MOMENT METHODS ON COMPACT GROUPS: WEINGARTEN CALCULUS AND ITS APPLICATIONS

BENOÎT COLLINS

ABSTRACT

A fundamental property of compact groups and compact quantum groups is the existence and uniqueness of a left and right invariant probability—the Haar measure. This is a very natural playground for classical and quantum probability, provided that it is possible to compute its moments. Weingarten calculus addresses this question in a systematic way. The purpose of this manuscript is to survey recent developments, describe some salient theoretical properties of Weingarten functions, as well as applications of this calculus to random matrix theory, quantum probability and algebra, mathematical physics, and operator algebras.

MATHEMATICS SUBJECT CLASSIFICATION 2020

Primary 15B52; Secondary 60B20, 28C10, 43A05, 46L54, 22D20

KEYWORDS

Haar measure on compact groups, Weingarten calculus, asymptotic freeness, quantum information theory, quantum groups, Schur–Weyl duality, Norm estimates, random tensors

INTERNATIONAL CONGRESS OF MATHEMATICIANS 2022 JULY 6–14

Proc. Int. Cong. Math. 2022, Vol. 4, pp. 3142–3164
DOI 10.4171/ICM2022/114

Published by EMS Press

1. INTRODUCTION

One of the key properties of a compact group G is that it admits a unique left and right invariant probability measure μ_G. It is called the *Haar measure*, and we refer to [15] for reference. In other words, $\mu_G(G) = 1$, and for any Borel subset A of G and $g \in G$, $\mu_G(Ag) = \mu_G(gA) = \mu_G(A)$, where $Ag = \{hg, h \in A\}$ and $gA = \{gh, h \in A\}$. The left and right invariance together with the uniqueness of μ_G readily imply that $\mu_G(A^{-1}) = \mu_G(A)$. The standard proofs of existence of the Haar measure are not constructive. In the more general context of locally compact groups, a left (resp. right) invariant measure exists, too. It is finite if and only if the group is compact and uniqueness is up to a nonnegative scalar multiple. In addition, the left and right Haar measures need not be the same. For locally compact groups, a classical proof of existence imitates the construction of the Lebesgue measure on $\mathbb{R}$ and resorts to outer measures. In the specific case of compact groups, a fixed-point argument can be applied. Either way, in both cases, the proof of existence is not constructive, in the sense that it does not tell us *how to integrate functions*. Weingarten calculus is about addressing this problem systematically. *Which functions* one wants to integrate needs, of course, to be clarified. We focus on the case of matrix groups, for which there are very natural candidates, namely polynomials in coordinate functions.

We recast this problem as the question of *computing the moments* of the Haar measure. Recall that, for a real random variable X, its moments are by definition the sequence $\mathbb{E}(X^k)$, $k \geq 0$—whenever they are defined. If the variable is vector-valued in $\mathbb{R}^n$, i.e., $X = (X_1, \ldots, X_n)$, then the moments are the numbers $\mathbb{E}(X_1^{k_1} \cdots X_n^{k_n})$, $k_1, \ldots k_n \geq 0$. Naturally, the existence of moments is not granted and is subject to the integrability of the functions. In the case of matrix compact groups, we have $G \subset \mathbb{M}_n(\mathbb{C}) = \mathbb{R}^{2n^2}$ therefore we may consider that the random variable we are studying is a random vector in $\mathbb{R}^{2n^2}$ whose distribution is the Haar measure with respect to the above inclusion. In this sense, we are really considering a moment problem. For this reason, we do not consider only coordinate functions, but also their complex conjugates in our moment problem.

The goal of this note is to provide an account of Weingarten calculus and in particular its multiple applications, with emphasis on the moment aspects and applications. From the point of view of the theory, there have been many approaches to computing integral of functions with respect to the Haar measure. We enumerate here a few important ones:

(1) Historically, the first nontrivial functions computed are arguably Fourier transforms, e.g., the Harish-Chandra integral [51]. The literature is huge and started from the initial papers of Harish-Chandra and Itzykson Zuber until now, however, we do not elaborate too much on this field as we focus on polynomial integrals. These techniques involve representation theory, symplectic geometry, and complex analysis. We refer to [62] for a recent approach and to the bibliography therein for references.

(2) Geometric techniques are natural because, when compact groups are manifolds, the measure can be described locally with differential geometry. They are effi-

cient for small groups. We refer, for example, to [3] for such techniques and gaussianization methods, with application to quantum groups. Geometry is also useful to compute specific functions, such as polynomials in one row or column with respect to orthogonal or unitary groups.

(3) Probability, changes of variables, and stochastic calculus are natural tools to try to compute the moments of Haar measures. For example, Rains in [68] used Brownian motion on compact groups and the fact that the Haar measure is the unique invariant measure to compute a complete set of relations. Subsequently, Lévy, Dahlqvist, Kemp, and the author have made progress on understanding the unitary multiplicative Brownian version of Weingarten calculus in [21,56].

(4) Representation theory has always been ubiquitous in the quest for calculating the Haar measure. A first important set of applications can be found by [45], but results were already available by [17,44,73].

(5) Combinatorial interpretations of the Haar measure in some specific cases were initiated in [18]. In another direction, there was the notable work of [18]. Subsequently, new combinatorial techniques were developed in [20,24], and we refer to [26] for substantial generalizations. We also refer [59] for modern interpretations and applications to geometric group theory.

As for the applications, they can be found in a considerable number of areas, including theoretical physics (2D quantum gravity, matrix integrals, random tensors), mathematical physics (quantum information theory, quantum spin chains), operator algebras (free probability), probability (limit theorems), representation theory, statistics, finance, machine learning, and group theory. The foundations of Weingarten calculus, as well as its applications, keep expanding rapidly, and this manuscript is a subjective snapshot of the-state-of-the-art. This introduction is followed by Section 2 that contains the foundations and theoretical results about the Weingarten functions. Section 3 investigates "simple" asymptotics of Weingarten functions and applications to random matrix theory. Section 4 deals with "higher order" asymptotics and applications to mathematical physics. Section 5 considers "uniform" asymptotics and applications to functional analysis, whereas the last section contains concluding remarks and perspectives.

2. WEINGARTEN CALCULUS

2.1. Notation

On the complex matrix algebra $\mathbb{M}_n(\mathbb{C})$, we denote by $\overline{A}$ the entrywise conjugate of a matrix A and $A^* = \overline{A}^t$ the adjoint. In the sequel we work with a compact matrix group G, i.e., a subgroup of $\mathrm{GL}_n(\mathbb{C})$ of invertible complex matrices, that is compact for the induced topology. It is known that such a group is conjugate inside $\mathrm{GL}_n(\mathbb{C})$ to the unitary group $\mathcal{U}_n = \{U, UU^* = U^*U = 1_n\}$. Writing an element U of $\mathcal{U}_n$ as a matrix

$U = (u_{ij})_{i,j\in\{1,\ldots,n\}}$, we view the entries u_{ij} as polynomial functions $\mathcal{U}_n \to \mathbb{C}$. As functions, they form a $*$-algebra—the $*$-operation being the complex conjugation. By construction, they are separating for $\mathcal{U}_n$, therefore, by Weierstrass' theorem, the $*$-algebra generated by $u_{ij}, i, j \in \{1, \ldots, n\}$, which is the algebra of polynomial functions on $\mathcal{U}_n$, is a dense subalgebra for the sup norm in the algebra of continuous functions on G.

By Riesz' theorem, understanding the Haar measure boils down to understanding $\int_{U\in G} f(U)d\mu_G(U)$ for any continuous function. By density and linearity, it is actually enough to be able to calculate systematically

$$\int_{U\in G} u_{i_1 j_1} \ldots u_{i_k j_k} \overline{u_{i'_1 j'_1} \ldots u_{i'_{k'} j'_{k'}}} d\mu_G(U).$$

No answer was known in full generality until a systematic development was initiated in [20, 40]. However, in the particular case of $\mathcal{U}_n, \mathcal{O}_n$, an algorithm to calculate a development in large n was devised in [44, 73], with further improvements by [66], and character expansions were obtained in [17], however, these approaches are largely independent. Likewise, Woronowicz obtained a formula for the moments of characters in the case of quantum groups in [74]. Interestingly, motivated by probability questions, the same formula was rediscovered independently by Diaconis–Shashahani [45] in the particular case of compact matrix groups.

2.2. Fundamental formula

Although the partial answers to the question of computing moments were rather involved, the general answer turns out, in hindsight, to be surprisingly simple, so we describe it here. We also refer to [32] for an invitation to the theory. We first start with the following notation: for an element $U = (u_{ij}) \in G \subset \mathbb{M}_n(\mathbb{C})$, $\overline{U}$ is the entrywise conjugate, i.e., $\overline{U} = (\overline{U_{ij}})$. Since U is unitary, $\overline{U}$ is unitary, too. We denote by $V = \mathbb{C}^n$ the fundamental representation of G, and $\overline{V}$ the contragredient representation. For a general representation W of G, $\mathrm{Fix}(G, W)$ is the vector subspace of W of fixed points under the action of G, i.e., $\mathrm{Fix}(G, W) = \{x \in W, \forall U \in G, Ux = x\}$. Finally, we fix two integers k, k', and set

$$Z_G = \int_{U\in G} U^{\otimes k} \otimes \overline{U}^{\otimes k'} d\mu_G(U),$$

and abbreviate $\mathrm{Fix}(G, V^{\otimes k} \otimes \overline{V}^{\otimes k'})$ into $\mathrm{Fix}(G, k, k')$.

Proposition 2.1. *The matrix Z_G is the orthogonal projection onto* $\mathrm{Fix}(G, k, k')$.

Proof. Since the distribution of U and UU' is the same for any fixed $U' \in G$, it implies that for any $U \in G$, $Z_G = Z_G \cdot U^{\otimes k} \otimes \overline{U^{\otimes k'}}$. Integrating once more over U gives the fact that Z_G is a projection. The fact that the map $U \to U^{-1} = U^*$ preserves the Haar measure implies that $Z_G = Z_G^*$. From the definition of invariance, for $x \in \mathrm{Fix}(G, k, k')$ and for any $U \in G$ one has $U^{\otimes k} \otimes \overline{U^{\otimes k'}} \cdot x = x$. Integrating with respect to the Haar measure of G gives $Z_G \cdot x = x$. Finally, take x outside $\mathrm{Fix}(G, k, k')$. It means that there exists U such that

$$U^{\otimes k} \otimes \overline{U^{\otimes k'}} x \neq x.$$

However, $||U^{\otimes k} \otimes \overline{U^{\otimes k'}} x||_2 = ||x||_2$. Thanks to the strict convexity of the Euclidean ball, after averaging over the Haar measure, we necessarily get $||Z_G x||_2 < ||x||_2$, which implies that x is not in $\mathrm{Im}(Z_G)$. Therefore we proved that $\mathrm{Im}(Z_G) = \mathrm{Fix}(G, k, k')$. ∎

From this, we can deduce an integration formula as soon as we have a generating family $y_1, \dots, y_l$ for $\mathrm{Fix}(G, k, k')$ (for any k, k'). Let

$$\mathrm{Gr} = (g_{ij})_{i,j \in \{1,\dots,l\}}$$

be its Gram matrix, i.e., $g_{ij} = \langle y_i, y_j \rangle$ and $W = (w_{ij})$ the pseudoinverse of Gr. Let $E_1, \dots, E_n$ be the canonical orthonormal basis of $V = \mathbb{C}^n$. Let k be a number and we consider the tensor space $V^{\otimes k}$ with its canonical orthogonal basis $E_I = e_{i_1} \otimes \cdots \otimes e_{i_k}$, where $I = (i_1, \dots, i_k)$ is a multiindex in $\{1, \dots, n\}^k$. Let $I = (i_1, \dots, i_k, i'_1, \dots, i'_{k'})$, $J = (j_1, \dots, j_k, j'_1, \dots, j'_{k'})$ be $k + k'$-indices, i.e.. elements of $\{1, \dots, n\}^{k+k'}$. Then

Theorem 2.2.

$$\begin{aligned}\int_{U \in G} u_{i_1 j_1} \dots u_{i_k j_k} \overline{u_{i'_1 j'_1} \dots u_{i'_{k'} j'_{k'}}} d\mu_G(U) &= \langle Z_G, E_I \otimes E_J \rangle \\ &= \sum_{i,j \in \{1,\dots,l\}} \langle E_I, y_i \rangle \langle y_j, E_J \rangle w_{ij}.\end{aligned}$$

2.3. Examples with classical groups

For interesting applications to be derived, the following conditions must be met:

(1) $y_1, \dots, y_l$ must be easy to describe.

(2) Gr should be easy to compute—and if possible, its inverse, the Weingarten matrix, too.

(3) $\langle E_I, y_i \rangle$ should be easy to compute.

Let us describe some fundamental examples. Let $P_2(k)$ be the collection of *pair partitions* on $\{1, \dots, k\}$ ($P_2(k)$ is empty if k is odd, and its cardinal is $1 \cdots (k-1) = k!!$ if k is even). Typically, a partition $\pi \in P_2(k)$ consists of $k/2$ blocks of cardinal 2, $\pi = \{V_1, \dots, V_{k/2}\}$, and we call $\delta_{\pi,I}$ the multiindex Kronecker function whose value is 1 if, for any block $V = \{k < k'\}$ of π, $i_k = i_{k'}$, and zero in all other cases. Likewise, we call $E_\pi = \sum_I E_I \delta_{\pi,I}$.

In [40], we obtained a complete solution to computing moments of Haar integrals for $\mathcal{O}_n, \mathcal{U}_n, \mathcal{S}p_n$. The following theorem describes this method. For convenience, we stick to the case of $\mathcal{O}_n, \mathcal{U}_n$.

Theorem 2.3. *The entries of* Gr *are* $\langle E_\pi, E_{\pi'} \rangle = n^{\mathrm{loops}(\pi,\pi')}$*, and we have* $\langle E_I, E_\pi \rangle = \delta_{\pi,I}$.

- *(The orthogonal case) For* $\mathcal{O}_n$*,* $E_\pi, \pi \in P_2(k)$ *is a generating family of the image of* $Z_{\mathcal{O}_n}$.
- *(The unitary case) Thanks to commutativity, and setting* $2k' = k$*, we consider the subset of* $P_2(k)$ *of pair partitions such that each block pairs one of the first* k'

elements with one of the last k' elements. This set is in natural bijection with the permutations $S_{k'}$, and it is the generating family of the image of $Z_{\mathcal{U}_n}$.

Proof. The first two points are direct calculations. The last two points are a reformulation of Schur–Weyl duality, respectively in the case of the unitary group and of the orthogonal group (see. e.g., [46]). ■

2.4. Example with Quantum groups

We finish the general theory of Weingarten calculus with a quick excursion through compact matrix quantum groups. For the theory of compact quantum groups, we refer to [74,75]. The subtlety for quantum groups is that in general we can not capture all representations with just $U^{\otimes k} \otimes \overline{U^{\otimes k'}}$ because U and $\overline{U}$ fail to commute in general. The theory of Tannaka–Krein duality for compact quantum groups is completely developed, and in order to get a completely general formula, we must instead consider

$$U^{\otimes k_1} \otimes \overline{U^{\otimes k_1'}} \otimes \cdots \otimes U^{\otimes k_p} \otimes \overline{U^{\otimes k_p'}}.$$

Let us just illustrate the theory with the *free quantum orthogonal group* O_n^+. It was introduced by Wang in [72], and its Tannaka–Krein dual was computed by Banica in [1]. Its algebra of polynomial functions $\mathbb{C}(O_n^+)$ is the noncommutative unital $*$-algebra generated by n^2 self-adjoint elements u_{ij} that satisfy the relation $\sum_k u_{ik}u_{jk} = \delta_{ij}1$ and $\sum_k u_{ki}u_{kj} = \delta_{ij}1$. Note that the abelianized version of this unital $*$-algebra is the $*$-algebra of polynomial functions on $\mathcal{O}_n$, which explains why it is called the free orthogonal quantum group. There exists a unital $*$-algebra homomorphism, called the coproduct $\Delta : \mathbb{C}(O_n^+) \to \mathbb{C}(O_n^+) \otimes \mathbb{C}(O_n^+)$ defined on generators by $\Delta u_{ij} = \sum_k u_{ik} \otimes u_{kj}$, and a unique linear functional $\mu : \mathbb{C}(O_n^+) \to \mathbb{C}$ such that $\mu(1) = 1$ and

$$(\mu \otimes Id)\Delta = 1\mu, \qquad (Id \otimes \mu)\Delta = 1\mu.$$

This functional is known as the Haar state, and it extends the notion of Haar measure on compact groups. Although the whole definition is completely algebraic, the proofs rely on functional analysis and operator algebras.

However, the calculation of the Haar state is purely algebraic and just relies on the notion of noncrossing pair partitions, denoted by $NC_2(k)$, which are a subset of $P_2(k)$ defined as follows. A partition π of $P_2(k)$ is noncrossing—and therefore in $NC_2(k)$ if any two of its blocks $\{i,j\}$ and $\{i',j'\}$ fail to satisfy the crossing relations $i < j, i' < j'$, $i < i' < j < j'$. This notion was found to be of crucial use for free probability by Speicher, see, e.g., [64]. The following theorem is a particular case of a series of results that can be found in [4]:

Theorem 2.4. *In the case of O_n^+, for $U^{\otimes k}$, the complete solution follows from the following result: $E_\pi, \pi \in NC_2(k)$ is a generating family of the image of $Z_{O_n^+}$.*

Note that, since $U = \overline{U}$, it is enough to consider $U^{\otimes k}$ to compute fully the Haar measure. We refer to [2,5,6] for applications of classical Weingarten functions to quantum groups, and to [3,4,8] for further developments of quantum Weingarten theory.

2.5. Representation theoretic formulas

A representation theoretic approach to Weingarten calculus is available for many families of groups, including unitary, orthogonal, and symplectic groups. Here we only describe the unitary group, and for the others, we refer to [30, 60].

Call S_k the symmetric group and consider its *group algebra* $\mathbb{C}[S_k]$—the unital $*$-algebra whose basis as a vector space is $\lambda_\sigma, \sigma \in S_k$, and endowed with the multiplication $\lambda_\sigma \lambda_\tau = \lambda_{\sigma\tau}$ and the $*$-structure $\lambda_\sigma^* = \lambda_{\sigma^{-1}}$. We follow standard representation-theoretic notation, see, e.g., [19] and $\lambda \vdash k$ denotes a Young diagram λ has k boxes; $\lambda \vdash k$ enumerates both the conjugacy classes of S_k and its irreducible representations. The symmetric group S_k acts on the set $\{1, \dots, k\}$, and in turn, by leg permutation on $(\mathbb{C}^n)^{\otimes k}$, which induces an algebra morphism $\mathbb{C}[S_k] \to \mathbb{M}_n(\mathbb{C})^{\otimes k}$. By Schur–Weyl duality, λ describes also irreducible polynomial representations of the unitary group $\mathcal{U}_n$ if its length is less than n and in this context, V_λ stands for the associated representation of the unitary group. For a permutation $\sigma \in S_k$, we call $\#\sigma$ the number of cycles (or loops) in its cycle product decomposition (counting fixed points). Consider the function

$$G = \sum_{\sigma \in S_k} n^{\#\sigma} \lambda_\sigma,$$

and its pseudoinverse $W = G^{-1} = \sum_{\sigma \in S_k} w(\sigma)\lambda_\sigma$. The following result was observed by the author and Śniady in [40] and it provides the link between representation theory and Weingarten calculus:

Theorem 2.5. *G is positive in $\mathbb{C}[S_k]$. In addition, we have $w(\sigma, \tau) = w(\tau\sigma^{-1})$, which we rename as $\mathrm{Wg}(n, \tau\sigma^{-1})$, and the following character expansion:*

$$\mathrm{Wg}(n, \sigma) = \frac{1}{k!^2} \sum_{\lambda \vdash k} \frac{\chi_\lambda(e)^2 \chi_\lambda(\sigma)}{\dim V_\lambda}.$$

Proof. Consider the action of S_k on $(\mathbb{C}^n)^{\otimes k}$ by leg permutation. It extends to a unital $*$-algebra morphism $\phi : \mathbb{C}[S_k] \to \mathbb{M}_n(\mathbb{C})^{\otimes k}$. By inspection, for $A \in \mathbb{C}[S_k]$, $\mathrm{Tr}[\phi(A)] = \tau(GA)$, where τ is the regular trace $\tau(\lambda_g) = \delta_{g,e}$. The positivity of τ implies that of G which proves positivity. The remaining points follow from the fact that G is central and by a character formula. ∎

2.6. Combinatorial formulations

Let us write formally $n^{-k} G = \lambda_e + \sum_{\sigma \in S_k - \{e\}} n^{\#\sigma - k} \lambda_\sigma$. It follows that as a power series in n^{-1},

$$n^k W = \lambda_e + \sum_{p \geq 1} (-1)^p \Big(\sum_{\sigma \in S_k - \{e\}} n^{\#\sigma - k} \lambda_\sigma \Big)^p.$$

Reading through the coefficients of this series gives a combinatorial formula for Wg in the unitary case. Such formulas were first found in [20], and we refer to [26] for substantial generalizations. See also [59] for other interpretations, as well as [18].

However, this formula is signed, and therefore impractical for the quest of uniform asymptotics. In a series of works, Novak and coworkers in [47–49,61] came with a very interesting solution to this problem which we describe below. It relies on Jucys Murphy elements, which are the following elements of $\mathbb{C}[S_k]$: $J_i = \sum_{j>i} \lambda_{(ij)}$. The following important result was observed:

$$G = (n + J_1) \cdots (n + J_{k-1}).$$

This follows from the fact that every permutation σ has a unique factorization as

$$\sigma = (i_1 j_1) \cdots (i_l j_l)$$

with the property $i_p < j_p$ and $j_p < j_{p+1}$.

This prompts us to define $P(\sigma, l)$ to be the set of solutions to the equation $\sigma = (i_1 j_1) \cdots (i_l j_l)$ with $i_p < j_p$, $j_p \le j_{p+1}$. The number of solutions to this problem is related to Hurwitz numbers, for details we refer, for example, to [26] and to the above references. From this we have the following theorem:

Theorem 2.6. *For $\sigma \in S_k$, we have the expansion*

$$\mathrm{Wg}(n, \sigma) = n^{-k} \sum_{l \ge 0} \# P(\sigma, l)(-n^{-1})^l. \tag{2.1}$$

The first strategy to compute the Weingarten formula was initiated in [73]. Let us outline it. We can write $\mathrm{Wg}(n, \sigma) = \int u_{11} \cdots u_{kk} \overline{u_{1\sigma 1} \cdots u_{k\sigma k}}$. Indeed, when considering the integral on the right-hand side in Theorems 2.2 and 2.3, the only pairing appearing corresponds to $\mathrm{Wg}(n, \sigma)$. Replacing the first row index of u and $\overline{u}$ by i and summing over i, we are to evaluate

$$\begin{aligned}\sum_{i=1}^{n} \int u_{i1} \cdots u_{kk} \overline{u_{i\sigma(1)} \cdots u_{k\sigma(k)}} &= \delta_{1\sigma(1)} \int u_{22} \cdots u_{kk} \overline{u_{2\sigma(2)} \cdots u_{k\sigma(k)}} \\ &= n\, \mathrm{Wg}(n, \sigma) + \sum_{i=2}^{l} \mathrm{Wg}\big(n, (1i)\sigma\big),\end{aligned} \tag{2.2}$$

where the first equality follows from orthogonality and the second from repeated use of the Weingarten formula. The second line provides an iterative technique to compute $\mathrm{Wg}(n, \sigma)$ both numerically and combinatorially. Historically, this is the idea of Weingarten, and in [73] he proved that the collection of all relations obtained above determine uniquely Wg for k fixed, n large enough.

In [31], we revisited his argument and figured out that these equations can be interpreted as a fixed point problem and a path counting formula, both formally and numerically. We got theoretical mileage from this approach and obtained new theoretical results, such as

Theorem 2.7. *All unitary Weingarten functions and their derivatives are monotone on* (k, ∞).

The unavoidability of Weingarten's historical argument becomes blatant when one studies quantum Weingarten function. Partial results about their asymptotics were obtained in [8], however, the asymptotics were not optimal for all entries. On the other hand, motivated by the study of planar algebras, Vaughan Jones asked us the following question: considering the canonical basis of the Temperley Lieb algebra $\mathrm{TL}_k(n)$, are the coefficients of the dual basis all nonzero when expressed in the canonical basis? For notations, we refer to our paper [16]. One motivation for this question is that the dual element of the identity is a multiple of the Jones–Wenzl projection.

Observing that this question is equivalent, up to a global factor, to the problem of computing the Weingarten function for O_n^+, and realizing that representation theory did not give tractable formulas in this case, we revisited the original idea of Weingarten and proved the following result, answering a series of open questions of Jones:

Theorem 2.8. *The quantum O_n^+ Weingarten function is never zero on the noncritical interval $[2, \infty)$, and it is monotone.*

Our proof actually provides explicit formulas for a Laurent expansion of the free Wg in the neighborhood of $n = \infty$, as a generating series of paths on graphs.

3. ASYMPTOTICS AND PROPERTIES OF WEINGARTEN FUNCTIONS

In this section, we are interested in the following problem. For a given permutation $\sigma \in S_k$, what is the behavior as $n \to \infty$ of $\mathrm{Wg}(n, \sigma)$? This function is rational as soon as $n \geq k$, and even elementary observations about its asymptotics have nontrivial applications in analysis. In the forthcoming subsections, we refine iteratively our study of the asymptotics, and derive each time new applications. Similar results have been obtained for most sequences of classical compact groups, but we focus here mostly on $\mathcal{U}_n$ and $\mathcal{O}_n$, and refer to the literature for other compact groups.

3.1. First order for identity Weingarten coefficients and Borel theorems

Let us first setup notations related to *noncommutative probability spaces* and of *convergence in distribution* in a noncommutative sense. A noncommutative probability space (NCPS) is a unital $*$-algebra $\mathcal{A}$ together with a state τ ($\tau : \mathcal{A} \to \mathbb{C}$ is linear, $\tau(1) = 1$, and $\tau(xx^*) \geq 0$ for any x). In general, we will assume *traciality*, $\tau(ab) = \tau(ba)$ for all a, b.

Assume we have a family of NCPS $(\mathcal{A}_n, \tau_n)$, a limiting object $(\mathcal{A}, \tau)$ and a d-tuple $(x_n^1, \ldots, x_n^d) \in \mathcal{A}_n^d$. We say that this d-tuple of *noncommutative random variables converges in distribution* to $(x^1, \ldots, x^d) \in \mathcal{A}^d$ iff for any sequence $i_1, \ldots, i_k$ of indices in $\{1, \ldots, d\}$,

$$\tau_n(x_n^{i_1} \cdots x_n^{i_k}) \to \tau(x^{i_1} \cdots x^{i_k}).$$

In the abelian case this corresponds to a convergence in moments (which is not in general the convergence in distribution); however, in the noncommutative framework, it is usually called convergence in noncommutative distribution, cf. [71]. The following result was proved in [42] in the classical case, and [4] in the quantum case:

Theorem 3.1. *Consider a sequence of vectors* $(A_1^n, \dots, A_r^n)$ *in* $\mathbb{M}_n(\mathbb{R})$ *such that the matrix* $(\mathrm{tr}(A_i A_j^t))$ *converges to* A*, and a* $\mathcal{O}_n$*-Haar distributed random variable* U_n*. Then, as* $n \to \infty$*, the sequence random vectors*

$$\big(\mathrm{Tr}(A_1^n U_n), \dots, \mathrm{Tr}(A_r^n U_n)\big)$$

converges in moments (and in distribution) to a Gaussian real vector of covariance A*. If we assume instead* U_n *to be in* O_n^+*, then* $(\mathrm{Tr}(A_1^n U_n), \dots, \mathrm{Tr}(A_r^n U_n))$ *converges in noncommutative distribution to a free semicircular family of covariance* A*.*

The proof relies on two ingredients. Firstly, for all examples considered so far, $\mathrm{Gr} = n^k \cdot 1_l(1 + O(n^{-1}))$, which implies that $W = \mathrm{Gr}^{-1} = n^{-k} 1_l(1 + O(n^{-1}))$. By inspection, it turns out that in the above theorem, the only entries of W that contribute asymptotically are the diagonal ones, and one can conclude with the classical (resp. the free) Wick theorem.

3.2. Other leading orders for Weingarten coefficients

The asymptotics obtained in the previous section are sharp only for the diagonal coefficients, however, they already yield nontrivial limit theorems. For more refined theorems, it is, however, necessary to obtain sharp asymptotics for all Weingarten coefficients. In the case of $\mathcal{U}_n$, sharp asymptotics can be deduced from the following

Theorem 3.2. *In the case of the full cycle in* S_k*, we have the following explicit formula:*

$$\mathrm{Wg}\big(n, (1 \cdots k)\big) = \frac{(-1)^{k+1} c_k}{(n-k+1)\cdots(n+k-1)},$$

where $c_k = (k+1)^{-1}\binom{2k}{k}$ *is the Catalan number. In addition,* Wg *is almost multiplicative in the following sense: if* σ *is a disjoint product of two permutations* $\sigma = \sigma_1 \sqcup \sigma_2$ *then*

$$\mathrm{Wg}(n, \sigma) = \mathrm{Wg}(n, \sigma_1)\,\mathrm{Wg}(n, \sigma_2)\big(1 + O(n^{-2})\big).$$

This result defines recursively a function $\mathrm{Moeb} : \bigsqcup_{k \geq 1} S_k \mapsto \mathbb{Z} - \{0\}$ satisfying

$$\mathrm{Wg}(n, \sigma) = n^{-k-|\sigma|}\mathrm{Moeb}(\sigma)\big(1 + O(n^{-2})\big).$$

This function was actually already introduced by Biane in [12], and it is closely related to Speicher's noncrossing Möbius function on the incidence algebra of the lattice of noncrossing partitions—see, e.g., [64]. Similar results are available for the orthogonal and symplectic group, we refer to [41]. Finally, let us mention that the asymptotics Weingarten function for the unitary group are the object of intense study; see, for example, [59, 69].

3.3. Classical asymptotic freeness

Weingarten calculus allows answering the following

Question 1. Given two families $(A_i^{(n)})_{i \in I}$ and $(B_j^{(n)})_{j \in J}$ of matrices in $\mathbb{M}_n(\mathbb{C})$, what is the joint behavior of $(A_i^{(n)})_{i \in I} \sqcup (U_n B_j^{(n)} U_n^*)_{j \in J}$, where U_n is invariant according to the Haar measure on $\mathcal{U}_n$?

The notion of behavior has to be clarified, and it will be refined in the same time as we refine our estimates of the Weingarten function. For now, we assume that $(A_i^{(n)})_{i\in I}$ and $(B_j^{(n)})_{j\in J}$ have asymptotic moments, namely, for any sequence $i_1, \dots, i_l$,

$$\operatorname{tr} A_{i_1}^{(n)} \cdots A_{i_l}^{(n)}$$

admits a finite limit, and likewise for $(B_j^{(n)})_{j\in J}$ (note that our standing notation is $\operatorname{tr} = n^{-1}\operatorname{Tr}$). In this specific context, the question becomes:

Question 2. Does the enlarged family $(A_i^{(n)})_{i\in I} \sqcup (U_n B_j^{(n)} U_n^*)_{j\in J}$ have asymptotic moments?

Let us note that since the moments are random, the question admits variants, namely, does the enlarged family have asymptotic moments *in expectation, almost surely*? The answer turns out to be *yes*—irrespective of the variant chosen—and the above asymptotics allow us to deduce the joint behavior of random matrices in large dimension. We recall that a family of unital $*$-subalgebras $\mathcal{A}_i, i \in I$ of an NCPS $(\mathcal{A}, \tau)$ is *free* iff for any $l \in \mathbb{N}_*, i_1, \dots, i_l \in I$, $i_1 \neq i_2, \dots, i_{l-1} \neq i_l, \tau(x_1 \cdots x_l) = 0$ as soon as (i) $\tau(x_j) = 0$ and (ii) $x_j \in A_{i_j}$. Asymptotic freeness holds when a family has a limit distribution and the limiting distribution generates free $*$-subalgebras.

Theorem 3.3. *The answer to Question 2 is yes. The limit of the union is determined by the relation of asymptotic freeness, and the convergence is almost sure.*

The proof relies on calculating moments, together with our knowledge of the asymptotics of the Weingarten function. In the next theorem, we observe that different types of "asymptotic behavior," such as the existence of a limiting point spectrum, are also preserved under enlargement of the family. The theorem below is a particular case of a results to be found in [27]:

Theorem 3.4. *Let $\lambda_{i,n}$ be sequences of complex numbers such that $\lim_n \lambda_{i,n} = 0$. Let $\Lambda_{i,n} = \operatorname{diag}(\lambda_{i,1}, \dots, \lambda_{i,n})$ and $A_{j,n}$ be random matrices with the property that (i) $(A_{j,n})_j$ converges in NC distribution as $n \to \infty$ and (ii) $(UA_{j,n}U^*)_j$ has the same distribution as $(A_{j,n})_j$ as a d-tuple of random matrices. Let P be a noncommutative polynomial. Then the eigenvalues of $P(\Lambda_{i,n}, A_{j,n})$ converge almost surely.*

The proof is also based on Weingarten calculus and moment formula. The limiting distribution is of a new type—involving pure point spectrum—and we call it *cyclic monotone convergence*.

3.4. Quantum asymptotic freeness

Finally, let us discuss another seemingly completely unrelated application, to asymptotic representation theory. The idea is to replace classical randomness by quantum randomness. To keep the exposition simple, we stick to the case of the unitary group, although more general results are true for more general Lie groups, see [41]. Call E_{ij} the canonical matrix entries of $\mathbb{M}_n(\mathbb{C})$, and e_{ij} the generators of the enveloping Lie algebra $\mathcal{U}(\mathrm{GL}_n(\mathbb{C}))$

of $\mathrm{GL}_n(\mathbb{C})$, namely, the unital $*$-algebra generated by e_{ij} and the relations $e_{ij}^* = e_{ji}$ and $[e_{ij}, e_{kl}] = \delta_{jk} e_{il} - \delta_{il} e_{kj}$. The map $E_{ij} \to e_{ij}$ can be factored through all Lie algebra representations of $\mathcal{U}_n$, and we are interested in the following variants of its Choi matrix

$$A_n^{(1)} = \sum_{ij} E_{ij} \otimes e_{ij} \otimes 1, \quad A_n^{(2)} = \sum_{ij} E_{ij} \otimes 1 \otimes e_{ij} \in \mathbb{M}_n(\mathbb{C}) \otimes \mathfrak{U}\big(\mathrm{GL}_n(\mathbb{C})\big)^{\otimes 2}.$$

In [39], thanks—among others—to asymptotics of Weingarten functions, we proved the following, extending considerably the results of [11].

Theorem 3.5. *For each n, take λ_n, μ_n two Young diagrams corresponding to a polynomial representations V_{λ_n}, V_{μ_n} of* $\mathrm{GL}_n(\mathbb{C})$. *Assume that both dimensions tend to infinity as $n \to \infty$ and consider the traces on $\chi_{\lambda_n}, \chi_{\mu_n}$ on* $\mathfrak{U}(\mathrm{GL}_n(\mathbb{C}))$. *Assume that A_n converges in noncommutative distribution in Voiculescu's sense both for* $\mathrm{tr} \otimes \chi_{\lambda_n}$ *and* $\mathrm{tr} \otimes \chi_{\mu_n}$. *Then $A_m^{(1)}, A_n^{(2)}$ are asymptotically free with respect to* $\mathrm{tr} \otimes \chi_{\lambda_n} \otimes \chi_{\mu_n}$.

4. MULTIPLICATIVITY AND APPLICATIONS TO MATHEMATICAL PHYSICS

4.1. Higher-order freeness

The asymptotic multiplicativity of the Weingarten function states that

$$\mathrm{Wg}(\sigma_1 \sqcup \sigma_1) = \mathrm{Wg}(\sigma_1)\,\mathrm{Wg}(\sigma_2)\big(1 + O(n^{-2})\big)$$

and it is very far reaching. The fact that the error term $O(n^{-2})$ is summable in n allows in [20] to use a Borel–Cantelli lemma and prove almost sure convergence of moments for random matrices, cf. [70] for the original proof.

A more systematic understanding of the error term is possible and has deep applications in random matrix theory. It requires the notion of classical cumulants that we recall now. Let X be a random variable, the cumulant $C_p(X)$ is defined formally by

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

For instance, the second cumulant $C_2(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$ is the variance of the probability distribution of X. Cumulant $C_p(X)$ is well defined as soon as X has moments up to order p, and it is an n-homogeneous function in X, therefore we can polarize it and define a p-linear symmetric function $(X_1, \ldots, X_p) \to C_p(X_1, \ldots, X_p)$. For any partition π of p elements with blocks $B \in \pi$, we define $C_\pi(X_1, \ldots, X_p) = \prod_{B \in \pi} C(\prod_{i \in B} X_i)$. We are now in the position to write the expectations in term of the cumulants:

$$\mathbb{E}\left(\prod_{i=1}^{p} X_i\right) = \sum_{\pi \in P(p)} C_\pi.$$

The equation can be inverted through the Möbius inversion formula. Asymptotic freeness considers the case where moments have a limit, whereas higher-order asymptotic freeness considers the case where things are known about the fluctuations of the moments: in addition to the existence of $\lim_n \operatorname{tr} A_{i_1}^{(n)} \cdots A_{i_l}^{(n)}$, we assume the existence of

$$\lim_n n^{2k-2} C_k(A_{i_{11}}^{(n)} \cdots A_{i_{l_1 1}}^{(n)}, \dots, A_{i_{1k}}^{(n)} \cdots A_{i_{l_k k}}^{(n)})$$

for any sequence of indices. We call this set of limits the *higher order limit*. In [33], we proved

Theorem 4.1. *The extended family $(A_i^{(n)})_{i\in I} \sqcup (UB_j^{(n)}U^*)_{j\in J}$ admits a higher order limit. In addition, there exists a combinatorial rule to construct the joint asymptotic correlations from the asymptotic correlations of each family.*

This rule extends freeness, is called *higher order freeness*. Subsequent work was done in the case of orthogonal invariance by Mingo and Redelmeier.

4.2. Matrix integrals

Historically, matrix integrals have been studied before higher order freeness. However, from the point of view of formal expansion, higher order freeness supersedes matrix integrals. In [20], we proved the following

Theorem 4.2. *Let A be a noncommutative polynomial in formal variables $(Q_i)_{i\in I}$, formal unitaries U_j, $j \in J$ and their adjoint. Consider in $\mathbb{M}_n(\mathbb{C})$ matrices $(Q_i^{(n)})_{i\in I}$ admitting a joint limiting distribution as $n \to \infty$, and in i.i.d. Haar distributed $(U_j^{(n)})_{j\in J}$ and their adjoint. Evaluating A in these matrices in the obvious sense, we obtain a random matrix A_n and consider the Taylor expansion around zero of the function*

$$z \to n^{-2} \log E\big(\exp(zn^2 A_n)\big) = \sum_{q\geq 1} a_q^{(n)} z^q.$$

Then, for all q, $\lim_n a_q^{(n)}$ exists and depends only on the polynomial and the limiting distribution of $Q_i^{(n)}$.

In [24], we upgraded this result in the case where A_n is selfadjoint, and proved that there exists a *real* neighborhood of zero on which the convergence holds uniformly. The complex convergence remains a difficult problem, as a uniform understanding of the higher genus expansion must be obtained. Novak made a recent breakthrough in this direction, in the case of the HCIZ integral, see [65].

4.3. Random tensors

Let us revisit Question 1, under the assumption that U has *more structure*, i.e., less randomness. Our model is a tensor structure, namely, $U = U_1 \otimes \cdots \otimes U_D$ where $U_i \in \mathbb{M}_n(\mathbb{C})$ are i.i.d. In other words, we are interested in the symmetries under conjugation by elements of the group ${\mathcal{U}_n}^{\otimes D}$. The joint moments of a matrix are a complete invariant of global symmetry under $\mathcal{U}_n$-conjugation in $\mathbb{M}_n(\mathbb{C})$, however, for $U_1 \otimes \cdots \otimes U_D$-invariance in $\mathbb{M}_n(\mathbb{C})^{\otimes D}$, one needs more invariants, generated, for $\sigma_1, \ldots, \sigma_D \in S_k$, by

$$\begin{aligned}\mathrm{Tr}_{\sigma_1,\ldots,\sigma_D}(A) = & \sum_{i_{11},\ldots,i_{Dk},j_{11},\ldots,j_{Dk}} A_{i_{11}\ldots i_{D1},j_{11}\ldots j_{D1}} \cdots A_{i_{1k}\ldots i_{Dk},j_{1k}\ldots j_{Dk}} \\ & \times \delta_{i_{11},j_{1\sigma_1(1)}} \cdots \delta_{i_{1k},j_{1\sigma_1(k)}} \cdots \delta_{i_{D1},j_{D\sigma_D(1)}} \cdots \delta_{i_{Dk},j_{D\sigma_D(k)}}. \end{aligned} \tag{4.1}$$

In the case of higher tensors, thanks to the Weingarten calculus, we unveil many new inequivalent asymptotic regimes for higher order tensors. These questions are addressed in a series of projects with Gurau and Lionni, starting with [26]. We study the asymptotic expansion of the Fourier transform of the tensor valued Haar measure—a tensor extension of the Harish-Chandra integral to tensors, and considerably extend the single tensor case. Just as the HCIZ integral can be seen as a generating function for monotone Hurwitz numbers, which count certain weighted branched coverings of the 2-sphere, the integral studied in [26] leads to a generalization of monotone Hurwitz numbers, which count weighted branched coverings of a collection of 2-spheres that “touch” at one common nonbranch node.

4.4. Quantum information theory

Quantum information theory has been a powerful source of problems in random matrix theory in the last two decades, and their tensor structure has made it necessary to resort to moment techniques. The goal of this section is to elaborate on a few salient cases. One starting point is the paper [53] where the authors compute moments of the output of random quantum channels. We just recall here strictly necessary definitions, and refer to [36] for details. A *quantum channel* Φ is a linear map $\mathbb{M}_n(\mathbb{C}) \to \mathbb{M}_k(\mathbb{C})$ that preserves the nonnormalized trace, and that is completely positive, i.e., $\Phi \otimes \mathrm{Id}_l : \mathbb{M}_n \otimes \mathbb{M}_l(\mathbb{C}) \to \mathbb{M}_k(\mathbb{C}) \otimes \mathbb{M}_l(\mathbb{C})$ is positive for any integer l. It follows from Stinespring theorem that for any quantum channel, there exists an integer p and an isometry $U : \mathbb{C}^n \to \mathbb{C}^k \otimes \mathbb{C}^p$ such that $\Phi(X) = (\mathrm{Id}_k \otimes \mathrm{Tr}_p) U X U^*$.

The set of *density matrices* D_n consists in the selfadjoint matrices whose eigenvalues are nonnegative and whose trace is 1. For $A \in D_n$, we define its *von Neumann entropy* $H(A)$ as $\sum_{i=1}^n -\lambda_i(A) \log \lambda_i(A)$ with the convention that $0 \log 0 = 0$ and the eigenvalues of A are $\lambda_1(A) \geq \cdots \geq \lambda_n(A)$. The *minimum output entropy* of a quantum channel Φ is defined as $H_{\min}(\Phi) = \min_{A \in D_n} H(\Phi(A))$, and a crucial question in QIT was whether one can find Φ_1, Φ_2 such that

$$H_{\min}(\Phi_1 \otimes \Phi_2) < H_{\min}(\Phi_1) + H_{\min}(\Phi_2).$$

For the statement, implications and background, we refer to [36]. An answer to this question was given in [52] and it relies on random methods, which motivates us to consider quantum channels obtained from Haar unitaries. A description of $\Phi(D_n)$ in some appropriate large n limit has been found in [9], and the minimum in the limit of the entropy was found in [10]. In the meantime, the image under tensor product of random channels $\Phi_1 \otimes \Phi_2$ of appropriate matrices (known as Bell states) had to be computed. To achieve this, we had to develop a graphical version of Weingarten calculus in [34].

We consider the case where k is a fixed integer, and $t \in (0, 1)$ is a fixed number. For each n, we consider a random unitary matrix $U \in \mathbb{M}_{nk}(\mathbb{C})$, and a projection q_n of $\mathbb{M}_{nk}(\mathbb{C})$

of rank p_n such that $p_n/(nk) \sim t$ as $n \to \infty$. Our model of a random quantum channel is $\Phi : \mathbb{M}_{p_n}(\mathbb{C}) \to \mathbb{M}_n(\mathbb{C})$ given by $\Phi(X) = \mathrm{tr}_k(UXU^*)$, where $\mathbb{M}_{p_n}(\mathbb{C}) \simeq q_n\mathbb{M}_{nk}(\mathbb{C})q_n$. By Bell we denote the Bell state on $\mathbb{M}_{p_n}(\mathbb{C})^{\otimes 2}$. In [34], we proved

Theorem 4.3. *Almost surely, as $n \to \infty$, the random matrix $\Phi \otimes \overline{\Phi}(\mathrm{Bell}) \in \mathbb{M}_{n^2}(\mathbb{C})$ has nonzero eigenvalues converging towards*

$$\gamma^{(t)} = \Bigg(t + \frac{1-t}{k^2}, \underbrace{\frac{1-t}{k^2}, \dots, \frac{1-t}{k^2}}_{k^2-1 \text{ times}}\Bigg).$$

This result plays an important result in the understanding of phenomena underlying the subadditivity of the minimum output entropy, and relies heavily on Weingarten calculus, and in particular a graphical interpretation thereof. Much more general results in related areas of quantum information theory have been obtained in [22,28,35,37,38,43].

5. UNIFORM ESTIMATES AND APPLICATIONS TO ANALYSIS

5.1. A motivating question

The previous sections show that when the degree of a polynomial is fixed, very precise asymptotics can be obtained in the limit of large dimension. For the purpose of analysis, an important question is whether such estimates hold uniformly. About 20 years ago, Gilles Pisier asked me the following question: given k i.i.d. Haar unitaries $U_1^{(n)}, \dots, U_k^{(n)} \in \mathcal{U}_n$, what is the large dimension behavior of the real random variable

$$t_n = ||U_1^{(n)} + \dots + U_k^{(n)}||_\infty,$$

where $||\cdot||_\infty$ stands for the operator norm? It follows from asymptotic freeness results that almost surely $\liminf t_n \geq 2\sqrt{k-1}$ as soon as $k \geq 2$. Setting $X_n = U_1^{(n)} + \dots + U_k^{(n)}$, it would be in principle enough to estimate

$$\mathbb{E}\big(\mathrm{Tr}\big((X_n X_n^*)^{l(n)}\big)\big)$$

for $l(n) \gg \log n$. However, there are two major hurdles: (i) uniform estimates of Weingarten calculus would be needed; (ii) unlike in the multimatrix model case, the combinatorics grow exponentially and a direct moment approach is not possible. Both hurdles require developing specific tools, which we describe in the sequel.

One notion on which we rely heavily is that of *strong convergence*. Given a multiatrix model that admits a joint limiting distribution in Voiculescu's sense, we say that it *converges strongly* iff the operator norm of any polynomial P, evaluated in the matrices of the model—thus yielding the random matrix P_n—satisfies

$$\lim_n ||P_n|| = \lim_\ell (\lim_n n^{-1}\,\mathrm{Tr}\big((P_n P_n^*)^\ell\big)^{(2\ell)^{-1}}.$$

In other words, the operator norm of any matrix model obtained from a noncommutative polynomial converges to the operator norm of the limiting object. Strong convergence was established in [50] in the case of Gaussian random matrices. Subsequently, the author and

Male solved the counterpart for Haar unitary matrices in [29], with no explicit speed of convergence. This result was refined further by Parraud [67] with explicit speeds of convergence, relying on ideas of [25]. The strongest result concerning strong convergence of random unitaries can be found in [14]:

Theorem 5.1. $(\overline{U}_i^{\otimes q_-} \otimes U_i^{\otimes q_+})_{i=1,\dots,d}$ *are strongly asymptotically free as* $n \to \infty$ *on the orthogonal of fixed point spaces.*

This means that strong asymptotic freeness does not hold at the sole level of the fundamental representation of $\mathcal{U}_n$, but with respect to any sequence of representation associated to a nontrivial (λ, μ). In other words, the only obstructions to strong freeness are the dimension one irreducible representations of $\mathcal{U}_n$. We need a *linearization step*, popularized by [50] to evaluate the norm of $\sum_{i=-d}^{d} a_i \otimes X_i^{(n)}$, where $X_{-i}^{(n)} = X_i^{(n)*}$ and $X_0^{(n)} = Id$. Although this first simplification step was sufficient to obtain strong convergence for i.i.d. GUE—i.e., matrices with high symmetries—thanks to analytic techniques, this turns out to be insufficient when one has to resort to moment methods. In [13], we initiated techniques based on a *operator version* of *nonbacktracking theory*, which we generalized in [14]. We outline one key feature here.

We consider $(b_1, \dots, b_l)$ elements in $\mathcal{B}(\mathcal{H})$ where $\mathcal{H}$ is a Hilbert space. We assume that the index set is endowed with an involution $i \mapsto i^*$ (and $i^{**} = i$ for all i). The *nonbacktracking operator* associated to the ℓ-tuple of matrices $(b_1, \dots, b_l)$ is the operator on $\mathcal{B}(\mathcal{H} \otimes \mathbb{C}^l)$ defined by

$$B = \sum_{j \neq i^*} b_j \otimes E_{ij}. \tag{5.1}$$

The following theorem allows leveraging moments techniques on linearization of noncommutative polynomials through the study of B:

Theorem 5.2. *Let* $\lambda \in \mathbb{C}$ *satisfy* $\lambda^2 \notin \cup_{i \in \{1,\dots,l\}} \operatorname{spec}(b_i b_{i^*})$. *Define the operator* A_λ *on* $\mathcal{H}$ *through*

$$A_\lambda = b_0(\lambda) + \sum_{i=1}^{\ell} b_i(\lambda), \quad b_i(\lambda) = \lambda b_i (\lambda^2 - b_{i^*} b_i)^{-1},$$

and

$$b_0(\lambda) = -1 - \sum_{i=1}^{\ell} b_i (\lambda^2 - b_{i^*} b_i)^{-1} b_{i^*}.$$

Then $\lambda \in \sigma(B)$ *if and only if* $0 \in \sigma(A_\lambda)$.

5.2. Centering and uniform Weingarten estimates

In order to use Theorem 5.2, one has to understand the spectral radius of the operator B and therefore, evaluate $\tau(B^T B^{*T})$ with T growing with the matrix dimension, and this can be done through moment methods as soon as we have uniform estimates on Weingarten functions. The first uniform estimate was obtained in [23] and had powerful applications to the study of area laws in mathematical physics, however, it was not sufficient for norm estimates, and it was superseded by [31]:

Theorem 5.3. *For any $\sigma \in S_k$ and $n > \sqrt{6}k^{7/4}$,*

$$\frac{1}{1-\frac{k-1}{n^2}} \leq \frac{n^{k+|\sigma|}\,\mathrm{Wg}(n,\sigma)}{\mathrm{Moeb}(\sigma)} \leq \frac{1}{1-\frac{6k^{7/2}}{n^2}}.$$

In addition, the left-hand side inequality is valid for any $n \geq k$.

This result already enables us to prove Theorem 5.1 in the case where $q_- \neq q_+$ because there are no fixed points there. Let us now outline how to tackle the case $q_- = q_+$, which is interesting because it has fixed points. To handle fixed points, we need to introduce the *centering* of a random variable X, namely $[X] = X - E(X)$. For a symbol $\varepsilon \in \{\cdot, -\}$ and $z \in \mathbb{C}$, we take the notation that $z^\varepsilon = z$ if $\varepsilon = \cdot$ and $z^\varepsilon = \bar{z}$ if $\varepsilon = -$. We want to compute, for $U = (U_{ij})$ Haar distributed on $\mathcal{U}_n$, expressions of the form $\mathbb{E}\prod_{t=1}^{T}[\prod_{l=1}^{k_t} U_{x_{tl}y_{tl}}^{\varepsilon_{tl}}]$ in a meaningful way. We can write a Weingarten formula:

$$\mathbb{E}\prod_{t=1}^{T}\left[\prod_{l=1}^{k_t} U_{x_{tl}y_{tl}}^{\varepsilon_{tl}}\right] = \sum_{\sigma,\tau\in P_2(k_1+\cdots+k_T)} \delta_{\sigma,x}\delta_{\tau,y}\,\mathrm{Wg}(\sigma,\tau;k_1,\ldots,k_T),$$

where the function Wg depends on the pairings and the partition. We say that a block of the partition $\{\{1,\ldots,k_1\},\ldots,\{k_1+\cdots+k_{T-1}+1,\ldots,k_1+\cdots+k_T\}\}$ is *lonesome* with respect to the pairing (σ,τ) iff the group generated by σ,τ stabilizes it. In **[14]**, we prove

Theorem 5.4. Wg *decays as n^{-k} where $k = (k_1+\cdots+k_T)/2 + d(\sigma,\tau) + 2\#$lonesome blocks, and this estimate is uniform on $k \sim \mathrm{Poly}(n)$.*

This theorem, together a comparison with Gaussian vectors, allows proving Theorem 5.1.

6. PERSPECTIVES

Understanding better how to integrate over compact groups is a fascinating problem which is connected to many questions in various branches of mathematics and other scientific fields. We conclude this manuscript by a brief and completely subjective list of perspectives:

(1) *Uniform measures on (quantum) symmetric spaces*
Viewing a group as a compact manifold, can one extend the Weingarten calculus to other surfaces? Some substantial work has been done algebraically in this direction by Matsumoto **[60]** in the case of symmetric spaces, see as well **[42]** for the asymptotic version. It would be interesting to study extensions of Matsumoto's results for compact quantum symmetric spaces.

(2) *Surfaces and geometric group theory*
An important observation by Magee and Puder is that if G is a compact subgroup of $\mathcal{U}_n$, the Haar measure on G^k yields a random representation of the free group F_k on $\mathcal{U}_n$ whose law is invariant under *outer automorphisms* of F_k.

This motivated them to compute, in [59], the expectation of the trace of nontrivial words in $(U_1, \dots, U_k) \in \mathcal{U}_n{}^k$. In addition to refining known asymptotics, they used the properties of the Weingarten function to solve nontrivial problems about the orbits of F_k under the action by its outer conjugacy group. In a different vein, Magee has very recent achieved a breakthrough by obtaining the first steps of Weingarten calculus for representations of some one relator groups [57, 58].

(3) *Other applications to representation theory*
The problem of calculating Weingarten functions on $\mathrm{SU}(n)$ efficiently is hard, and even more so when the degree is high in comparison to n. A striking example is $\int_{U \in \mathrm{SU}(n)} \prod_{i,j=1}^n u_{ij}$. It was established in [55] that proving that this integral is nonzero is equivalent to the Alon–Tarski conjecture.
More generally, this raises the question of computing efficiently integrals of high degree on classical groups (typically, of degree $\geq n$ or $\geq n^2$). Weingarten calculus as developed in this manuscript is not well adapted to this task. Some results in this direction have been obtained by Novak [65] the author, and Cioppa.

(4) *More tensors and norm estimates*
In [13, 14], we obtained strong convergence for an arbitrary finite number of tensors of random unitaries or random permutations. It turns out that the result can be relaxed a bit to allow the number of legs to vary slowly to infinity as the dimension of the group goes to infinity. This points to a double limit problem, and we wonder to which extent the number of legs of tensors and the size of the matrix can be independent. In the extreme case, could strong freeness hold for a given finite group but a number of tensors tending to infinity? Many variants of this problem exist, e.g., taking i.i.d. copies of unitaries instead of the same.
Likewise, an important question is the behavior of $U_i \otimes U_j$ for arbitrary indices—not only $i = j$ as in [13, 14]. As observed by Hayes in [54], this is a possible approach towards the Peterson–Thom conjecture in operator algebras, and it seems plausible that Weingarten calculus could help to solve this problem.

(5) *Maximizing functionals over groups*
Given a polynomial function $f : G \to \mathbb{C}$, finding $m = \max_{U \in G} |f(U)|$ could provide approaches to various conjectures in analysis or algebra. In general, finding the argmax is a problem intricately linked to the conjectures, and Haar integration could yield nonconstructive approaches. Indeed,

$$l^{-1} \log \int \big(f(U)\overline{f(U)}\big)^l \, d\mu_G(U) \sim 2 \log m,$$

and the left-hand side could in principle be approached with Weingarten calculus. Let us mention the example of the Hadamard conjecture. It states that for any $4/n$, there exists an orthogonal matrix in $\mathbb{M}_n(\mathbb{R})$ whose entries are ± 1.

An approach to this problem would be to show that the minimum of the polynomial function $f(U) = \sum_{ij} u_{ij}^4$ on $\mathcal{O}_n$ is 1. We refer to [7] for attempts with Weingarten calculus. We also believe that some important problems in operator algebra could be approached that way (e.g., the problem of the nonexistence of hyperlinear group).

ACKNOWLEDGMENTS

It has always been extremely stimulating to work with people from very diverse backgrounds: coauthors, graduate students, as well as postdoctoral fellows. I want to thank them all for our collaborations. I would like to thank Charles Bordenave, Mike Brannan, Luca Lionni, Sho Matsumoto, Akihiro Miyagawa, Ion Nechita, and Jonathan Novak for reading carefully my manuscript and for many suggestions of improvements.

FUNDING

This work was partially supported JSPS Kakenhi 17H04823, 20K20882, 21H00987. Most of the original results presented in this note have been supported by JSPS Kakenhi, NSERC grants and ANR grants, while the author was working at either of the following places: CNRS (Lyon 1), the University of Ottawa, or Kyoto University.

REFERENCES

[1] T. Banica, Le groupe quantique compact libre $U(n)$. *Comm. Math. Phys.* **190** (1997), no. 1, 143–172.

[2] T. Banica, S. T. Belinschi, M. Capitaine, and B. Collins, Free Bessel laws. *Canad. J. Math.* **63** (2011), no. 1, 3–37.

[3] T. Banica, J. Bichon, and B. Collins, Quantum permutation groups: a survey. In *Noncommutative harmonic analysis with applications to probability*, pp. 13–34, Banach Center Publ. 78, Polish Acad. Sci. Inst. Math, Warsaw, 2007.

[4] T. Banica and B. Collins, Integration over compact quantum groups. *Publ. Res. Inst. Math. Sci.* **43** (2007), no. 2, 277–302.

[5] T. Banica and B. Collins, Integration over quantum permutation groups. *J. Funct. Anal.* **242** (2007), no. 2, 641–657.

[6] T. Banica and B. Collins, Integration over the Pauli quantum group. *J. Geom. Phys.* **58** (2008), no. 8, 942–961.

[7] T. Banica, B. Collins, and J.-M. Schlenker, On polynomial integrals over the orthogonal group. *J. Combin. Theory Ser. A* **118** (2011), no. 3, 778–795.

[8] T. Banica, S. Curran, and R. Speicher, Stochastic aspects of easy quantum groups. *Probab. Theory Related Fields* **149** (2011), no. 3–4, 435–462.

[9] S. Belinschi, B. Collins, and I. Nechita, Eigenvectors and eigenvalues in a random subspace of a tensor product. *Invent. Math.* **190** (2012), no. 3, 647–697.

[10] S. T. Belinschi, B. Collins, and I. Nechita, Almost one bit violation for the additivity of the minimum output entropy. *Comm. Math. Phys.* **341** (2016), no. 3, 885–909.

[11] P. Biane, Representations of unitary groups and free convolution. *Publ. Res. Inst. Math. Sci.* **31** (1995), no. 1, 63–79.

[12] P. Biane, Representations of symmetric groups and free probability. *Adv. Math.* **138** (1998), no. 1, 126–181.

[13] C. Bordenave and B. Collins, Eigenvalues of random lifts and polynomials of random permutation matrices. *Ann. of Math. (2)* **190** (2019), no. 3, 811–875.

[14] C. Bordenave and B. Collins, Strong asymptotic freeness for independent uniform variables on compact groups associated to non-trivial representations. 2020, arXiv:2012.08759.

[15] N. Bourbaki, *Integration. II. Chapters 7–9*. Elem. Math. (Berlin), Springer, Berlin, 2004. Translated from the 1963 and 1969 French originals by S. K. Berberian.

[16] M. Brannan and B. Collins, Dual bases in Temperley–Lieb algebras, quantum groups, and a question of Jones. *Quantum Topol.* **9** (2018), no. 4, 715–748.

[17] E. Brézin and D. J. Gross, The external field problem in the large N limit of QCD. *Phys. Lett. B* **97** (1980), no. 1, 120–124.

[18] P. W. Brouwer and C. W. J. Beenakker, Diagrammatic method of integration over the unitary group, with applications to quantum transport in mesoscopic systems. *J. Math. Phys.* **37** (1996), no. 10, 4904–4934.

[19] T. Ceccherini-Silberstein, F. Scarabotti, and F. Tolli, *Representation theory of the symmetric groups*. Cambridge Stud. Adv. Math. 121, Cambridge University Press, Cambridge, 2010. The Okounkov–Vershik approach, character formulas, and partition algebras.

[20] B. Collins, Moments and cumulants of polynomial random variables on unitary groups, the Itzykson–Zuber integral, and free probability. *Int. Math. Res. Not.* **17** (2003), 953–982.

[21] B. Collins, A. Dahlqvist, and T. Kemp, The spectral edge of unitary Brownian motion. *Probab. Theory Related Fields* **170** (2018), no. 1–2, 49–93.

[22] B. Collins, M. Fukuda, and I. Nechita, On the convergence of output sets of quantum channels. *J. Operator Theory* **73** (2015), no. 2, 333–360.

[23] B. Collins, C. E. González-Guillén, and D. Pérez-García, Matrix product states, random matrix theory and the principle of maximum entropy. *Comm. Math. Phys.* **320** (2013), no. 3, 663–677.

[24] B. Collins, A. Guionnet, and E. Maurel-Segala, Asymptotics of unitary and orthogonal matrix integrals. *Adv. Math.* **222** (2009), no. 1, 172–215.

[25] B. Collins, A. Guionnet, and F. Parraud, On the operator norm of non-commutative polynomials in deterministic matrices and iid GUE matrices. 2019, arXiv:1912.04588. Accepted in *Camb. J. Math.*

[26] B. Collins, R. Gurau, and L. Lionni, The tensor Harish-Chandra–Itzykson–Zuber integral I: Weingarten calculus and a generalization of monotone Hurwitz numbers. 2020, arXiv:2010.13661.

[27] B. Collins, T. Hasebe, and N. Sakuma, Free probability for purely discrete eigenvalues of random matrices. *J. Math. Soc. Japan* **70** (2018), no. 3, 1111–1150.

[28] B. Collins, P. Hayden, and I. Nechita, Random and free positive maps with applications to entanglement detection. *Int. Math. Res. Not. IMRN* **3** (2017), 869–894.

[29] B. Collins and C. Male, The strong asymptotic freeness of Haar and deterministic matrices. *Ann. Sci. Éc. Norm. Supér. (4)* **47** (2014), no. 1, 147–163.

[30] B. Collins and S. Matsumoto, On some properties of orthogonal Weingarten functions. *J. Math. Phys.* **50** (2009), no. 11, 113516, 14 pp.

[31] B. Collins and S. Matsumoto, Weingarten calculus via orthogonality relations: new applications. *ALEA Lat. Am. J. Probab. Math. Stat.* **14** (2017), no. 1, 631–656.

[32] B. Collins, S. Matsumoto, and J. Novak, The Weingarten calculus. 2021, arXiv:2109.14890.

[33] B. Collins, J. A. Mingo, P. Śniady, and R. Speicher, Second order freeness and fluctuations of random matrices. III. Higher order freeness and free cumulants. *Doc. Math.* **12** (2007), 1–70.

[34] B. Collins and I. Nechita, Random quantum channels I: Graphical calculus and the Bell state phenomenon. *Comm. Math. Phys.* **297** (2010), no. 2, 345–370.

[35] B. Collins and I. Nechita, Gaussianization and eigenvalue statistics for random quantum channels (III). *Ann. Appl. Probab.* **21** (2011), no. 3, 1136–1179.

[36] B. Collins and I. Nechita, Random matrix techniques in quantum information theory. *J. Math. Phys.* **57** (2016), no. 1, 015215, 34 pp.

[37] B. Collins, I. Nechita, and D. Ye, The absolute positive partial transpose property for random induced states. *Random Matrices Theory Appl.* **1** (2012), no. 3, 1250002, 22 pp.

[38] B. Collins, I. Nechita, and K. Życzkowski, Area law for random graph states. *J. Phys. A* **46** (2013), no. 30, 305302, 18 pp.

[39] B. Collins, J. Novak, and P. Śniady, Semiclassical asymptotics of $\mathrm{GL}_N(C)$ tensor products and quantum random matrices. *Selecta Math. (N.S.)* **24** (2018), no. 3, 2571–2623.

[40] B. Collins and P. Śniady, Integration with respect to the Haar measure on unitary, orthogonal and symplectic group. *Comm. Math. Phys.* **264** (2006), no. 3, 773–795.

[41] B. Collins and P. Śniady, Representations of Lie groups and random matrices. *Trans. Amer. Math. Soc.* **361** (2009), no. 6, 3269–3287.

[42] B. Collins and M. Stolz, Borel theorems for random matrices from the classical compact symmetric spaces. *Ann. Probab.* **36** (2008), no. 3, 876–895.

[43] B. Collins, Z. Yin, and P. Zhong, The PPT square conjecture holds generically for some classes of independent states. *J. Phys. A* **51** (2018), no. 42, 425301, 19 pp.

[44] B. de Wit and G. 't Hooft, Nonconvergence of the $1/N$ expansion for $\mathrm{SU}(N)$ gauge fields on a lattice. *Phys. Lett. B* **69** (1977), no. 1, 61–64.

[45] P. Diaconis and M. Shahshahani, On the eigenvalues of random matrices. J. Appl. Probab. **31A** (1994), 49–62.

[46] R. Goodman and N. R. Wallach, *Symmetry, representations, and invariants*. Grad. Texts in Math. 255, Springer, Dordrecht, 2009.

[47] I. P. Goulden, M. Guay-Paquet, and J. Novak, Monotone Hurwitz numbers in genus zero. *Canad. J. Math.* **65** (2013), no. 5, 1020–1042.

[48] I. P. Goulden, M. Guay-Paquet, and J. Novak, Polynomiality of monotone Hurwitz numbers in higher genera. *Adv. Math.* **238** (2013), 1–23.

[49] I. P. Goulden, M. Guay-Paquet, and J. Novak, Monotone Hurwitz numbers and the HCIZ integral. *Ann. Math. Blaise Pascal* **21** (2014), no. 1, 71–89.

[50] U. Haagerup and S. Thorbjørnsen, A new application of random matrices: $\mathrm{Ext}(C^*_{\mathrm{red}}(F_2))$ is not a group. *Ann. of Math. (2)* **162** (2005), no. 2, 711–775.

[51] Harish-Chandra, Differential operators on a semisimple Lie algebra. *Amer. J. Math.* **79** (1957), 87–120.

[52] M. B. Hastings, Superadditivity of communication capacity using entangled inputs. *Nat. Phys.* **5** (2009), no. 4, 255–257.

[53] P. Hayden and A. Winter, Counterexamples to the maximal p-norm multiplicity conjecture for all $p > 1$. *Comm. Math. Phys.* **284** (2008), no. 1, 263–280.

[54] B. Hayes, A random matrix approach to the Peterson–Thom conjecture. 2020, arXiv:2008.12287.

[55] S. Kumar and J. M. Landsberg, Connections between conjectures of Alon–Tarsi, Hadamard–Howe, and integrals over the special unitary group. *Discrete Math.* **338** (2015), no. 7, 1232–1238.

[56] T. Lévy, Schur–Weyl duality and the heat kernel measure on the unitary group. *Adv. Math.* **218** (2008), no. 2, 537–575.

[57] M. Magee, Random unitary representations of surface groups I: asymptotic expansions. 2021, arXiv:2101.00252.

[58] M. Magee, Random unitary representations of surface groups II: the large n limit. 2021, arXiv:2101.03224.

[59] M. Magee and D. Puder, Matrix group integrals, surfaces, and mapping class groups I: $U(n)$. *Invent. Math.* **218** (2019), no. 2, 341–411.

[60] S. Matsumoto, Weingarten calculus for matrix ensembles associated with compact symmetric spaces. *Random Matrices Theory Appl.* **2** (2013), no. 2, 1350001, 26 pp.

[61] S. Matsumoto and J. Novak, Jucys–Murphy elements and unitary matrix integrals. *Int. Math. Res. Not. IMRN* **2** (2013), 362–397.

[62] C. McSwiggen, A new proof of Harish-Chandra's integral formula. *Comm. Math. Phys.* **365** (2019), no. 1, 239–253.

[63] J. A. Mingo and R. Speicher, *Free probability and random matrices*. Fields Inst. Monogr. 35, Springer, New York; Fields Institute for Research in Mathematical Sciences, Toronto, ON, 2017.

[64] A. Nica and R. Speicher, *Lectures on the combinatorics of free probability*. London Math. Soc. Lecture Note Ser. 335, Cambridge University Press, Cambridge, 2006.

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

[66] K. H. O'Brien and J. B. Zuber, A note on $U(N)$ integrals in the large N limit. *Phys. Lett. B* **144** (1984), no. 5–6, 407–408.

[67] F. Parraud, On the operator norm of non-commutative polynomials in deterministic matrices and iid Haar unitary matrices. 2020, arXiv:2005.13834. Accepted in PTRF.

[68] E. M. Rains, Combinatorial properties of Brownian motion on the compact classical groups. *J. Theoret. Probab.* **10** (1997), no. 3, 659–679.

[69] D. Stanford, Z. Yang, and S. Yao, Subleading Weingartens. 2021, arXiv:2107.10252.

[70] D. Voiculescu, A strengthened asymptotic freeness result for random matrices with applications to free entropy. *Int. Math. Res. Not.* **1** (1998), 41–63.

[71] D. V. Voiculescu, K. J. Dykema, and A. Nica, *A noncommutative probability approach to free products with applications to random matrices, operator algebras and harmonic analysis on free groups*. Free random variables, CRM Monogr. Ser. 1, American Mathematical Society, Providence, RI, 1992.

[72] S. Wang, Free products of compact quantum groups. *Comm. Math. Phys.* **167** (1995), no. 3, 671–692.

[73] D. Weingarten, Asymptotic behavior of group integrals in the limit of infinite rank. *J. Math. Phys.* **19** (1978), no. 5, 999–1001.

[74] S. L. Woronowicz, Compact matrix pseudogroups. *Comm. Math. Phys.* **111** (1987), no. 4, 613–665.

[75] S. L. Woronowicz, Tannaka–Kreĭn duality for compact matrix pseudogroups. Twisted SU(N) groups. *Invent. Math.* **93** (1988), no. 1, 35–76.

BENOÎT COLLINS

Mathematics Department, Kyoto University, Kyoto, Japan, collins@math.kyoto-u.ac.jp