathcal{A}}, \phi_{\mathcal{A}})$ and $(H_{\mathcal{B}}, \phi_{\mathcal{B}})$, then if $\otimes_2$ is the tensor product of Hilbert spaces, $\mathcal{A} \otimes_{\min} \mathcal{B}$ is the completion of the image of $\phi_{\mathcal{A}} \otimes \phi_{\mathcal{B}}$ in $B(H_{\mathcal{A}} \otimes_2 H_{\mathcal{B}})$ for the operator norm in this space. This definition is independent of the representations that we fixed.*

While we will not always be in this situation during this paper, it is important to note that if $\mathcal{A} = \mathbb{M}_N(\mathbb{C})$, then up to isomorphism $\mathcal{A} \otimes_{\min} \mathcal{A}$ is simply $\mathbb{M}_{N^2}(\mathbb{C})$ with the usual operator norm. If $P \in \mathcal{A}_{d,q}$, $z = (z_1, \ldots, z_{d+q})$ belongs to a $\mathcal{C}^*$-algebra $\mathcal{A}$, then $(\partial_i P^k)(z)$ belongs to $\mathcal{A} \otimes_{\min} \mathcal{A}$, and $\left\|(\partial_i P^k)(z)\right\| \leq C_P k \left\|P(z)\right\|^{k-1}$ for some constant C_P independent of k. Thus we can define

$$(\partial_i e^P)(z) = \sum_{k \in \mathbb{N}} \frac{1}{k!} (\partial_i P^k)(z). \tag{9}$$

We have now defined the non-commutative differential of the exponential of a polynomial twice, in (8) and (9). However those two definitions are compatible thanks to the following proposition (see [9], Proposition 2.2 for the proof).

Proposition 2.7. *Let $P \in \mathcal{A}_{d,q}$, $z = (z_1, \dots, z_{d+q})$ elements of a $\mathcal{C}^*$-algebra $\mathcal{A}$, then with $(\partial_i e^P)(z)$ defined as in* (9),

$$\left(\partial_i e^P\right)(z) = \int_0^1 e^{\alpha P(z)} \otimes 1 \; \partial_i P(z) \; 1 \otimes e^{(1-\alpha)P(z)} \; d\alpha.$$

We explained why (8) was well-defined when we evaluate our polynomials in a $\mathcal{C}^*$-algebra. However in order to be perfectly rigorous we need to give the following definition for the non-commutative differential that we use in the rest of this paper.

Definition 2.8. *For $\alpha \in [0,1]$, let $\partial_{\alpha,i} : \mathcal{F}_{d,q} \to \mathcal{F}_{d,q} \otimes \mathcal{F}_{d,q}$ which satisfies* (7) *and such that for any $P \in \mathcal{A}_{d,q}$ self-adjoint,*

$$\partial_{\alpha,i} e^{\mathbf{i}P} = \mathbf{i} e^{\mathbf{i}\alpha P} \otimes 1 \; \partial_i P \; 1 \otimes e^{\mathbf{i}(1-\alpha)P}.$$

And then, given $z = (z_1, \dots, z_{d+q})$ elements of a $\mathcal{C}^*$-algebra, we define for any $Q \in \mathcal{F}_{d,q}$,

$$\partial_i Q(z) = \int_0^1 \partial_{\alpha,i} Q(z) \; d\alpha.$$

In particular, it means that we can define rigorously the composition of those applications. Since the map $\partial_{\alpha,i}$ goes from $\mathcal{F}_{d,q}$ to $\mathcal{F}_{d,q} \otimes \mathcal{F}_{d,q}$ it is very easy to do so. In particular we will use the very specific operator (see Definition 2.10 for the notation ∂_i^1, ∂_i^2 and $\boxtimes$).

Definition 2.9. *Let $Q \in \mathcal{F}_{d,q}$, given $z = (z_1, \dots, z_{d+q})$ elements of a $\mathcal{C}^*$-algebra, let $i, j \in [1, d]$, we define*

$$D_j \partial_i^1 D_i Q \boxtimes D_j \partial_i^2 D_i Q(z) = \int_{[0,1]^4} (m \circ \partial_{\alpha_3,j}) \circ \partial^1_{\alpha_2,i} \circ (m \circ \partial_{\alpha_1,i}) Q(z) \boxtimes (m \circ \partial_{\alpha_4,j}) \circ \partial^2_{\alpha_2,i} \circ (m \circ \partial_{\alpha_1,i}) Q(z) \; d\alpha.$$

For the sake of clarity, we introduce the following notation which is close to Sweedler's convention. Its interest will be clear in section 3.

Definition 2.10. *Let $Q \in \mathcal{F}_{d,q}$, $\mathcal{C}$ be a $\mathcal{C}^*$-algebra. Then let $\alpha : \mathcal{F}_{d,q} \to \mathcal{C}$ and $\beta : \mathcal{F}_{d,q} \to \mathcal{C}$ be morphisms. We also set $n : A \otimes B \in \mathcal{C} \otimes \mathcal{C} \mapsto AB \in \mathcal{C}$. Then we use the following notation,*

$$\alpha(\delta_i^1 P) \boxtimes \beta(\delta_i^2 P) = n \circ (\alpha \otimes \beta(\delta_i P)).$$

This notation is especially useful when our applications α and β are simply evaluation of P as it is the case in section 3. Indeed we typically denote, $\delta_i^1 P(X) \boxtimes \delta_i^2 P(Y)$, rather than define $h_X : P \to P(X)$ and use the more cumbersome and abstract notation, $n \circ (h_X \otimes h_Y(\delta_i P))$.

The map ∂_i is related to the so-called Schwinger-Dyson equations on semicircular variable thanks to the following proposition. One can find a proof for polynomials in [1], Lemma 5.4.7, and then extend it to $\mathcal{F}_{d,q}$ thanks to Definition (9).

Proposition 2.11. *Let $x = (x_1, \dots, x_d)$ be a free semicircular system, $y = (y_1, \dots, y_r)$ be non-commutative random variables free from x, if the family (x, y) belongs to the $\mathcal{C}^*$-probability space $(\mathcal{A}, *, \tau, \|.\|)$, then for any $Q \in \mathcal{F}_{d,q}$,*

$$\tau(Q(x, y, y^*) \; x_i) = \tau \otimes \tau(\partial_i Q(x, y, y^*)) \; .$$

Now that we have defined the usual non-commutative polynomial spaces, we build a very specific one which we need to define properly the coefficients of the topological expansion.

Definition 2.12. *Let $(c_n)_n$ be the sequence such that $c_0 = 0$, $c_{n+1} = 2c_n + 2$. We define by induction, $J_0 = \{\emptyset\}$ and for $n \geq 0$,*

$$J_{n+1} = \{ \; \{I, 2c_n + 1\} \; | I \in J_n\} \cup \{ \; \{s_1 + c_n, \dots, s_n + c_n, 2c_n + 2\} \; | I = \{s_1, \dots, s_n\} \in J_n\} \, .$$

Then we define $\mathcal{A}^n_{d,q} = \mathbb{C}\langle X_{i,I}, \; 1 \leq i \leq p, I \in J_n, Y_1, \dots, Y_r, Y_1^*, \dots, Y_r^* \rangle$. We also define $\mathcal{F}^n_{d,q}$ as the $*$-algebra generated by $\mathcal{A}^n_{d,q}$ and the family $\left\{ e^{\mathbf{i}Q} \mid Q \in \mathcal{A}^n_{d,q} \text{ self-adjoint} \right\}$.

Definition 2.13. *Similarly to Definition 2.8, we define ∂_i and $\partial_{i,I}$ on $\mathcal{F}^n_{d,q}$ which satisfies* (7) *and* (8) *and*

$$\forall i,j \in [1,p],\ I,K \in J_n, \quad \partial_{i,I} X_{j,K} = \delta_{i,j}\delta_{I,K} 1 \otimes 1, \quad \partial_i X_{j,K} = \delta_{i,j} 1 \otimes 1.$$

We then define $D_i = m \circ \partial_i$ and $D_{i,I} = m \circ \partial_{i,I}$ on $\mathcal{F}^n_{d,q}$.

In particular, $\mathcal{F}^0_{d,q} = \mathcal{F}_{d,q}$ and the two definitions of ∂_i coincide.

Remark 2.14. *Note that given $s \in [1, c_n]$, then there exists a unique $l \leq n$ such that for any $I = \{s_1, \dots, s_n\} \in J_n$, either $s_l = s$ or $s \notin I$. Besides for any $I = \{s_1, \dots, s_n\}, K = \{k_1, \dots, k_n\} \in J_n$, if there exists l such that $s_l = k_l$, then for any $q \geq l$, $s_q = k_q$.*

This prompts us to define the following,

Definition 2.15. *Let $l \in [0,n]$, we define $J^l_n = \{\ \{s_{l+1}, \dots, s_n\} \mid \{s_1, \dots, s_n\} \in J_n\ \}$, then we set*

$$\forall A \in J^l_n, \quad \partial_{i,A} = \sum_{I \in J_n \text{ such that } A \subset I} \partial_{i,I}.$$

And finally we define $D_{i,A} = m \circ \partial_{i,A}$ on $\mathcal{F}^n_{d,q}$.

In particular $J^0_n = J_n$, $J^n_n = \{\emptyset\}$ and $\partial_{i,\emptyset} = \partial_i$. It is worth noting that thanks to remark 2.14, $A \subset I$ is equivalent to having the last l coefficients of I equal to those of A.

2.3 GUE random matrices

We conclude this section by reminding the definition of Gaussian random matrices and stating a few useful properties about them.

Definition 2.16. *A GUE random matrix X^N of size N is a self-adjoint matrix whose coefficients are random variables with the following laws:*

- *For $1 \leq i \leq N$, the random variables $\sqrt{N} X^N_{i,i}$ are independent centered Gaussian random variables of variance 1.*
- *For $1 \leq i < j \leq N$, the random variables $\sqrt{2N}\ \Re X^N_{i,j}$ and $\sqrt{2N}\ \Im X^N_{i,j}$ are independent centered Gaussian random variables of variance 1, independent of $\left(X^N_{i,i}\right)_i$.*

When doing computations with Gaussian variables, the main tool that we use is Gaussian integration by part. It can be summarized into the following formula, if Z is a centered Gaussian variable with variance 1 and f a $\mathcal{C}^1$ function, then

$$\mathbb{E}[Zf(Z)] = \mathbb{E}[\partial_Z f(Z)]\ . \tag{10}$$

A direct consequence of this, is that if x and y are centered Gaussian variable with variance 1, and $Z = \frac{x + \mathbf{i}y}{\sqrt{2}}$, then

$$\mathbb{E}[Zf(x,y)] = \mathbb{E}[\partial_Z f(x,y)] \quad \text{and} \quad \mathbb{E}[\overline{Z} f(x,y)] = \mathbb{E}[\partial_{\overline{Z}} f(x,y)]\ , \tag{11}$$

where $\partial_Z = \frac{1}{2}(\partial_x + \mathbf{i}\partial_y)$ and $\partial_{\overline{Z}} = \frac{1}{2}(\partial_x - \mathbf{i}\partial_y)$. When working with GUE matrices, an important consequence of this are the so-called Schwinger-Dyson equations, which we summarize in the following proposition. For more information about these equations and their applications, we refer to [1], Lemma 5.4.7.

Proposition 2.17. *Let X^N be GUE matrices of size N, $Q \in \mathcal{F}_{d,q}$, then for any i,*

$$\mathbb{E}\left[\frac{1}{N}\mathrm{Tr}_N(X^N_i\ Q(X^N))\right] = \mathbb{E}\left[\left(\frac{1}{N}\mathrm{Tr}_N\right)^{\otimes 2}(\partial_i Q(X^N))\right].$$

Proof. Let us first assume that $Q \in \mathcal{A}_{d,q}$. One can write $X_i^N = \frac{1}{\sqrt{N}}(x_{r,s}^i)_{1\leq r,s\leq N}$ and thus

$$\begin{aligned}\mathbb{E}\left[\frac{1}{N}\operatorname{Tr}_N(X_i^N\ Q(X^N))\right] &= \frac{1}{N^{3/2}}\sum_{r,s}\mathbb{E}\left[x_{r,s}^i\ \operatorname{Tr}_N(E_{r,s}\ Q(X^N))\right] \\ &= \frac{1}{N^{3/2}}\sum_{r,s}\mathbb{E}\left[\operatorname{Tr}_N(E_{r,s}\ \partial_{x_{r,s}^i}Q(X^N))\right] \\ &= \frac{1}{N^2}\sum_{r,s}\mathbb{E}\left[\operatorname{Tr}_N(E_{r,s}\ \partial_i Q(X^N)\# E_{s,r})\right] \\ &= \mathbb{E}\left[\left(\frac{1}{N}\operatorname{Tr}_N\right)^{\otimes 2}(\partial_i Q(X^N))\right].\end{aligned}$$

If $Q \in \mathcal{F}_{d,q}$, then the proof is pretty much the same but we need to use Duhamel's formula which states that for any matrices A and B,

$$e^B - e^A = \int_0^1 e^{\alpha B}(B-A)e^{(1-\alpha)A}\ d\alpha. \tag{12}$$

Thus this let us prove that

$$\partial_{x_{r,s}^i} e^{\mathbf{i}P(X^N)} = \mathbf{i}\int_0^1 e^{\mathbf{i}\alpha P(X^N)}\ \partial_i P(X^N)\# E_{s,r}\ e^{\mathbf{i}(1-\alpha)P(X^N)}\ d\alpha.$$

And the conclusion follows.

□

Now to finish this section we state a property that we use several times in this paper. For the proof we refer to Proposition 2.11 in [8].

Proposition 2.18. *There exist constants C, D and α such that for any $N \in \mathbb{N}$, if X^N is a GUE random matrix of size N, then for any $u \geq 0$,*

$$\mathbb{P}\left(\|X^N\| \geq u + D\right) \leq e^{-\alpha u N}.$$

Consequently, for any $k \leq \alpha N/2$,

$$\mathbb{E}\left[\|X^N\|^k\right] \leq C^k.$$

3 Proof of Theorem 1.1

3.1 A Poincaré type equality

One of the main tool when dealing with GUE random matrices is the Poincaré inequality (see Definition 4.4.2 from [1]), which gives us a sharp upper bound of the variance of a function in these matrices. Typically this inequality shows that the variance of a trace of a polynomial in GUE random matrices, which a priori is of order $\mathcal{O}(1)$, is of order $\mathcal{O}(N^{-2})$. In this paper we use the same kind of argument which are used to prove the Poincaré inequality to get an exact formula for the variances we are interested in.

Proposition 3.1. *Let $P, Q \in \mathcal{F}_{d,q}$, R^N, S^N, T^N be independent vectors of d independent GUE matrices of size N. Let A^N be a vector of deterministic matrices and their adjoints. With convention* $\operatorname{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$*, for any $t \geq 0$, we have:*

$$\begin{aligned}&\operatorname{Cov}\left(\operatorname{Tr}_N\left(P\left((1-e^{-t})^{1/2}R^N, A^N\right)\right), \operatorname{Tr}_N\left(Q\left((1-e^{-t})^{1/2}R^N, A^N\right)\right)\right) \\ &= \frac{1}{N}\sum_i\int_0^t \mathbb{E}\Big[\operatorname{Tr}_N\Big(D_i P\left((e^{-s}-e^{-t})^{1/2}R^N + (1-e^{-s})^{1/2}S^N, A^N\right) \\ &\qquad\qquad \times D_i Q\left((e^{-s}-e^{-t})^{1/2}R^N + (1-e^{-s})^{1/2}T^N, A^N\right)\Big)\Big]ds.\end{aligned}$$

Proof. We define the following function,

$$h(s) = \mathbb{E}\Big[\operatorname{Tr}_N \Big(P \Big((e^{-s} - e^{-t})^{1/2} R^N + (1 - e^{-s})^{1/2} S^N, A^N \Big) \Big) \\ \operatorname{Tr}_N \Big(Q \Big((e^{-s} - e^{-t})^{1/2} R^N + (1 - e^{-s})^{1/2} T^N, A^N \Big) \Big) \Big].$$

To simplify notations, we set

$$S_s^N = \Big((e^{-s} - e^{-t})^{1/2} R^N + (1 - e^{-s})^{1/2} S^N, A^N \Big),\ T_s^N = \Big((e^{-s} - e^{-t})^{1/2} R^N + (1 - e^{-s})^{1/2} T^N, A^N \Big).$$

Then we have,

$$\operatorname{Cov}\Big(\operatorname{Tr}_N \Big(P \Big((1 - e^{-t})^{1/2} R^N, A^N \Big) \Big), \operatorname{Tr}_N \Big(Q \Big((1 - e^{-t})^{1/2} R^N, A^N \Big) \Big) \Big) = - \int_0^t \frac{dh}{ds}(s)\ ds.$$

Thanks to Duhamel's formula (see (12)) we find

$$\frac{dP\left(S_s^N, A^N\right)}{ds} = -\frac{e^{-s}}{2} \sum_{i=1}^{d} \partial_i P\left(S_s^N, A^N\right) \# \left(\frac{R_i^N}{(e^{-s} - e^{-t})^{1/2}} - \frac{S_i^N}{(1 - e^{-s})^{1/2}} \right).$$

Since $\operatorname{Tr}_N(\partial_i P \# B) = \operatorname{Tr}_N(D_i P \times B)$, we compute,

$$\frac{dh}{ds}(s) = -\frac{e^{-s}}{2} \sum_i \mathbb{E}\Bigg[\operatorname{Tr}_N \left(D_i P\left(S_s^N\right) \left(\frac{R_i^N}{(e^{-s} - e^{-t})^{1/2}} - \frac{S_i^N}{(1 - e^{-s})^{1/2}} \right) \right) \operatorname{Tr}_N\left(Q\left(T_s^N\right)\right) \\ + \operatorname{Tr}_N\left(P\left(S_s^N\right)\right) \operatorname{Tr}_N \left(D_i Q\left(S_s^N\right) \left(\frac{R_i^N}{(e^{-s} - e^{-t})^{1/2}} - \frac{T_i^N}{(1 - e^{-s})^{1/2}} \right) \right) \Bigg].$$

But by using integration by part formula (11), we get that

$$\mathbb{E}\Bigg[\operatorname{Tr}_N \left(D_i P\left(S_s^N\right) \frac{R_i^N}{(e^{-s} - e^{-t})^{1/2}} \right) \operatorname{Tr}_N\left(Q\left(T_s^N\right)\right) \Bigg] \\ = \frac{1}{N} \sum_{1 \le a,b \le N} \mathbb{E}\Bigg[\operatorname{Tr}_N\left(E_{a,b}\ \partial_i D_i P\left(S_s^N\right) \# E_{b,a}\right) \times \operatorname{Tr}_N\left(Q\left(T_s^N\right)\right) \\ + \operatorname{Tr}_N\left(D_i P\left(S_s^N\right) E_{a,b}\right) \times \operatorname{Tr}_N\left(D_i Q\left(T_s^N\right) E_{b,a}\right) \Bigg].$$

And similarly

$$\mathbb{E}\Bigg[\operatorname{Tr}_N \left(D_i P\left(S_s^N\right) \frac{S_i^N}{(1 - e^{-s})^{1/2}} \right) \operatorname{Tr}_N\left(Q\left(T_s^N\right)\right) \Bigg] \\ = \frac{1}{N} \sum_{1 \le a,b \le N} \mathbb{E}\Bigg[\operatorname{Tr}_N\left(E_{a,b}\ \partial_i D_i P\left(S_s^N\right) \# E_{b,a}\right) \times \operatorname{Tr}_N\left(Q\left(T_s^N\right)\right) \Bigg].$$

Therefore with similar computations we conclude,

$$\frac{dh}{ds}(s) = -\frac{1}{N} e^{-s} \sum_i \mathbb{E}\Bigg[\operatorname{Tr}_N \left(D_i P\left(S_s^N\right) D_i P\left(S_s^N\right) \right) \Bigg].$$

Hence the conclusion.

□

3.2 A first rough formulation of the coefficients

In this subsection we prove the following lemma which will be the backbone of the proof of the topological expansion. The heuristic behind this lemma is that if $Q \in \mathcal{F}_{d(p+1),q}$, X^N independent GUE matrices, $(y_i)_{i\geq 1}$ systems of free semicirculars free between each other. Then we can find $R \in \mathcal{F}_{2d(p+1)+1,q}$ such that

$$\mathbb{E}\left[\tau_N\Big(Q\left(X^N,(y_i)_{1\leq i\leq p}\right),Z^N\Big)\right]-\tau_N\Big(Q\left(x,(y_i)_{1\leq i\leq p}\right),Z^N\Big)=\frac{1}{N^2}\mathbb{E}\left[\tau_N\Big(R\left(X^N,(y_i)_{1\leq i\leq 2p+2},Z^N\right)\Big)\right].$$

Then we will only need to apply this lemma recursively to build the topological expansion. Note that thanks to the definition of $\mathcal{A}_N$ in Definition 2.3, it makes sense to consider matrices and free semicirculars in the same space. One can also assume that those matrices are random thanks to Proposition 2.7 of [8].

Lemma 3.2. *Let the following objects be given,*

- $X^N = (X_1^N, \dots, X_d^N)$ *independent GUE matrices of size* N,
- $x = (x_1, \dots, x_d)$, $y_s = (y_{s,1}, \dots, y_{s,d})$, *systems of free semicircular variables, free between each other,*
- $Z^N = (Z_1^N, \dots, Z_q^N)$ *deterministic matrices and their adjoints,*
- $Y^N = \big(X^N, (1-e^{-t_1})^{1/2}y_1, \dots, (1-e^{-t_d})^{1/2}y_p, Z^N\big)$,
- $Y = \big(x, (1-e^{-t_1})^{1/2}y_1, \dots, (1-e^{-t_d})^{1/2}y_p, Z^N\big)$,
- $Y_t^N = \big(e^{-t/2}X^N + (1-e^{-t})^{1/2}x, (1-e^{-t_1})^{1/2}y_1, \dots, (1-e^{-t_d})^{1/2}y_p, Z^N\big)$,
- $\widetilde{Y}_t^N$ *a copy of* Y_t^N *where we replaced every semicircular variable by a free copy,*
- $Q \in \mathcal{F}_{d(p+1),q}$.

Then, for any N, *with* $\partial_{s,j}$ *the non-commutative differential as defined in 2.4 but with respect to* $(1-e^{-t_s})^{1/2}y_{s,j}$

$$\begin{aligned}
&\mathbb{E}\left[\tau_N\Big(Q\left(Y^N\right)\Big)\right]-\tau_N\Big(Q\left(Y\right)\Big)\\
&=\frac{1}{2N^2}\sum_{1\leq i,j\leq d}\int_0^\infty te^{-t}\ \mathbb{E}\Big[\tau_N\Big(\Big(D_j\left(\partial_i^1 D_i Q\right)\left(Y_t^N\right)\Big)\boxtimes\Big(D_j\left(\partial_i^2 D_i Q\right)\big(\widetilde{Y}_t^N\big)\Big)\Big)\Big]\\
&\qquad\qquad+e^{-t}\sum_s t_s\ \mathbb{E}\Big[\tau_N\Big(\Big(D_{s,j}\left(\partial_i^1 D_i Q\right)\left(Y_t^N\right)\Big)\boxtimes\Big(D_{s,j}\left(\partial_i^2 D_i Q\right)\big(\widetilde{Y}_t^N\big)\Big)\Big)\Big]dt.
\end{aligned}$$

First we need to prove the following technical lemma.

Lemma 3.3. *If* $R^{kN}, Y_1^{kN}, \dots, Y_p^{kN}$ *are* $p+1$ *independent vectors of* d *independent GUE matrices of size* kN, *then let*

$$S_k = \big(R^{kN}, Y_1^{kN}, \dots, Y_p^{kN}, X^N \otimes I_k, Z^N \otimes I_k\big)$$

With $P_{1,2} = I_N \otimes E_{1,2}$, $\mathbb{E}_k$ *the expectation with respect to* $(R^{kN}, Y_1^{kN}, \dots, Y_d^{kN})$ *given* $Q \in \mathcal{F}_{d(p+2),q}$, *we have that*

$$\lim_{k\to\infty} k^{3/2}\mathbb{E}_k\left[\tau_{kN}\left(Q(S_k)P_{1,2}\right)\right] = 0$$

Proof. Given $A_1, \dots, A_l, B_1, \dots, B_l \in \mathcal{A}_{d(p+2),q}$, we define the following quantity,

$$f_A(y) = \mathbb{E}_k\left[\tau_{kN}\left((A_1 e^{\mathbf{i}yB_1}\dots A_l e^{\mathbf{i}yB_l})(S_k)P_{1,2}\right)\right],$$

$$d_n(y) = \sup_{\sum_i \deg A_i \leq n,\ A_i \text{ monomials}} |f_A(y)|.$$

Thanks to Proposition 2.18, we know that there exists constants γ and D (depending on N,$\|X^N\|$ and $\|Z^N\|$) such that for any $n \leq \gamma k$, $|d_n(y)| \leq D^n$. Consequently we define

$$g(a,y) = \sum_{n \leq \gamma k/2} d_n(y) a^n.$$

Let $m = \sup_i \deg B_i$ and A be such that $\sum_i \deg A_i \leq n$, there exists a constant C_B which only depends on the coefficients of the B_i such that

$$\left| \frac{df_A(y)}{dy} \right| \leq C_B \; d_{n+m}(y).$$

Naturally we get that for any $y \in [0,1]$

$$|f_A(y)| \leq |f_A(0)| + C_B \int_0^y d_{n+m}(y) dy.$$

And by taking the supremum over A, we get that

$$d_n(y) \leq d_n(0) + C_B \int_0^y d_{n+m}(y).$$

Hence by summing over n, we have for a small enough,

$$g(a,y) \leq g(a,0) + C_B a^{-m} \left(\sum_{1 \leq i \leq m} (aD)^{\gamma k/2 + i - 1} + \int_0^y g(a,y) dy \right).$$

Thanks to Grönwall's inequality we get that there exist a constant κ such that for any $y \in [0,1]$,

$$g(a,y) \leq \left(g(a,0) + C_B a^{-m} \sum_{1 \leq i \leq m} (aD)^{\gamma k/2 + i - 1} \right) e^{y C_B a^{-m}}.$$

Thus for $a < 1/D$, we have

$$\limsup_{k \to \infty} k^{3/2} g(a,y) \leq e^{y C_B a^{-m}} \limsup_{k \to \infty} k^{3/2} g(a,0).$$

However, we have

$$g(a,0) = \sum_{n \leq \gamma k/2} a^n \sup_{A \text{ monomial, } \deg A \leq n} \left| \mathbb{E}_k \left[\tau_{kN} \left(A(S_k) P_{1,2} \right) \right] \right|.$$

We refer to the proof of Lemma 3.7 of [8] to prove that $\limsup_{k \to \infty} k^{3/2} g(a,0) = 0$ (with the notations of [8], it is the same thing as to show that $k^{3/2} f_{\gamma k/2}(a)$ converges towards 0). Hence for any A, B,

$$\limsup_{k \to \infty} k^{3/2} \left| \mathbb{E}_k \left[\tau_{kN} \left((A_1 e^{\mathbf{i} y B_1} \dots A_l e^{\mathbf{i} y B_l})(S_k) P_{1,2} \right) \right] \right| \leq a^{-\sum_i \deg A_i} \limsup_{k \to \infty} k^{3/2} g(a,1) = 0.$$

Hence the conclusion. ◻

Proof of Lemma 3.2. We have,

$$\mathbb{E}\left[\tau_N \Big(Q\left(Y^N\right) \Big) \right] - \tau_N \Big(Q\left(Y\right) \Big) = - \int_0^\infty \mathbb{E}\left[\frac{d}{dt} \tau_N \Big(Q\left(Y_t^N\right) \Big) \right] dt.$$

We can compute

$$\frac{d}{dt} \tau_N \Big(Q\left(Y_t^N\right) \Big) = \frac{e^{-t}}{2} \sum_i \tau_N \left(D_i Q\left(Y_t^N\right) \left(\frac{x_i}{(1-e^{-t})^{1/2}} - e^{t/2} X_i^N \right) \right)$$

Thus thanks to Gaussian integration by part (see (11)) and Schwinger-Dyson equations (see Proposition 2.11), we get that

$$\mathbb{E}\left[\frac{d}{dt}\tau_N\Big(Q\left(Y_t^N\right)\Big)\right] = \mathbb{E}\left[\frac{e^{-t}}{2}\sum_i\left(\tau_N\otimes\tau_N\Big(\partial_i D_i Q\left(Y_t^N\right)\Big) - \frac{1}{N}\sum_{u,v}\tau_N\Big(E_{u,v}\ \partial_i D_i Q\left(Y_t^N\right)\#E_{v,u}\Big)\right)\right]. \tag{13}$$

Let

$$\Lambda_{N,t} = \tau_N\otimes\tau_N\Big(\partial_i D_i Q\left(Y_t^N\right)\Big) - \frac{1}{N}\sum_{u,v}\tau_N\Big(E_{u,v}\ \partial_i D_i Q\left(Y_t^N\right)\#E_{v,u}\Big).$$

Thanks to Theorem 5.4.5 of [1], we have that if

$$Z_k = \Big(e^{-t}X^N\otimes I_k + (1-e^{-t})^{1/2}R^{kN}, (1-e^{-t_1})^{1/2}Y_1^{kN},\dots,(1-e^{-t_d})^{1/2}Y_d^{kN}, Z^N\otimes I_k\Big)$$

with R^{kN} and Y_s^{kN} being independent vectors of d independent GUE matrices. Then with $\mathbb{E}_k$ the expectation with respect to R^{kN} and Y_s^{kN},

$$\begin{aligned}\Lambda_{N,t} = \lim_{k\to\infty}\quad & E_k[\tau_{kN}]\otimes\mathbb{E}_k[\tau_{kN}]\Big(\partial_i D_i Q\left(Z_k\right)\Big) \\ & - \mathbb{E}_k\left[\frac{1}{N}\sum_{1\leq u,v\leq N}\tau_{kN}\Big(E_{u,v}\otimes I_k\ \partial_i D_i Q\left(Z_k\right)\#E_{v,u}\otimes I_k\Big)\right].\end{aligned} \tag{14}$$

For more information, we refer to Proposition 3.5 of [8]. See also the definition of $\mathcal{A}_N$ in Definition 2.3. Let A, B be matrices of $\mathbb{M}_N(\mathbb{C})\otimes\mathbb{M}_k(\mathbb{C})$, since $I_k = \sum_l E_{l,l}$ and $\tau_{kN}(M) = \sum_{a,b} g_a^*\otimes f_b^* M g_a\otimes f_b$,

$$\begin{aligned}&\frac{1}{N}\sum_{1\leq u,v\leq N}\tau_{kN}\Big(E_{u,v}\otimes I_k\ A\ E_{v,u}\otimes I_k\ B\Big)\\ &= \frac{1}{N}\sum_{1\leq u,v\leq N}\sum_{1\leq l,l'\leq k}\tau_{kN}\left(E_{u,v}\otimes E_{l,l}\ A\ E_{v,u}\otimes E_{l',l'}\ B\right)\\ &= \frac{1}{N^2k}\sum_{1\leq l,l'\leq k}\sum_{1\leq v\leq N} g_v^*\otimes f_l^*\ A\ g_v\otimes f_{l'}\sum_{1\leq u\leq N} g_u^*\otimes f_{l'}^*\ B\ g_u\otimes f_l\\ &= \frac{1}{k}\sum_{1\leq l,l'\leq k}\tau_N(I_N\otimes f_l^*\ A\ I_N\otimes f_{l'})\ \tau_N(I_N\otimes f_{l'}^*\ B\ I_N\otimes f_l)\\ &= k\sum_{1\leq l,l'\leq k}\tau_{kN}\big(A\ I_N\otimes E_{l',l}\big)\ \tau_{kN}\big(B\ I_N\otimes E_{l,l'}\big).\end{aligned}$$

Hence with convention $P_{l,l'} = I_N\otimes E_{l,l'}$, we have

$$\frac{1}{N}\sum_{1\leq u,v\leq N}\tau_{kN}\Big(E_{u,v}\otimes I_k\ \partial_i D_i Q\left(Z_k\right)\#E_{v,u}\otimes I_k\Big) = k\sum_{1\leq l,l'\leq k}\tau_{kN}\otimes\tau_{kN}\Big(\partial_i D_i Q\left(Z_k\right)\times P_{l',l}\otimes P_{l,l'}\Big) \tag{15}$$

Let $U\in\mathbb{M}_k(\mathbb{C})$ be a unitary matrix, then since for any i,

$$I_N\otimes U\ X_i^N\otimes I_k\ I_N\otimes U^* = X_i^N\otimes I_k,$$

$$I_N\otimes U\ Z_i^N\otimes I_k\ I_N\otimes U^* = Z_i^N\otimes I_k,$$

and that the law of R^{kN} and Y_j^{kN} is invariant by conjugation by a unitary matrix. We get that

$$\mathbb{E}_k\left[\tau_{kN}\Big(\partial_i^1 D_i Q\left(Z_k\right)P_{l',l}\Big)\right] = \mathbb{E}_k\left[(\tau_{kN}\otimes I_M)\Big(\partial_i^1 D_i Q\left(Z_k\right)I_N\otimes U^* E_{l',l}U\otimes I_M\Big)\right].$$

Thus if $l = l'$, we can pick U such that $U^* E_{l',l} U = E_{1,1}$, and if $l \neq l'$, we can pick U such that $U^* E_{l',l} U = E_{1,2}$. Hence,

$$\begin{aligned} & k \sum_{1 \leq l,l' \leq k} \mathbb{E}_k[\tau_{kN}] \otimes \mathbb{E}_k[\tau_{kN}] \Big(\partial_i D_i Q(Z_k) P_{l',l} \otimes P_{l,l'} \Big) \\ &= k^2 \mathbb{E}_k[\tau_{kN}] \otimes \mathbb{E}_k[\tau_{kN}] \Big(\partial_i D_i Q(Z_k) P_{1,1} \otimes P_{1,1} \Big) \\ &\quad + k^2(k-1) \mathbb{E}_k[\tau_{kN}] \otimes \mathbb{E}_k[\tau_{kN}] \Big(\partial_i D_i Q(Z_k) P_{1,2} \otimes P_{1,2} \Big). \end{aligned} \tag{16}$$

Similarly we also have,

$$\begin{aligned} & \mathbb{E}_k[\tau_{kN}] \otimes \mathbb{E}_k[\tau_{kN}] \Big(\partial_i D_i Q(Z_k) \Big) \\ &= \sum_{1 \leq l,l' \leq k} \mathbb{E}_k[\tau_{kN}] \otimes \mathbb{E}_k[\tau_{kN}] \Big(\partial_i D_i Q(Z_k) \ P_{l,l} \otimes P_{l',l'} \Big) \\ &= k^2 \ \mathbb{E}_k[\tau_{kN}] \otimes \mathbb{E}_k[\tau_{kN}] \Big(\partial_i D_i Q(Z_k) \ P_{1,1} \otimes P_{1,1} \Big). \end{aligned} \tag{17}$$

By combining equations (14),(15),(16) and (17), we get that

$$\begin{aligned} \Lambda_{N,t} = \lim_{k \to \infty} \quad - \Bigg\{ k \sum_{1 \leq l,l' \leq k} \mathbb{E}_k \Big[\tau_{kN} \otimes \tau_{kN} \Big(\partial_i D_i Q(Z_k) P_{l',l} \otimes P_{l,l'} \Big) \Big] \\ - \mathbb{E}_k[\tau_{kN}] \otimes \mathbb{E}_k[\tau_{kN}] \Big(\partial_i D_i Q(Z_k) P_{l',l} \otimes P_{l,l'} \Big) \\ + k^2(k-1) \mathbb{E}_k[\tau_{kN}] \otimes \mathbb{E}_k[\tau_{kN}] \Big(\partial_i D_i Q(Z_k) P_{1,2} \otimes P_{1,2} \Big) \Bigg\}. \end{aligned}$$

Thanks to Lemma 3.3, the last term converges towards 0. Consequently,

$$\begin{aligned} \Lambda_{N,t} = \lim_{k \to \infty} \quad - k \Bigg\{ \sum_{1 \leq l,l' \leq k} \mathbb{E}_k \Big[\tau_{kN} \otimes \tau_{kN} \Big(\partial_i D_i Q(Z_k) P_{l',l} \otimes P_{l,l'} \Big) \Big] \\ - \mathbb{E}_k[\tau_{kN}] \otimes \mathbb{E}_k[\tau_{kN}] \Big(\partial_i D_i Q(Z_k) P_{l',l} \otimes P_{l,l'} \Big) \Bigg\}. \end{aligned} \tag{18}$$

Let $T^{kN}, S^{kN}, T_j^{kN}, S_j^{kN}$ be independent vectors of d independent GUE random matrices. We set the following notations,

$$\begin{aligned} Z_{k,r}^1 = \Big(& e^{-t} X^N \otimes I_k + (e^{-r} - e^{-t})^{1/2} R^{kN} + (1 - e^{-r})^{1/2} S^{kN}, \\ & (1 - e^{-t_1})^{1/2} Y_1^{kN}, \dots, (1 - e^{-t_d})^{1/2} Y_d^{kN}, Z^N \otimes I_k \Big), \end{aligned}$$

$$\begin{aligned} Z_{k,r}^{1,s} = \Big(& e^{-t} X^N \otimes I_k + (1 - e^{-t})^{1/2} R^{kN}, (1 - e^{-t_1})^{1/2} Y_1^{kN}, \dots, \\ & (e^{-r} - e^{-t_s})^{1/2} Y_s^{kN} + (1 - e^{-r})^{1/2} S_s^{kN}, \dots, (1 - e^{-t_d})^{1/2} Y_d^{kN}, Z^N \otimes I_k \Big), \end{aligned}$$

And similarly we define $Z_{k,r}^2$ (respectively $Z_{k,r}^{2,s}$) but with T^{kN} (respectively T_j^{kN}) instead of S^{kN} (respectively S_j^{kN}). Thanks to Proposition 3.1, we get that

$$\begin{aligned} & k \sum_{1 \leq l,l' \leq k} \mathbb{E}_k \Big[\tau_{kN} \otimes \tau_{kN} \Big(\partial_i D_i Q(Z_k) P_{l',l} \otimes P_{l,l'} \Big) \Big] - \mathbb{E}_k[\tau_{kN}] \otimes \mathbb{E}_k[\tau_{kN}] \Big(\partial_i D_i Q(Z_k) P_{l',l} \otimes P_{l,l'} \Big) \\ &= \frac{1}{k^2 N^3} \sum_{1 \leq l,l' \leq k} \sum_{1 \leq j \leq d} \int_0^t \mathbb{E}\Big[\mathrm{Tr}_{kN} \Big(D_j \left(\partial_i^1 D_i Q\right) (Z_{k,r}^1) \ P_{l',l} \ \boxtimes \ D_j \left(\partial_i^2 D_i Q\right) (Z_{k,r}^2) \ P_{l,l'} \Big) \Big] dr \\ &\qquad + \sum_s \int_0^{t_s} \mathbb{E}\Big[\mathrm{Tr}_{kN} \Big(D_{s,j} \left(\partial_i^1 D_i Q\right) (Z_{k,r}^{1,j}) \ P_{l',l} \ \boxtimes \ D_{s,j} \left(\partial_i^2 D_i Q\right) (Z_{k,r}^{2,j}) \ P_{l,l'} \Big) \Big] dr, \end{aligned}$$

where $\partial_{s,j}$ is the non-commutative differential as defined in 2.4 but with respect to $(1-e^{-t_s})^{1/2} Y_{s,j}^{kN}$. Besides if A, B are matrices of $\mathbb{M}_N(\mathbb{C})\otimes\mathbb{M}_k(\mathbb{C})$, then $\sum_{1\leq l,l'\leq k} \mathrm{Tr}_{kN}\otimes I_M(AP_{l',l}BP_{l,l'}) = \mathrm{Tr}_N(\mathrm{Tr}_k\otimes I_N(A)\ \mathrm{Tr}_k\otimes I_N(B))$. Hence

$$
\begin{aligned}
& k \sum_{1\leq l,l'\leq k} \mathbb{E}_k\Big[\tau_{kN}\otimes\tau_{kN}\Big(\partial_i D_i Q(Z_k)P_{l',l}\otimes P_{l,l'}\Big)\Big] - \mathbb{E}_k[\tau_{kN}]\otimes\mathbb{E}_k[\tau_{kN}]\Big(\partial_i D_i Q(Z_k)P_{l',l}\otimes P_{l,l'}\Big) \\
&= \frac{1}{N^2}\sum_{1\leq j\leq d}\int_0^t \mathbb{E}_k\Big[\tau_N\Big((\tau_k\otimes I_N)\Big(D_j\left(\partial_i^1 D_i Q\right)(Z_{k,r}^1)\Big)\boxtimes(\tau_k\otimes I_N)\Big(D_j\left(\partial_i^2 D_i Q\right)(Z_{k,r}^2)\Big)\Big)\Big]dr \\
&\qquad + \sum_s \int_0^{t_s} \mathbb{E}_k\Big[\tau_N\Big((\tau_k\otimes I_N)\Big(D_{s,j}\left(\partial_i^1 D_i Q\right)(Z_{k,r}^{1,s})\Big)\boxtimes(\tau_k\otimes I_N)\Big(D_{s,j}\left(\partial_i^2 D_i Q\right)(Z_{k,r}^{2,s})\Big)\Big)\Big]dr.
\end{aligned}
$$

If $Q, T\in\mathcal{F}_{d(p+1),q}$ are evaluated in $Z_{k,r}^1, Z_{k,r}^2, Z_{k,r}^{1,s}$ or $Z_{k,r}^{2,s}$, then thanks to Proposition 3.1,

$$
\begin{aligned}
& \mathbb{E}_k\left[\tau_k\otimes I_N(Q)\tau_k\otimes I_N(T)\right] \\
&= \frac{1}{k^2}\sum_{1\leq i,j,r,s\leq N}\mathbb{E}_k\left[\mathrm{Tr}_{kN}(QI_k\otimes E_{j,i})\times E_{i,j}\ \ \mathrm{Tr}_{kN}(TI_k\otimes E_{s,r})\times E_{r,s}\right] \\
&= \frac{1}{k^2}\sum_{1\leq i,j,r,s\leq N}\mathbb{E}_k\left[\mathrm{Tr}_{kN}(QI_k\otimes E_{j,i})\ \ \mathrm{Tr}_{kN}(TI_k\otimes E_{s,r})\right]\times(E_{i,j}E_{r,s}) \\
&= \mathcal{O}(k^{-2}) + \frac{1}{k^2}\sum_{1\leq i,j,r,s\leq N}\Big(\mathbb{E}_k\left[\mathrm{Tr}_{kN}(QI_k\otimes E_{j,i})\right]\mathbb{E}_k\left[\mathrm{Tr}_{kN}(TI_k\otimes E_{s,r})\right]\Big)\otimes(E_{i,j}E_{r,s}) \\
&= \mathcal{O}(k^{-2}) + \mathbb{E}_k\left[\tau_k\otimes I_N(Q)\right]\mathbb{E}_k\left[\tau_k\otimes I_N(T)\right].
\end{aligned}
$$

Hence we get that

$$
\begin{aligned}
& k \sum_{1\leq l,l'\leq k} \mathbb{E}_k\Big[\tau_{kN}\otimes\tau_{kN}\Big(\partial_i D_i Q(Z_k)P_{l',l}\otimes P_{l,l'}\Big)\Big] - \mathbb{E}_k[\tau_{kN}]\otimes\mathbb{E}_k[\tau_{kN}]\Big(\partial_i D_i Q(Z_k)P_{l',l}\otimes P_{l,l'}\Big) \\
&= \frac{1}{N^2}\sum_{1\leq j\leq d}\int_0^t \tau_N\Big(\mathbb{E}_k\Big[(\tau_k\otimes I_N)\Big(D_j\left(\partial_i^1 D_i Q\right)(Z_{k,r}^1)\Big)\Big]\boxtimes\mathbb{E}_k\Big[(\tau_k\otimes I_N)\Big(D_j\left(\partial_i^2 D_i Q\right)(Z_{k,r}^2)\Big)\Big]\Big)dr \\
&\qquad + \sum_s \int_0^{t_s} \tau_N\Big(\mathbb{E}_k\Big[(\tau_k\otimes I_N)\Big(D_{s,j}\left(\partial_i^1 D_i Q\right)(Z_{k,r}^{1,s})\Big)\Big]\boxtimes\mathbb{E}_k\Big[(\tau_k\otimes I_N)\Big(D_{s,j}\left(\partial_i^2 D_i Q\right)(Z_{k,r}^{2,s})\Big)\Big]\Big)dr. \\
&\quad + \mathcal{O}(k^{-2}).
\end{aligned}
$$

However, $Z_{k,r}^1, Z_{k,r}^2, Z_{k,r}^{1,s}$ and $Z_{k,r}^{2,s}$ all have the same law as Z_k, thus

$$
\begin{aligned}
& k \sum_{1\leq l,l'\leq k} \mathbb{E}_k\Big[\tau_{kN}\otimes\tau_{kN}\Big(\partial_i D_i Q(Z_k)P_{l',l}\otimes P_{l,l'}\Big)\Big] - \mathbb{E}_k[\tau_{kN}]\otimes\mathbb{E}_k[\tau_{kN}]\Big(\partial_i D_i Q(Z_k)P_{l',l}\otimes P_{l,l'}\Big) \\
&= \frac{1}{N^2}\sum_{1\leq j\leq d} t\ \tau_N\Big(\mathbb{E}_k\Big[(\tau_k\otimes I_N)\Big(D_j\left(\partial_i^1 D_i Q\right)(Z_k)\Big)\Big]\boxtimes\mathbb{E}_k\Big[(\tau_k\otimes I_N)\Big(D_j\left(\partial_i^2 D_i Q\right)(Z_k)\Big)\Big]\Big) \\
&\qquad + \sum_s t_s\ \tau_N\Big(\mathbb{E}_k\Big[(\tau_k\otimes I_N)\Big(D_{s,j}\left(\partial_i^1 D_i Q\right)(Z_k)\Big)\Big]\boxtimes\mathbb{E}_k\Big[(\tau_k\otimes I_N)\Big(D_{s,j}\left(\partial_i^2 D_i Q\right)(Z_k)\Big)\Big]\Big). \\
&\quad + \mathcal{O}(k^{-2}).
\end{aligned}
$$

We can view GUE matrices of size kN as a matrix of size N with matrix coefficients. The diagonal coefficients are independent GUE matrices of size k multiplied by $N^{-1/2}$. The upper non-diagonal coefficients are independent random matrices of size k which have the same law as $(2N)^{-1/2}(X+\mathbf{i}Y)$ where X and Y are independent GUE matrices of size k, and the lower non-diagonal coefficients are the adjoints of the upper coefficients. We then define $\mathbf{x}^N$ and $\mathbf{y_j}^N$ as vectors of matrices whose diagonal coefficients are free semicirculars multiplied by $N^{-1/2}$, the upper non-diagonal coefficients are free between each other, free from the diagonal one, and they are of the form $(2N)^{-1/2}(a+\mathbf{i}b)$ where a and b are free semicirculars. Finally the lower non-diagonal coefficients are the adjoints of the upper coefficients. We also assume that semicirculars from different matrices are free and that all of those semicirculars live in a $\mathcal{C}^*$-algebra endowed with a trace τ. Thus if we define Z just like Z_k but with $\mathbf{x}^N, \mathbf{y_j}^N$ instead of R^{kN}, Y_j^{kN}, thanks to Theorem 5.4.2 from [1], we get that

$$\Lambda_{N,t} = -\frac{1}{N^2} \sum_{1\leq j\leq d} t\ \tau_N\Big((\tau\otimes I_N)\Big(D_j\left(\partial_i^1 D_i Q\right)(Z)\Big) \boxtimes (\tau\otimes I_N)\Big(D_j\left(\partial_i^2 D_i Q\right)(Z)\Big)\Big)$$
$$+\sum_s t_s\ \tau_N\Big((\tau\otimes I_N)\Big(D_{s,j}\left(\partial_i^1 D_i Q\right)(Z)\Big) \boxtimes (\tau\otimes I_N)\Big(D_{s,j}\left(\partial_i^2 D_i Q\right)(Z)\Big)\Big).$$

Let $\widetilde{Z}$ be a free copy of Z, that is such that we replaced every semicircular variable with a free copy. Then,

$$\Lambda_{N,t} = -\frac{1}{N^2} \sum_{1\leq j\leq d} t\ \tau\otimes\tau_N\Big(\Big(D_j\left(\partial_i^1 D_i Q\right)(Z)\Big) \boxtimes \Big(D_j\left(\partial_i^2 D_i Q\right)(\widetilde{Z})\Big)\Big)$$
$$+\sum_s t_s\ \tau\otimes\tau_N\Big(\Big(D_{s,j}\left(\partial_i^1 D_i Q\right)(Z)\Big) \boxtimes \Big(D_{s,j}\left(\partial_i^2 D_i Q\right)(\widetilde{Z})\Big)\Big).$$

But then if $\widetilde{Z_k}$ is an independent copy of Z_k, once again thanks to Theorem 5.4.2 from [1], we get that

$$\Lambda_{N,t} = -\frac{1}{N^2} \sum_{1\leq j\leq d} \lim_{k\to\infty} t\ \mathbb{E}_k\left[\tau_{kN}\Big(\Big(D_j\left(\partial_i^1 D_i Q\right)(Z_k)\Big) \boxtimes \Big(D_j\left(\partial_i^2 D_i Q\right)(\widetilde{Z_k})\Big)\Big)\right]$$
$$+\sum_s t_s\ \mathbb{E}_k\left[\tau_{kN}\Big(\Big(D_{s,j}\left(\partial_i^1 D_i Q\right)(Z_k)\Big) \boxtimes \Big(D_{s,j}\left(\partial_i^2 D_i Q\right)(\widetilde{Z_k})\Big)\Big)\right].$$

Hence if we define $\widetilde{Y}_t^N$ a copy of Y_t^N where we replaced every semicircular variables by a free copy, then by Theorem 5.4.5 from [1],

$$\Lambda_{N,t} = -\frac{1}{N^2} \sum_{1\leq j\leq d} t\ \tau_N\Big(\Big(D_j\left(\partial_i^1 D_i Q\right)(Y_t^N)\Big) \boxtimes \Big(D_j\left(\partial_i^2 D_i Q\right)(\widetilde{Y}_t^N)\Big)$$
$$+\sum_s t_s\ \tau_N\Big(\Big(D_{s,j}\left(\partial_i^1 D_i Q\right)(Y_t^N)\Big) \boxtimes \Big(D_{s,j}\left(\partial_i^2 D_i Q\right)(\widetilde{Y}_t^N)\Big)\Big).$$

Thus by using this result in equation (13), we have in conclusion

$$\mathbb{E}\left[\frac{d}{dt}\tau_N\Big(Q\left(Y_t^N\right)\Big)\right] = -\frac{e^{-t}}{2N^2} \sum_{1\leq i,j\leq d} t\ \mathbb{E}\Big[\tau_N\Big(\Big(D_j\left(\partial_i^1 D_i Q\right)(Y_t^N)\Big) \boxtimes \Big(D_j\left(\partial_i^2 D_i Q\right)(\widetilde{Y}_t^N)\Big)\Big]$$
$$+\sum_s t_s\ \mathbb{E}\Big[\tau_N\Big(\Big(D_{s,j}\left(\partial_i^1 D_i Q\right)(Y_t^N)\Big) \boxtimes \Big(D_{s,j}\left(\partial_i^2 D_i Q\right)(\widetilde{Y}_t^N)\Big)\Big)\Big].$$

□

3.3 Proof of Theorem 1.1

In this section we focus on proving Theorem 1.1 from which we deduce all of the important corollaries. It will mainly be a corollary of the following theorem, which is slightly stronger but less explicit. We refer to Definition 3.6 for the definition of L^{T_i} and Lemma 3.5 for the one of x^{T_i}.

Theorem 3.4. *Let the following objects be given,*

- $X^N = (X_1^N, \ldots, X_d^N)$ *independent GUE matrices of size* N*,*
- $Z^N = (Z_1^N, \ldots, Z_q^N)$ *deterministic matrices and their adjoints,*
- $P \in \mathcal{A}_{d,q}$ *a polynomial that we assume to be self-adjoint,*

- $f:\mathbb{R}\mapsto\mathbb{R}$ *such that there exists a measure on the real line* μ *with* $\int(1+y^{4(k+1)})\ \ d|\mu|(y)\ <+\infty$ *and for any* $x\in\mathbb{R}$,
$$f(x)=\int_{\mathbb{R}} e^{\mathbf{i}xy}\ d\mu(y)\ . \tag{19}$$

Then if we set,

$$\alpha_i^P(f,Z^N)=\int_{\mathbb{R}}\int_{[0,+\infty)^i}\tau_N\Big(\left(L^{T_1}\dots L^{T_i}\right)(e^{\mathbf{i}yP})(x^{T_i},Z^N)\Big)\ \ dt_1\dots dt_i\ d\mu(y),$$

and that we write $P=\sum_{1\leq i\leq Nb(P)}c_iM_i$ *where the* M_i *are monomials and* $c_i\in\mathbb{C}$, *if we set* $C_{\max}(P)=\max\{1,\max_i|c_i|\}$, *then there exist constants* C,K *and* c *independent of* P *such that with* $K_N=\max\{\left\|Z_1^N\right\|,\dots,\left\|Z_q^N\right\|,K\}$, *for any* N *and* $k\leq cN(\deg P)^{-1}$,

$$\left|\mathbb{E}\left[\tau_N\Big(f(P(X^N,Z^N))\Big)\right]-\sum_{0\leq i\leq k}\frac{1}{N^{2i}}\alpha_i^P(f,Z^N)\right| \tag{20}$$
$$\leq\frac{1}{N^{2k+2}}\int_{\mathbb{R}}(|y|+y^{4(k+1)})d|\mu|(y)\times\Big(C\times K_N^{\deg P}C_{\max}(P)Nb(P)(\deg P)^2\Big)^{4(k+1)}\times k^{3k}.$$

Besides if we define $\widehat{K}_N$ *like* K_N *but with* 2 *instead of* K, *then we have that for any* $j\in\mathbb{N}^*$,

$$\left|\alpha_j^P(f,Z^N)\right|\leq\int_{\mathbb{R}}(|y|+y^{4j})d|\mu|(y)\times\Big(C\times\widehat{K}_N^{\deg P}C_{\max}(P)Nb(P)(\deg P)^2\Big)^{4j}\times j^{3j}. \tag{21}$$

Finally if f *and* g *satisfies* (19) *and are bounded functions equal on a neighborhood of the spectrum of* $P(x,Z^N)$, *where* x *is a free semicircular system free from* $\mathbb{M}_N(\mathbb{C})$, *then for any* i, $\alpha_i^P(f,Z^N)=\alpha_i^P(g,Z^N)$. *In particular if* f *is a bounded function such that its support and the spectrum of* $P(x,Z^N)$ *are disjoint, then for any* i, $\alpha_i^P(f,Z^N)=0$.

The following lemma allows us to define the coefficients of the topological expansion by induction. It is basically a reformulation of Lemma 3.2 with the notations of Definitions 2.12 and 2.13. Although the notations in this formula are a bit heavy, such a formulation is necessary in order to get a better upper bound on the remainder term. It is the first step of the proof of Theorem 3.4.

Lemma 3.5. *Let* $x,y^1,\dots,y^{c_n}$ *be free semicircular system of* d *variables. Then with* $T_n=(t_1,\dots,t_n)$ *and* $I=\{s_1,\dots,s_n\}$,

$$X_{i,I}^{N,T_n}=e^{-\sum_{i=1}^n t_i/2}X_i^N+\sum_{l=0}^{n-1}e^{-\sum_{i=1}^l t_i/2}(1-e^{-t_{l+1}})^{1/2}y_i^{s_{l+1}},$$

$$x_{i,I}^{T_n}=e^{-\sum_{i=1}^n t_i/2}x_i+\sum_{l=0}^{n-1}e^{-\sum_{i=1}^l t_i/2}(1-e^{-t_{l+1}})^{1/2}y_i^{s_{l+1}}.$$

Given $Q\in\mathcal{F}_{d,q}^n$, *with* $X_1^{N,T_{n+1}}=\Big(X_{\{I,2c_n+1\}}^{N,T_{n+1}}\Big)_{I\in J_n}$ *and* $X_2^{N,T_{n+1}}=\Big(X_{\{I+c_n,2c_n+2\}}^{N,T_{n+1}}\Big)_{I\in J_n}$, *then with the convention of Definition 2.15,*

$$\begin{aligned}&\mathbb{E}\left[\tau_N\Big(Q\left(X^{N,T_n},Z^N\right)\Big)\right]-\tau_N\Big(Q\left(x^{T_n},Z^N\right)\Big)\\ &=\frac{1}{2N^2}\sum_{1\leq i,j\leq d}\int_0^\infty e^{-\sum_{i=1}^{n+1}t_i}\sum_{0\leq l\leq n}t_{l+1}\ e^{-\sum_{i=1}^l t_i}\sum_{A\in J_n^l}\mathbb{E}\Big[\tau_N\Big(\Big(D_{j,A}\left(\partial_i^1 D_iQ\right)(X_1^{N,T_{n+1}},Z^N)\Big)\\ &\qquad\boxtimes\Big(D_{j,A}\left(\partial_i^2 D_iQ\right)(X_2^{N,T_{n+1}},Z^N)\Big)\Big)\Big]dt_{n+1}.\end{aligned}$$

Proof. Let Y_t^N and $\widetilde{Y}_t^N$ be as in Lemma 3.2, $S \in \mathcal{F}_{d(p+1),q}$ such that $Q(X^{N,T_n}, Z^N) = S(Y_t^N)$, then $\partial_i D_i S(Y_t^N) = e^{-\sum_{i=1}^n t_i}\ \partial_i D_i Q(X^{N,T}, Z^N)$. Thus with the convention of Definition 2.15, we also have

$$e^{\sum_{i=1}^n t_i}\ D_j \partial_i^1 D_i S(Y_t^N) \boxtimes D_j \partial_i D_i S(\widetilde{Y}_t^N) = e^{-\sum_{i=1}^n t_i}\ D_{j,\emptyset} \partial_i^1 D_i S(X_1^{N,\{T_n,t\}}, Z^N) \boxtimes D_{j,\emptyset} \partial_i^2 D_i S(X_2^{N,\{T_n,t\}}, Z^N).$$

Thanks to remark 2.14, we know that given $s \in [1, c_n]$, there exists a unique $l \in [0, n-1]$ and $k_{l+2}, \dots, k_n$ such that for any $I = \{s_1, \dots, s_n\} \in J_n$, either for $i > l+1$, $s_i = k_i$ and $s_{l+1} = s$ or $s \notin I$. Hence for any s, there exists $A = \{s, k_{l+2}, \dots, k_n\} \in J_n^l$ such that

$$e^{\sum_{i=1}^n t_i}\ D_{s,j} \partial_i^1 D_i S(Y_t^N) \boxtimes D_{s,j} \partial_i D_i S(\widetilde{Y}_t^N) = e^{-\sum_{i=1}^l t_i}\ D_{j,A} \partial_i^1 D_i S(X_1^{N,\{T_n,t\}}, Z^N) \boxtimes D_{j,A} \partial_i^2 D_i S(X_2^{N,\{T_n,t\}}, Z^N).$$

Thus thanks to Lemma 3.2,

$$\begin{aligned}
&\mathbb{E}\left[\tau_N\Big(Q\left(X^{N,T_n}, Z^N\right)\Big)\right] - \tau_N\Big(Q\left(x^{N,T_n}, Z^N\right)\Big) \\
&= \frac{1}{2N^2} \sum_{1 \le i,j \le p} \int_0^\infty t e^{-t-2\sum_{i=1}^n t_i}\ \mathbb{E}\Big[\tau_N\Big(\Big(D_{j,\emptyset}\left(\partial_i^1 D_i Q\right)(X_1^{N,\{T_n,t\}}, Z^N)\Big) \\
&\qquad \boxtimes \Big(D_{j,\emptyset}\left(\partial_i^2 D_i Q\right)(X_2^{N,\{T_n,t\}}, Z^N)\Big)\Big)\Big] \\
&\qquad + e^{-t-\sum_{i=1}^n t_i} \sum_{0 \le l \le n-1} t_l e^{-\sum_{i=1}^l t_i} \sum_{A \in J_n^l} \mathbb{E}\Big[\tau_N\Big(\Big(D_{j,A}\left(\partial_i^1 D_i Q\right)(X_1^{N,\{T_n,t\}}, Z^N)\Big) \\
&\qquad \boxtimes \Big(D_{j,A}\left(\partial_i^2 D_i Q\right)(X_2^{N,\{T_n,t\}}, Z^N)\Big)\Big)\Big] dt.
\end{aligned}$$

Hence by renaming t in t_{n+1}, we get the conclusion.

□

This prompts us to define the following operator:

Definition 3.6. *We set $X_1 = \left(X_{\{I, 2c_n+1\}}\right)_{I \in J_n}$ and $X_2 = \left(X_{\{I+c_n, 2c_n+2\}}\right)_{I \in J_n}$. Then we define $L^{T_{n+1}} : \mathcal{F}_{d,q}^n \to \mathcal{F}_{d,q}^{n+1}$ by:*

$$L^{T_{n+1}}(Q) = \frac{1}{2} \sum_{\substack{1 \le i,j \le p \\ 0 \le l \le n}} t_{l+1}\ e^{-\sum_{i=1}^{n+1} t_i - \sum_{i=1}^l t_i} \sum_{A \in J_n^l} D_{j,A}\left(\partial_i^1 D_i Q\right)(X_1, Z) \boxtimes D_{j,A}\left(\partial_i^2 D_i Q\right)(X_2, Z). \quad (22)$$

In order to be perfectly rigorous, as in Definition 2.8, we have to define $\widetilde{L}^{T_{s+1}}_{\alpha_s,\beta_s,\gamma_s,\delta_s} : \mathcal{F}_{d,q}^n \to \mathcal{F}_{d,q}^{n+1}$ as in (22) but with $D_{\alpha_s,j,A}\left(\partial^1_{\gamma_s,i} D_{\delta_s,i} Q\right)(X_1, Z) \boxtimes D_{\beta_s,j,A}\left(\partial^2_{\gamma_s,i} D_{\delta_s,i} Q\right)(X_2, Z)$ (where we used notations of Definition 2.8) instead of $\left(\partial_i^1 D_i Q\right)(X_1, Z) \boxtimes D_{j,A}\left(\partial_i^2 D_i Q\right)(X_2, Z)$. Then given $x = (x_I)_{I \in J_n}$ and z elements of a $\mathcal{C}^*$-algebra, we define $L^{T_{n+1}}(Q)(x,z)$ as the integral of $\widetilde{L}^{T_{s+1}}_{\alpha_s,\beta_s,\gamma_s,\delta_s}(Q)(x,z)$ over $\alpha_s, \beta_s, \gamma_s$ and δ_s.

Thus we get directly the following proposition.

Proposition 3.7. *Let x be a free semicircular system, $(y^i)_{i \ge 1}$ be free semicircular systems free from x, and X^N be independent GUE matrices. We define X^{N,T_n} and x^{T_n} as in 3.5, then for any $Q \in \mathcal{F}_{d,q}$,*

$$\begin{aligned}
\mathbb{E}\left[\tau_N\Big(Q(X^N, Z^N)\Big)\right] &= \sum_{0 \le i \le k} \frac{1}{N^{2i}} \int_{[0,+\infty)^i} \tau_N\Big(\left(L^{T_i} \dots L^{T_1}\right)(Q)(x^{T_i}, Z^N)\Big)\ dt_1 \dots dt_i \\
&\quad + \frac{1}{N^{2(k+1)}} \int_{[0,+\infty)^{k+1}} \mathbb{E}\left[\tau_N\Big(\left(L^{T_{k+1}} \dots L^{T_1}\right)(Q)(X^{N,T_{k+1}}, Z^N)\Big)\right]\ dt_1 \dots dt_{k+1}.
\end{aligned}$$

Before giving the proof of Theorem 3.4, as mentioned in the introduction, the former proposition gives some insight in map enumeration.

Remark 3.8. *We say that a graph on a surface is a map if it is connected and its faces are homeomorphic to discs. It is of genus g if it can be embedded in a surface of genus g but not $g-1$. For an edge-colored graph on an orientated surface we say that a vertex is of type $q = X_{i_1} \dots X_{i_p}$ if it has degree p and when we look at the half-edges going out of it, starting from a distinguished one and going in the clockwise order the first half-edge is of color i_1, the second i_2, and so on. If $\mathcal{M}_g(X_{i_1} \dots X_{i_p})$ is the number of such maps of genus g with a single vertex, then given X_i^N independent GUE matrices*

$$\mathbb{E}\left[\frac{1}{N}\operatorname{Tr}_N\left(X_{i_1}^N \dots X_{i_p}^N\right)\right] = \sum_{g\in\mathbb{N}} \frac{1}{N^{2g}} \mathcal{M}_g(X_{i_1} \dots X_{i_p}).$$

For a proof we refer to [31] for the one matrix case and [5], chapter 22, for the multimatrix case. Thanks to Proposition 3.7, we immediately get that

$$\mathcal{M}_g(X_{i_1} \dots X_{i_p}) = \int\limits_{[0,+\infty)^g} \tau\Big(L^{T_g} \dots L^{T_1}\left(X_{i_1} \dots X_{i_p}\right)(x^{T_g})\Big)\ \ dt_1 \dots dt_g\ .$$

We can now prove Theorem 3.4.

Proof of Theorem 3.4. Thanks to Proposition 3.7, we immediately get that

$$\begin{aligned}\mathbb{E}\Big[\tau_N\Big(f(P(X^N, Z^N))\Big)\Big] &= \sum_{0\le i\le k} \frac{1}{N^{2i}} \alpha_i^P(f, Z^N) \\ &+ \frac{1}{N^{2(k+1)}} \int_{\mathbb{R}} \int\limits_{[0,+\infty)^{k+1}} \mathbb{E}\left[\tau_N\Big(\left(L^{T_{k+1}} \dots L^{T_1}\right)(e^{\mathbf{i}yP})(X^{N,T_{k+1}}, Z^N)\Big)\right]\ \ dt_1 \dots dt_{k+1}\ \ d\mu(y).\end{aligned}$$

All we need to do from now on is to get an estimate on the last line. To do so we use the following remark. Let $Q \in \mathcal{F}_{d,q}^n$, then we can write

$$Q = \sum_{1\le i\le Nb(Q)} c_i M_i$$

where $c_i \in \mathbb{C}$ and $M_i \in \mathcal{F}_{d,q}^n$ are monomials (not necessarily distinct). We also define $C_{max}(Q) = \max\{1, \sup_i |c_i|\}$. Since for any $I \in J^n$, $\left\|X_{i,I}^{N,T_n}\right\| \le 2 + \|X_i^N\|$, given $\mathcal{B}$ the union of $\left\{2 + \|X_i^N\|\right\}_{1\le i\le p}$ and $\left\{\|Z_j^N\|\right\}_{1\le j\le q}$, and D_N the maximum of this family, we get that

$$\left\|Q(X^{N,T_n}, Z^N)\right\| \le Nb(Q) \times C_{max}(Q) \times D_N^{\deg(Q)}. \tag{23}$$

It is worth noting that this upper bound is not optimal at all and heavily dependent on the decomposition chosen. Now let us consider $\widetilde{L}^{T_{s+1}}_{\alpha_s,\beta_s,\gamma_s,\delta_s}$ defined as in Definition 3.6. We also consider $\widetilde{\mathcal{F}}_{d,q}^n$ the $*$-algebra generated by $\mathcal{A}_{d,q}^n$ and the family

$$\left\{e^{\mathbf{i}\lambda y P(X_I)} \mid I \in J_n, \lambda \in [0,1]\right\}.$$

Then $\widetilde{L}^{T_{s+1}}_{\alpha_s,\beta_s,\gamma_s,\delta_s}$ send $\widetilde{\mathcal{F}}_{d,q}^s$ to $\widetilde{\mathcal{F}}_{d,q}^{s+1}$. Let $Q \in \widetilde{\mathcal{F}}_{d,q}^s$, then we get that

$$\deg\left(\widetilde{L}^{T_{s+1}}_{\alpha_s,\beta_s,\gamma_s,\delta_s}(Q)\right) \le \deg Q + 4\deg P,$$

$$C_{\max}\left(\widetilde{L}^{T_{s+1}}_{\alpha_s,\beta_s,\gamma_s,\delta_s}(Q)\right) \le \frac{\sum_{k=1}^s t_k}{2} e^{-\sum_{k=1}^{s+1} t_k}(1+|y|)^4\ C_{\max}(P)^4\ C_{\max}(Q),$$

$$Nb\left(\widetilde{L}^{T_{s+1}}_{\alpha_s,\beta_s,\gamma_s,\delta_s}(Q)\right) \le \deg(Q)(\deg Q + \deg P)(\deg Q + 2\deg P)(\deg Q + 3\deg P) \times (Nb(P)\deg P)^4 \times Nb(Q).$$

Thus if we define by induction $Q_0 = e^{\mathbf{i}yP}$, and $Q_{s+1} = \widetilde{L}^{T_{s+1}}_{\alpha_s,\beta_s,\gamma_s,\delta_s} Q_s$, by a straightforward induction we get that

$$\deg Q_s \le 4s \deg P \tag{24}$$

$$C_{\max}(Q_s) \leq \frac{\prod_{r=1}^s \sum_{k=1}^r t_k e^{-\sum_{k=1}^r t_k}}{2^s} (1+|y|)^{4s}\ C_{\max}(P)^{4s} \tag{25}$$

$$Nb(Q_s) \leq \Big(Nb(P)(\deg P)^2 \Big)^{4s} (4s)! \tag{26}$$

Actually since we have $D_{\delta_1,i} e^{\mathbf{i}yP} = \mathbf{i}y\ \partial_{\delta_1,i} P \tilde{\#} e^{\mathbf{i}yP}$, one can replace $(1+|y|)^{4s}$ in equation (25) by $|y|(1+|y|)^{4s-1}$. Thus thanks to (23), we get that

$$\begin{aligned}
&\left\| \widetilde{L}^{T_{k+1}}_{\alpha_{k+1},\beta_{k+1},\gamma_{k+1},\delta_{k+1}} \dots \widetilde{L}^{T_1}_{\alpha_1,\beta_1,\gamma_1,\delta_1} Q(X^{N,T_{k+1}}, Z^N) \right\| \\
&\leq \prod_{r=1}^{k+1} \frac{\sum_{s=1}^r t_s e^{-\sum_{s=1}^r t_s}}{2} \times \frac{|y|}{1+|y|} \times \Big((1+|y|) C_{\max}(P) Nb(P) (\deg P)^2 \Big)^{4(k+1)} (4(k+1))! \times D_N^{4(k+1)\deg P}
\end{aligned}$$

Consequently after integrating over $\alpha_s, \beta_s, \gamma_s, \delta_s$, we get that

$$\begin{aligned}
&\left| \int_{\mathbb{R}} \int_{[0,+\infty)^{k+1}} \mathbb{E}\left[\tau_N\Big(\left(L^{T_{k+1}} \dots L^{T_1} \right) (e^{\mathbf{i}yP})(X^{N,T_{k+1}}, Z^N) \Big) \right] \ dt_1 \dots dt_{k+1} \ \ d\mu(y) \right| \\
&\leq \int_{[0,+\infty)^{k+1}} \prod_{r=1}^{k+1} \frac{\sum_{s=1}^r t_s e^{-\sum_{s=1}^r t_s}}{2} dt_1 \dots dt_{k+1} \times \int_{\mathbb{R}} |y|(1+|y|)^{4k+3} d|\mu|(y) \\
&\qquad \times \Big(C_{\max}(P) Nb(P) (\deg P)^2 \Big)^{4(k+1)} (4(k+1))! \times \mathbb{E}\left[D_N^{4(k+1)\deg P} \right].
\end{aligned}$$

Besides

$$\begin{aligned}
\int_{[0,+\infty)^{k+1}} \prod_{r=1}^{k+1} \frac{\sum_{s=1}^r t_s e^{-\sum_{s=1}^r t_s}}{2} dt_1 \dots dt_{k+1} &= 2^{-k-1} \int_{0\leq t_1 \leq \dots \leq t_{k+1}} \prod_{r=1}^{k+1} t_r e^{-t_r} \quad dt_1 \dots dt_{k+1} \\
&\leq 2^{-k-1} \int_{0\leq t_1 \leq \dots \leq t_{k+1}} \prod_{r=1}^{k+1} e^{-t_r/2} \quad dt_1 \dots dt_{k+1} \\
&= \int_{0\leq t_1 \leq \dots \leq t_{k+1}} \prod_{r=1}^{k+1} e^{-t_r} \quad dt_1 \dots dt_{k+1} \\
&= \frac{1}{(k+1)!} \ ,
\end{aligned}$$

and

$$\int_{\mathbb{R}} |y|(1+|y|)^{4k+3} d|\mu|(y) \leq 2^{4k+3} \int_{\mathbb{R}} (|y| + y^{4(k+1)}) d|\mu|(y).$$

Thanks to Proposition 2.18 we can find constants K and c such that with $K_N = \max\{K, \left\|Z_1^N\right\|, \dots, \left\|Z_q^N\right\|\}$, then for any $k \leq c(\deg P)^{-1} N$,

$$\mathbb{E}\left[D_N^{4(k+1)\deg P} \right] \leq K_N^{4(k+1)\deg P}.$$

Thus thanks to Stirling formula, there exists a constant C such that

$$\begin{aligned}
&\left| \int_{\mathbb{R}} \int_{[0,+\infty)^{k+1}} \mathbb{E}\left[\tau_N\Big(\left(L^{T_{k+1}} \dots L^{T_1} \right) (e^{\mathbf{i}yP})(X^{N,T_{k+1}}, Z^N) \Big) \right] \ dt_1 \dots dt_{k+1} \ \ d\mu(y) \right| \\
&\leq \int_{\mathbb{R}} (|y| + y^{4(k+1)}) d|\mu|(y) \times \Big(C \times K_N^{\deg P} C_{\max}(P) Nb(P) (\deg P)^2 \Big)^{4(k+1)} \times k^{3k} \ .
\end{aligned}$$

Hence we get equation (20). We get equation (21) very similarly. Finally to prove the last affirmation, we only need to consider a function which takes value 0 on a neighborhood of the spectrum of $P(x, Z^N)$. Let X^{lN} be independent GUE matrices of size lN, then we get that for any k such that f is smooth enough, thanks to equation (20),

$$\mathbb{E}\left[\tau_N\Big(f(P(X^{lN}, Z^N \otimes I_l))\Big)\right] = \sum_{0\leq i\leq k} \frac{1}{(lN)^{2i}}\alpha_i^P(f, Z^N \otimes I_l) + \mathcal{O}(l^{-2(k+1)}).$$

But in the sense of Definition 2.1, for any i, $(x^{T_i}, Z^N \otimes I_l)$ and (x^{T_i}, Z^N) have the same distribution, hence

$$\mathbb{E}\left[\tau_N\Big(f(P(X^{lN}, Z^N \otimes I_l))\Big)\right] = \sum_{0\leq i\leq k} \frac{1}{(lN)^{2i}}\alpha_i^P(f, Z^N) + \mathcal{O}(l^{-2(k+1)}).$$

Consequently, if there exists i such that $\alpha_i^P(f, Z^N) \neq 0$, then we can find constants c and k (dependent on N) such that

$$\mathbb{E}\left[\tau_N\Big(f(P(X^{lN}, Z^N \otimes I_l))\Big)\right] \sim_{l\to\infty} c \times l^{-2k}. \tag{27}$$

We are going to show that the left hand side decays exponentially fast in l, hence proving a contradiction. Now if we set E the support of f, then

$$\left|\mathbb{E}\left[\tau_N\Big(f(P(X^{lN}, Z^N \otimes I_l))\Big)\right]\right| \leq \|f\|_\infty \mathbb{P}\left(\sigma\left(P(X^{lN}, Z^N \otimes I_l)\right) \cap E \neq \emptyset\right).$$

However thanks to Proposition 2.18, there exist constants A and B such that for any l,

$$\mathbb{P}\left(\left\|P(X^{lN}, Z^N \otimes I_l)\right\| \geq A\right) \leq e^{-Bl}.$$

Thus,

$$\left|\mathbb{E}\left[\tau_N\Big(f(P(X^{lN}, Z^N \otimes I_l))\Big)\right]\right| \leq \|f\|_\infty \Big(\mathbb{P}\left(\sigma\left(P(X^{lN}, Z^N \otimes I_l)\right) \cap E \cap [-A, A] \neq \emptyset\right) + e^{-Bl}\Big).$$

Let g be a $\mathcal{C}^\infty$-function, with compact support disjoint from the spectrum of $P(x, Z^N)$ such that $g_{|E\cap[-A,A]} = 1$. Then,

$$\left|\mathbb{E}\left[\tau_N\Big(f(P(X^{lN}, Z^N \otimes I_l))\Big)\right]\right| \leq \|f\|_\infty \mathbb{P}\left(\left\|g\left(P(X^{lN}, Z^N \otimes I_l)\right)\right\| \geq 1\right) + e^{-Bl}.$$

Since g is $\mathcal{C}^\infty$ and has compact support, thanks to the Fourier transform it satisfies (19) for any k, thus for any self-adjoint matrices U and V,

$$\begin{aligned}\|g(U) - g(V)\| &= \left\|\int y \int_0^1 e^{\mathbf{i}yU\alpha}(U - V)e^{\mathbf{i}yV(1-\alpha)} d\alpha d\mu(y)\right\| \\ &\leq \|U - V\| \int |y| d|\mu|(y).\end{aligned}$$

Hence there is a constant C_B such that for any self-adjoint matrices $X_i, Y_i \in \mathbb{M}_{lN}(\mathbb{C})$ whose operator norm is bounded by B,

$$\left\|g(P(X, Z^N)) - g(P(Y, Z^N))\right\| \leq C_B \sum_i \|X_i - Y_i\|.$$

Consequently, with a proof very similar to the one of Proposition 4.6 of [8], we get that there exist constant D and S such that for any $\delta > 0$,

$$\mathbb{P}\left(\left|\left\|g\left(P(X^{lN}, Z^N \otimes I_l)\right)\right\| - \mathbb{E}\left[\left\|g\left(P(X^{lN}, Z^N \otimes I_l)\right)\right\|\right]\right| \geq \delta + De^{-N}\right) \leq pe^{-2N} + e^{-S\delta^2 l}.$$

But then thanks to Theorem 1.6 of [3] and Weierstrass theorem, we know that almost surely $\left\|g\big(P(X^{lN}, Z^N \otimes I_l)\big)\right\|$ converges towards 0. Hence thanks to Proposition 2.18 and dominated convergence theorem, we get that

$\mathbb{E}\left[\left\|g\left(P(X^{lN}, Z^N \otimes I_l)\right)\right\|\right]$ also converges towards 0. Hence for l large enough, there exist a constant S such that

$$\mathbb{P}\left(\left\|g\left(P(X^{lN}, Z^N \otimes I_l)\right)\right\| \geq 1\right) \leq e^{-Sl}.$$

Consequently we get that there exist constants A and B such that

$$\left|\mathbb{E}\left[\tau_N\Big(f(P(X^{lN}, Z^N \otimes I_l))\Big)\right]\right| \leq Ae^{-Bl},$$

which is in contradiction with equation (27). Hence the conclusion.

□

We can now prove Theorem 1.1, the only difficulty of the proof is to use the hypothesis of smoothness to replace our function f by a function which satisfies (19) without losing too much on the constants.

Proof of Theorem 1.1. To begin with, let

$$h : x \to \begin{cases} e^{-x^{-4}-(1-x)^{-4}} & \text{if } x \in (0,1), \\ 0 & \text{else.} \end{cases}$$

Let H be the primitive of h which takes value 0 on $\mathbb{R}^-$ and renormalized such that it takes value 1 for $x \geq 1$. Then given a constant m one can define the function $g : x \to H(m+1-x)H(m+1+x)$ which takes value 1 on $[-m, m]$ and 0 outside of $(-m-1, m+1)$. Let B be the union over i of the events $\{\|X_i^N\| \geq D+\alpha^{-1}\}$ where D and α where defined in Proposition 2.18. Thus $\mathbb{P}(B) \leq pe^{-N}$. By adjusting the constant K defined in Theorem 3.4 we can always assume that it is larger than $D+\alpha^{-1}$, thus if for any i, $\|X_i^N\| \leq D+\alpha^{-1}$, $\|P(X^N, Z^N)\| \leq \mathbf{m}C_{\max}K_N^n$. We fix $m = \mathbf{m}C_{\max}K_N^n$, thus if $P(X^N, Z^N)$ has an eigenvalue outside of $[-m, m]$, necessarily $X^N \in B$. Thus

$$\mathbb{E}\left[\tau_N\Big(f(1-g)(P(X^N, Z^N))\Big)\right] \leq \|f\|_\infty \mathbb{P}(B) \leq \|f\|_\infty p \times e^{-N}. \tag{28}$$

Since fg has compact support and is a function of class $\mathcal{C}^{4(k+1)+2}$, we can take its Fourier transform and then invert it so that with the convention $\hat{h}(y) = \frac{1}{2\pi}\int_\mathbb{R} h(x)e^{-\mathbf{i}xy}dx$, we have

$$\forall x \in \mathbb{R}, \quad (fg)(x) = \int_\mathbb{R} e^{\mathbf{i}xy}\widehat{fg}(y)\, dy\ .$$

Besides, since if h has compact support bounded by $m+1$ then $\left\|\hat{h}\right\|_0 \leq \frac{1}{\pi}(m+1)\|h\|_0$, we have

$$\begin{aligned}
\int_\mathbb{R} (|y| + y^{4(k+1)}) \left|\widehat{fg}(y)\right|\, dy &\leq \int_\mathbb{R} \frac{\sum_{i=0}^{4(k+1)+2} |y|^i}{1+y^2} \left|\widehat{fg}(y)\right|\, dy \\
&\leq \int_\mathbb{R} \frac{\sum_{i=0}^{4(k+1)+2} \left|\widehat{(fg)^{(i)}}(y)\right|}{1+y^2}\, dy \\
&\leq \frac{1}{\pi}(m+1)\|fg\|_{\mathcal{C}^{4(k+1)+2}} \int_\mathbb{R} \frac{1}{1+y^2}\, dy \\
&\leq (m+1)\|fg\|_{\mathcal{C}^{4(k+1)+2}}\ ,
\end{aligned}$$

Hence fg satisfies the hypothesis of Theorem 3.4 with $\mu(dy) = \widehat{fg}(y)dy$. Therefore, combining with equation (28), by adjusting the constant C, we get that

$$\begin{aligned}
&\left|\mathbb{E}\left[\tau_N\Big(f(P(X^N, Z^N))\Big)\right] - \sum_{0\leq i\leq k} \frac{1}{N^{2i}}\alpha_i^P(fg, Z^N)\right| \\
&\leq \frac{1}{N^{2k+2}}\|fg\|_{\mathcal{C}^{4(k+1)+2}} \times \Big(C \times K_N^{\deg P} C_{\max}(P) N b(P)(\deg P)^2\Big)^{4(k+1)+1} \times k^{3k}.
\end{aligned}$$

Since the norm of the family Z^N is uniformly bounded over N, the second line is of order N^{-2k-2}. Hence if $(\beta_i)_{1\leq i\leq k}$ is a family of scalar such that $\mathbb{E}\left[\tau_N\Big(f(P(X^N,Z^N))\Big)\right] - \sum_{0\leq i\leq k} N^{-2i}\beta_i$ is also of order N^{-2k-2}, then so is $\sum_{0\leq i\leq k}(\alpha_i^P(fg,Z^N)-\beta_i)N^{-2i}$. Thus for any i, $\beta_i = \alpha_i^P(fg,Z^N)$. Thus since it does not depends on g, one can set $\alpha_i^P(f,Z^N) = \alpha_i^P(fg,Z^N)$.
Finally, one can write the j-th derivative of $x\to e^{-x^{-4}}$ on $\mathbb{R}^+$ as $x\to Q_j(x^{-1})e^{-x^{-4}}$ for some polynomial Q_j. By studying $Nb(Q_j), C_{\max}(Q_j)$ and $\deg(Q_j)$, as in the proof of Theorem 3.4, we get that the infinity norm of the j-th derivative of this function is smaller than $20^j j!(5j/4)^{5j/4}$. Hence by adjusting C and using Stirling formula,

$$\begin{aligned}&\left|\mathbb{E}\left[\tau_N\Big(f(P(X^N,Z^N))\Big)\right] - \sum_{0\leq i\leq k}\frac{1}{N^{2i}}\alpha_i^P(fg,Z^N)\right| \\ &\leq \frac{1}{N^{2k+2}}\|f\|_{\mathcal{C}^{4(k+1)+2}} \times \Big(C\times K_N^{\deg P}C_{\max}(P)Nb(P)(\deg P)^2\Big)^{4(k+1)+1}\times k^{12k}.\end{aligned}$$

The other points of the Theorem are a direct consequence of Theorem 3.4.

□

4 Consequences of Theorem 3.4

4.1 Proof of corollary 1.2

Let g be a non-negative $\mathcal{C}^\infty$-function which takes value 0 on $(-\infty,1/2]$, 1 on $[1,\infty)$ and in $[0,1]$ elsewhere. For any $a,b\in\mathbb{R}$, we define $h^\varepsilon_{[a,b]}: x\mapsto g(\varepsilon^{-1}(x-a))g(-\varepsilon^{-1}(x-b))$. Then let $\mathcal{I}$ be the collection of connected components of the complementary set of $\sigma(P(x,A^N))$. Then we define

$$h^\varepsilon = \sum_{I\in\mathcal{I}} h^\varepsilon_I.$$

This function is well-defined since the spectrum of $P(x,A^N)$ is compact, hence its complementary set has a finite number of connected components of measure larger than ε. And since if $b-a\leq\varepsilon$, $h^\varepsilon_{[a,b]}=0$, the sum over $I\in\mathcal{I}$ is actually a finite sum. Besides, we have that

$$\mathbb{P}\left(\sigma(P(X^N,A^N))\not\subset\sigma(P(x,A^N))+\varepsilon\right) \;\leq\; \mathbb{P}\left(\left\|h^\varepsilon(P(X^N,A^N))\right\|\geq 1\right) \;\leq\; \mathbb{E}\left[\mathrm{Tr}_N\left(h^\varepsilon(P(X^N,A^N))\right)\right].$$

Besides thanks to Theorem 1.1 since the spectrum of $P(x,A^N)$ and the support of h^ε are disjoint, and that the operator norm of the matrices A^N is uniformly bounded over N, for any $k\in\mathbb{N}$, we get that there is a constant C_k such that for any ε and for N large enough,

$$\mathbb{E}\left[\mathrm{Tr}_N\left(h^\varepsilon(P(X^N,A^N))\right)\right]\leq C_k\frac{\varepsilon^{-4k-2}}{N^{2k-1}}.$$

Thus if we set $\varepsilon = N^{-\alpha}$ with $\alpha<1/2$, then by fixing k large enough we get that

$$\mathbb{P}\left(\sigma(P(X^N,A^N))\not\subset\sigma(P(x,A^N))+N^{-\alpha}\right) = \mathcal{O}(N^{-2}).$$

Hence the conclusion by Borel-Cantelli lemma.

4.2 Proof of Corollary 1.3

Firstly, we need the following lemma.

Lemma 4.1. *Let g be a $\mathcal{C}^\infty$ function which takes value 0 on $(-\infty,1/2]$ and value 1 on $[1,\infty)$, and in $[0,1]$ otherwise. We set $f_\varepsilon: t\mapsto g(\varepsilon^{-1}(t-\alpha))$ with $\alpha=\left\|PP^*(x,A^N)\right\|$, then there exist constants C and c such that for any $k\leq cN$, $\varepsilon>0$ and N,*

$$\mathbb{E}\left[\mathrm{Tr}_N\left(f_\varepsilon(PP^*(X^N,A^N))\right)\right]\leq N\times C^k\left(\frac{\varepsilon^{-2}}{N}\right)^{2k}k^{12k}\ .$$

Proof. To estimate the above expectation we once again want to use the Fourier transform with a few refinements to have an optimal estimate with respect to ε. We set $f_\varepsilon^\kappa : t \mapsto g(\varepsilon^{-1}(t-\alpha))g(\varepsilon^{-1}(\kappa-t)+1)$ with $\kappa > \alpha$. Since g has compact support and is sufficiently smooth we can apply Theorem 3.2. Setting $h : t \mapsto g(t - \varepsilon^{-1}\alpha)g(\varepsilon^{-1}\kappa + 1 - t) = f_\varepsilon^\kappa(\varepsilon t)$, we have for $k \in \mathbb{N}^*$,

$$\begin{aligned}
\int y^{4k} |\hat{f_\varepsilon^\kappa}(y)| \, dy &= \frac{1}{2\pi} \int y^{4k} \left| \int g(\varepsilon^{-1}(t-\alpha))g(\varepsilon^{-1}(\kappa-t)+1) e^{-\mathbf{i}yt} \, dt \right| \, dy \\
&= \frac{1}{2\pi} \int y^{4k} \left| \int h(t) e^{-\mathbf{i}y\varepsilon t} \, \varepsilon dt \right| \, dy \\
&= \frac{\varepsilon^{-4k}}{2\pi} \int y^{4k} \left| \int h(t) e^{-\mathbf{i}yt} \, dt \right| \, dy \\
&\leq \frac{\varepsilon^{-4k}}{2\pi} \int \frac{1}{1+y^2} \, dy \int (|h^{(4k)}(t)| + |h^{(4k+2)}(t)|) \, dt \\
&\leq \varepsilon^{-4k} \left(\left\| h^{(4k)} \right\|_\infty + \left\| h^{(4k+2)} \right\|_\infty \right) .
\end{aligned}$$

In the last line we used the fact the support of the derivatives of h are included in $[\varepsilon^{-1}\alpha, \varepsilon^{-1}\alpha + 1] \cup [\varepsilon^{-1}\kappa, \varepsilon^{-1}\kappa + 1]$. Thus thanks to Theorem 3.4 and by using the function g defined in the proof of Theorem 1.1, we get that there exist constants C and c such that for any $k \leq cN$, for any $\kappa > \alpha$,

$$\mathbb{E}\left[\mathrm{Tr}_N \left(f_\varepsilon^\kappa(PP^*(X^N, A^N)) \right) \right] \leq N \times C^k \left(\frac{\varepsilon^{-2}}{N} \right)^{2k} k^{12k} .$$

Hence the conclusion by dominated convergence theorem.

□

Consequently, with $x_+ = \max(x, 0)$, for any $r > 0$,

$$\begin{aligned}
\mathbb{E}\left[\left(\|PP^*(X^N, A^N)\| - \|PP^*(x, A^N)\| \right)_+ \right] &\leq r + \int_r^\infty \mathbb{P}\left(\|PP^*(X^N, A^N)\| \geq \|PP^*(x, A^N)\| + \varepsilon \right) \, d\varepsilon \\
&\leq r + \int_r^\infty \mathbb{P}\left(\mathrm{Tr}_N \left(f_\varepsilon(PP^*(X^N, A^N)) \right) \geq 1 \right) d\varepsilon \\
&\leq r + \int_r^\infty \mathbb{E}\left[\mathrm{Tr}_N \left(f_\varepsilon(PP^*(X^N, A^N)) \right) \right] d\varepsilon \\
&\leq r + r \times N \times C^k \left(\frac{r^{-2}}{N} \right)^{2k} k^{12k}.
\end{aligned}$$

Thus by taking $r = N^{-a}$, we get that

$$\mathbb{E}\left[\left(\|PP^*(X^N, A^N)\| - \|PP^*(x, A^N)\| \right)_+ \right] \leq N^{-a} \times \left(1 + N^{1+2k(2a-1)} \times C^k k^{12k} \right).$$

Now we want to pick a and k such that $N^{1+2k(2a-1)} \times C^k k^{12k}$ is bounded by 1 uniformly over N (while keeping in mind that k has to be an integer). It is sufficient to pick a and k such that,

$$\ln C + \frac{\ln N}{k} + 12 \ln k \leq 2 \times (1 - 2a) \ln N.$$

We fix $k = \lceil \ln N \rceil$, then we need to pick a such that

$$\ln C + 1 + 12 \ln \lceil \ln N \rceil \leq 2 \times (1 - 2a) \ln N.$$

Which means that we can pick $a = \frac{1}{2} - 4 \frac{\ln \ln N}{\ln N}$, and for N large enough,

$$\mathbb{E}\left[\left(\|PP^*(X^N, A^N)\| - \|PP^*(x, A^N)\| \right)_+ \right] \leq 2N^{-a} = \frac{2 \ln^4 N}{\sqrt{N}}.$$

Thanks to Proposition 4.6 from [8] we have that for $N \geq \ln(p)$, there exist constants K and D such that

$$\mathbb{P}\Big(\mid \|P^*P(X^N,A^N)\| - \mathbb{E}\left[\|P^*P(X^N,A^N)\|\right] \mid \geq \delta + Ke^{-N} \Big) \leq e^{-N} + e^{-D\delta^2 N}.$$

Thus with $x_+ = \max(x,0)$, we immediately get that

$$\begin{aligned}
&\mathbb{P}\left(\left(\|P(X^N,A^N)\| - \|P(x,A^N)\|\right) \geq \frac{\delta + Ke^{-N} + \frac{2\ln^4 N}{\sqrt{N}}}{\|P(x,A^N)\|} \right) \\
&\leq \mathbb{P}\left(\left(\|P(X^N,A^N)\| - \|P(x,A^N)\|\right) \times \left(\|P(X^N,A^N)\| + \|P(x,A^N)\|\right) \geq \delta + Ke^{-N} + \frac{2\ln^4 N}{\sqrt{N}} \right) \\
&\leq \mathbb{P}\left(\|P^*P(X^N,A^N)\| - \|P^*P(x,A^N)\| \geq \delta + Ke^{-N} + \frac{2\ln^4 N}{\sqrt{N}} \right) \\
&\leq \mathbb{P}\Big(\|P^*P(X^N,A^N)\| - \|P^*P(x,A^N)\| \geq \delta + Ke^{-N} + \mathbb{E}\left[\left(\|P^*P(X^N,A^N)\| - \|P^*P(x,A^N)\|\right)_+ \right] \Big) \\
&\leq \mathbb{P}\Big(\|P^*P(X^N,A^N)\| - \|P^*P(x,A^N)\| \geq \delta + Ke^{-N} + \mathbb{E}\left[\|P^*P(X^N,A^N)\| - \|P^*P(x,A^N)\| \right] \Big) \\
&\leq \mathbb{P}\Big(\|P^*P(X^N,A^N)\| - \mathbb{E}\left[\|P^*P(X^N,A^N)\|\right] \geq \delta + Ke^{-N} \Big) \\
&\leq e^{-N} + e^{-D\delta^2 N}.
\end{aligned}$$

Since the family $(A^N)_N$ is uniformly bounded over N, so is the sequence $(\|P(x,A^N)\|)_N$, hence by replacing δ by $D^{-1/2}\delta$, we get that there is a constant C such that

$$\mathbb{P}\left(\|P(X^N,A^N)\| - \|P(x,A^N)\| \geq C\left(\delta + \frac{\ln^4 N}{\sqrt{N}}\right) \right) \leq e^{-N} + e^{-\delta^2 N}.$$

Finally by replacing δ by $\frac{\ln^4 N}{\sqrt{N}}\delta$, we get that

$$\mathbb{P}\left(\frac{\sqrt{N}}{\ln^4 N}\left(\|P(X^N,A^N)\| - \|P(x,A^N)\|\right) \geq C\,(\delta+1) \right) \leq e^{-N} + e^{-\delta^2 \ln^8 N}.$$

Acknowledgements

The author would like to thanks his PhD supervisors Benoît Collins and Alice Guionnet for proof-reading this paper and their continuous help, as well as Mikael de la Salle for helpful discussion. The author was partially supported by a MEXT JASSO fellowship and Labex Milyon (ANR-10-LABX-0070) of Université de Lyon.

References

[1] G. W. Anderson, A. Guionnet, and O. Zeitouni, *An introduction to random matrices*, volume 118 of *Cambridge Studies in Advanced Mathematics*, Cambridge University Press, Cambridge, 2010.

[2] U. Haagerup and S. Thorbjørnsen, A new application of random matrices: $\mathrm{Ext}(C^*_{\mathrm{red}}(\mathbb{F}_2))$ is not a group, *Ann. of Math.*, 162(2), 711–775, 2005.

[3] C. Male, The norm of polynomials in large random and deterministic matrices. With an appendix by Dimitri Shlyakhtenko, *Probab. Theory Related Fields* 154(3-4), 477-532, 2012.

[4] G.J. Murphy, *C*-Algebras and Operator Theory*, Elsevier Science, 1990.

[5] A. Nica and R. Speicher, *Lectures on the combinatorics of free probability*, volume 335 of *London Mathematical Society Lecture Note Series*, Cambridge University Press, Cambridge, 2006.

[6] C. A. Tracy and H. Widom, Level spacing distributions and the Airy kernel, *Comm. Math. Phys.*, 159, 151-174, 1994.

[7] D. Voiculescu, Limit laws for random matrices and free products, *Invent. Math.*, 104(1), 201–220, 1991.

[8] B. Collins, A. Guionnet and F. Parraud, On the operator norm of non-commutative polynomials in deterministic matrices and iid GUE matrices, 2019.

[9] F. Parraud, On the operator norm of non-commutative polynomials in deterministic matrices and iid Haar unitary matrices, 2020.

[10] U. Haagerup and S. Thorbjørnsen, Asymptotic expansions for the Gaussian Unitary Ensemble, *Infinite Dimensional Analysis Quantum Probability and Related Topics*, 15(1), 2010.

[11] S. Albeverio, L. A. Pastur and M. Shcherbina, On the $1/n$ expansion for some unitary invariant ensembles of random matrices, *Comm. Math. Physics*, 224, 271–305, 2001.

[12] G. Borot and A. Guionnet, Asymptotic expansion of β-matrix models in the one-cut regime, *Comm. Math. Phys.*, 317(2), 447–483, 2013.

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

[14] G. Borot, B. Eynard and N. Orantin, Abstract loop equations, topological recursion and new applications, *Commun. Number The-ory Phys.*, 9(1), 51–187, 2015.

[15] G. Borot, A. Guionnet and K. Kozlowski, Large N asymptotic expansion for mean field models with Coulomb gas interaction, *Int. Math. Res. Not. IMRN*, 20, 10451–10524, 2015.

[16] G. Borot, A. Guionnet and K. Kozlowski, *Asymptotic expansion of a partition function related to the sinh-model*, Mathematical Physics Studies, Springer, 2016.

[17] L. O. Chekhov, B. Eynard and O. Marchal, Topological expansion of the β-ensemble model and quantum algebraic geometry in the sectorwise approach, *Theoret. and Math. Phys.*, 166(2), 141–185, 2011.

[18] L. Chekhov and B. Eynard, Matrix eigenvalue model: Feynman graph technique for all genera, *J. High Energy Phys.*, (12), 2006.

[19] B. Collins, A. Guionnet and E. Maurel-Segala, Asymptotics of unitary and orthogonal matrix integrals, *Adv. Math.*, 222(1), 172–215, 2009.

[20] G. Parisi, E. Brézin, C. Itzykson and J. B. Zuber, Planar diagrams, *Comm. Math. Phys.*, 59, 35–51, 1978.

[21] B. Eynard and N. Orantin, Invariants of algebraic curves and topological expansion, *Commun. Number Theory Phys.*, 1(2), 347–452, 2007.

[22] B. Eynard, Topological expansion for the 1-Hermitian matrix model correlation functions, *J. High Energy Phys.*, (11), 2004.

[23] A. Guionnet and E. Maurel-Segala, Combinatorial aspects of matrix models, *Alea* 1, 241–279, 2006.

[24] A. Guionnet and E. Maurel-Segala, Second order asymptotics for matrix models, *Ann. Probab.*, 35(6), 2160–2212, 2007.

[25] A. Guionnet and J. Novak, Asymptotics of unitary multimatrix models: The Schwinger-Dyson lattice and topological recursion, *Journal of Functional Analysis*, 268, 2014.

[26] E. Maurel-Segala, High order asymptotics of matrix models and enumeration of maps, arXiv:math/0608192v1 [math.PR], 2006.

[27] M. Shcherbina, *Asymptotic expansions for β-matrix models and their applications to the universality conjecture, Random matrix theory, interacting particle systems, and integrable systems*, volume 65 of *Mathematical Sciences Research Institute Publications*, Cambridge University Press, New York, 463–482, 2014.

[28] G. t'Hooft, Magnetic monopoles in unified gauge theories, *Nuclear Phys.*, B79, 276–284, 1974.

[29] F. Bekerman, A. Figalli andA. Guionnet, Transport Maps for β-Matrix Models and Universality, *Commun. Math. Phys.*, 338, 589–619, 2015.

[30] A. Figalli and A. Guionnet, Universality in several-matrix models via approximate transport maps, *Acta Mathematica*, 217, 2014.

[31] J. Harer and D. Zagier, The Euler characteristic of the moduli space of curves, *Invent. Math.*, 85, 457–485, 1986.

[32] A. Zvonkin, Matrix integrals and map enumeration: an accessible introduction, *Math. Comput. Modelling*, 26, 281–304, 1997.

[33] L. Chekhov and B. Eynard, Hermitian matrix model free energy: Feynman graph technique for all genera, *J. High Energy Phys.*, 14, 2006.

[34] B. Eynard and N. Orantin, Topological expansion of the 2-matrix model correlation functions: diagrammatic rules for a residue formula, *J. High Energy Phys.*, 034, 2005.

[35] F. David, Loop equations and non perturbative effects in two-dimensional quantum gravity, *Mod. Phys. Lett.*, 1990.

[36] V. A. Kazakov, The appearance of matter fields from quantum fluctuations of 2D-gravity, *Mod. Phys. Lett.*, 4(22), 2125–2139, 1989.

[37] J. Ambjørn, L. Chekhov, C. F. Kristjansen and Y. Makeenko, Matrix model calculations beyond the spherical limit, *Nucl.Phys.*, B404, 127–172, 1993.

[38] N. Ercolani and K. McLaughlin, Asymptotics of the partition function for random matrices via Riemann-Hilbert techniques, and applications to graphical enumeration, *Int. Math. Res.*, 14. 2002.