SEMiclassical Asymptotics Of GLN(ℂ) Tensor Products Via Quantum Random Matrices

BENOİT COLLINS, JONATHAN NOVAK, AND PIOTR ŚNIADY

To Philippe Biane, for his 55th birthday.

Contents

1. Introduction 2
  1.1. Asymptotic representation theory 2
  1.2. Connection with RMT 3
  1.3. Biane-Perelomov-Popov quantization 4
  1.4. The semiclassical limit 6
  1.5. Mean values and asymptotic freeness 6
  1.6. Proof strategy 8
  1.7. Higher correlators and higher order freeness 9
2. Computations at the Planck Scale 10
  2.1. Unitary invariance 10
  2.2. The Weingarten function 11
  2.3. Casimirs and Biasimirs 12
  2.4. Classical/Quantum decomposition 15
3. Asymptotic Computations 17
  3.1. The Weingarten function 18
  3.2. Quantum asymptotics 18
  3.3. Classical asymptotics 21
  3.4. Semiclassical asymptotics and freeness 21
4. Covariance of Traces of BPP Matrices 22
  4.1. Covariance setup 22
  4.2. More computations at the Planck scale 23
  4.3. More asymptotics 29

References 31

Appendix A. BPP Matrices and Geometric Quantization 32
  A.1. Toy example 32

Date: October 10, 2016.

2010 Mathematics Subject Classification. 22E46 (Primary) 60B20, 46L54, 34L20 (Secondary).

Key words and phrases. Random matrix theory, asymptotic representation theory, free probability.
1. Introduction

1.1. Asymptotic representation theory. The rational representations of the complex general linear group $GL_N(\mathbb{C})$ were classified by Schur more than a century ago, see e.g. [Wey39]. This classification may be stated, albeit in a slightly nonstandard way, as follows: the irreducible rational representations of $GL_N(\mathbb{C})$ are parameterized, up to isomorphism, by configurations of $N$ hard particles on the one-dimensional lattice $\mathbb{Z}[\hbar N]$, where $\hbar N > 0$ is an arbitrary lattice constant specifying the spacing between adjacent sites. Once Schur’s classification is known, one may ask which particle configurations occur, and with what multiplicity, as the signature of an irreducible component of a representation obtained from irreducibles by means of standard operations. In this paper, we focus on tensor products.

Given a sequence

\[ h_1, h_2, h_3, \ldots \]

of lattice constants and two triangular arrays

\[
\begin{align*}
& a_1^{(1)} \quad a_2^{(1)} \\
& a_1^{(2)} \quad a_2^{(2)} \quad a_3^{(2)} \\
& a_1^{(3)} \quad a_2^{(3)} \quad a_3^{(3)} \\
& \vdots \quad \vdots \quad \vdots \\
& b_1^{(1)} \quad b_2^{(1)} \\
& b_1^{(2)} \quad b_2^{(2)} \quad b_3^{(2)} \\
& b_1^{(3)} \quad b_2^{(3)} \quad b_3^{(3)} \\
& \vdots \quad \vdots \quad \vdots \\
\end{align*}
\]

such that, for each $N \in \mathbb{N}^*$,

\[ a_1^{(N)} > \cdots > a_N^{(N)} \quad \text{and} \quad b_1^{(N)} > \cdots > b_N^{(N)} \]

specify a pair of particle configurations on $\mathbb{Z}[\hbar N]$, let $V_N$ and $W_N$ be the corresponding irreducible representations of $GL_N(\mathbb{C})$. The problem is to determine the multiplicity $\text{mult}_N(c_1, \ldots, c_N)$ of the irreducible representation $X^{(c_1, \ldots, c_N)}$ indexed by the configuration $\{c_1 > \cdots > c_N\} \subset \mathbb{Z}[\hbar N]$ in the tensor product $V_N \otimes W_N$.

In principle, this question is answered by the famous Littlewood-Richardson rule (see e.g. [Ful97]), which gives a combinatorial algorithm for computing $\text{mult}_N(c_1, \ldots, c_N)$. Unfortunately, the complexity of this algorithm is
such that, for generic input data (2), it quickly becomes unusable as \( N \) increases. In lieu of satisfactory exact formulas, one may pursue approximate, statistical answers to questions of this sort. Indeed, the data (2) determines a natural sequence of probability measures \( P_N \) on configurations of \( N \) hard particles on \( \mathbb{Z}[\hbar] \):

\[
P_N(c_1 > \cdots > c_N) = \frac{\text{mult}_N(c_1, \ldots, c_N) \dim X^{(c_1, \ldots, c_N)}}{\dim V_N \dim W_N}, \quad N \in \mathbb{N}^*.
\]

One may seek a probabilistic understanding of the random point process (4) on \( \mathbb{Z}[\hbar] \) whose law is \( P_N \). This line of investigation was opened twenty years ago by Biane [Bia95], who was the first to realize its intimate connection with random matrix theory (RMT).

1.2. Connection with RMT. Biane’s philosophical point of departure was the principle that the discrete point process (4) quantizes a certain classical continuous random point process, namely the ensemble

\[
z_1^{(N)} \geq \cdots \geq z_N^{(N)}, \quad N \in \mathbb{N}^*,
\]

of eigenvalues of the random Hermitian matrix \( Z_N = X_N + Y_N \) whose summands \( X_N, Y_N \) are \( N \times N \) random Hermitian matrices with independent, uniformly random eigenvectors and deterministic eigenvalues given by the configurations (3). That (4) is indeed a quantization of (5) in some meaningful sense may be understood in the general context of the Kirillov-Kostant orbit method, see e.g. [KT01] for a discussion along these lines.

The process (5) admits a very natural family of global observables, namely the real random variables

\[
P_d^{(N)} := \frac{1}{N} p_d(z_1^{(N)}, \ldots, z_N^{(N)}), \quad d \in \mathbb{N}^*,
\]

where \( p_d(x_1, \ldots, x_N) = x_1^d + \cdots + x_N^d \) is the Newton power sum symmetric function of degree \( d \). For any \( r, d_1, \ldots, d_r \in \mathbb{N}^* \), one has a corresponding obvious equality of \( r \)-point correlation functions, namely

\[
\langle P_{d_1}^{(N)} \cdots P_{d_r}^{(N)} \rangle = \mathbb{E} \left[ \text{tr}(Z_{N}^{d_1}) \cdots \text{tr}(Z_{N}^{d_r}) \right];
\]

here \( \text{tr} = N^{-1} \text{Tr} \) is the normalized matrix trace, \( \langle \cdot \rangle \) denotes expectation with respect to the law of (5), and \( \mathbb{E} \) denotes expectation with respect to the distribution of \( Z_N \) in the space of \( N \times N \) Hermitian matrices. This formula allows one to analyze the point process (5) by working with mixed moments of the matrix elements of \( Z_N \), and these can in turn be attacked by leveraging the stochastic independence of the matrix elements of \( X_N \) and
Y_N; this is an instance of the moment method, a ubiquitous and powerful technique in random matrix theory, see e.g. [AGZ].

1.3. Biane-Perelomov-Popov quantization. Biane followed the quantum vs. classical analogy between the particle systems (4) and (5) to its natural conclusion: he discovered that a formula identical in form to (7) holds for the discrete particle ensemble (4), provided one broadens the scope of RMT to include quantum random matrices. Thus the problem of decomposing GL_N(C) tensor products becomes a part of RMT once quantum random matrices are brought into the game.

Let us describe Biane’s fundamental insight in more detail. The first step is to understand how to quantize the classical random Hermitian matrices X_N, Y_N. Up to minor modifications, the required quantization was constructed in the 1960s by Perelomov and Popov [PP67], see also Želobenko [Žel73]. It is as follows. For each \(N \in \mathbb{N}^*\), introduce two \(N \times N\) matrices defined by

\[
A_N := \begin{bmatrix}
\ldots & \ldots \\
\rho_N(h_N e_{ij}) \otimes I_{W_N} & \ldots \\
\vdots & \vdots 
\end{bmatrix},
\]

and

\[
B_N := \begin{bmatrix}
\ldots & \ldots \\
I_{V_N} \otimes \sigma_N(h_N e_{ij}) & \ldots \\
\vdots & \vdots 
\end{bmatrix},
\]

where \(\{e_{ij}\}\) are the standard generators of the universal enveloping algebra \(\mathcal{U}(\text{gl}_N(\mathbb{C}))\) and \(\rho_N, \sigma_N\) are the actions of \(\mathcal{U}(\text{gl}_N(\mathbb{C}))\) on \(V_N, W_N\) induced by the respective linear actions of GL_N(\mathbb{C}) on these vector spaces. The matrices \(A_N, B_N\) so defined are quantum random matrices in the sense that their entries are quantum random variables living in the noncommutative probability space \((\mathcal{A}_N, \mathbb{E})\), where \(\mathcal{A}_N\) is the algebra

\[
\mathcal{A}_N := \text{End} V_N \otimes \text{End} W_N
\]

and \(\mathbb{E}: \mathcal{A}_N \to \mathbb{C}\) is the quantum expectation functional defined by

\[
\mathbb{E} := \text{tr}_{V_N} \otimes \text{tr}_{W_N},
\]

with

\[
\text{tr}_{V_N} := \frac{1}{\dim V_N} \text{Tr}_{V_N}
\]

the normalized trace on \(\text{End} V_N\) and \(\text{tr}_{W_N}\) the analogous normalized trace on \(\text{End} W_N\).
We shall refer to the quantum random matrices $A_N$, $B_N$ as Biane-Perelomov-Popov matrices, or BPP matrices for short. In Appendix A, we present a self-contained argument, in the spirit of geometric quantization, which explains why the BPP matrices $A_N$, $B_N$ are the canonical quantization of the classical random Hermitian matrices $X_N$, $Y_N$. A more pedestrian justification is simply that “it works” — BPP matrices provide us with a quantum analogue of the formula (7), which can be harnessed to analyze the point process.

As shown by Perelomov and Popov, traces of powers of $A_N$ and $B_N$ are scalar operators in $A_N —$ this is the quantum analogue of the fact that the classical random matrices $X_N$, $Y_N$ have deterministic spectra. In fact, Perelomov and Popov showed that, for each $d \in \mathbb{N}^*$, one has

\[
\text{Tr}(A_N^d) = \wp_d(a_1^{(N)}, \ldots, a_N^{(N)}) I_{V_N} \otimes I_{W_N}
\]

\[
\text{Tr}(B_N^d) = \wp_d(b_1^{(N)}, \ldots, b_N^{(N)}) I_{V_N} \otimes I_{W_N},
\]

where

\[
(8) \quad \wp_d(x_1, \ldots, x_N) = \sum_{i=1}^N \prod_{j \neq i} \left(1 - \frac{\hbar N}{x_i - x_j}\right)x_i^d
\]

is a quantum deformation of the power sum $p_d$, to which it visibly degenerates in the limit $\hbar N \to 0$.

Biane realized that the deformed power sums $\wp_d$ yield the “right” family of global observables of the discrete point process (4). To this end, let us introduce the real random variables

\[
(9) \quad \wp_d^{(N)}(c_1^{(N)}, \ldots, c_N^{(N)}) := \frac{1}{N} \wp_d(c_1^{(N)}, \ldots, c_N^{(N)}), \quad d \in \mathbb{N}^*.
\]

We will refer to these random variables as the Biane-Perelomov-Popov observables of the discrete random point process (4), or BPP observables for short. The relationship between BPP observables and BPP matrices is completely analogous to the relationship between the Newton observables (7) of the continuous random point process and random Hermitian matrices. More precisely, for any $r, d_1, \ldots, d_r \in \mathbb{N}$, we have

\[
(10) \quad \langle \wp_d^{(N)}(\ldots) \rangle = \mathbb{E} \left[ \text{tr}(C_N^{d_1}) \cdots \text{tr}(C_N^{d_r}) \right],
\]

where $\langle \cdot \rangle$ denotes expectation with respect to $\mathbb{P}_N$ on the left hand side, and on the right hand side $\mathbb{E}$ is the quantum expectation functional $\mathbb{E} : A_N \to \mathbb{C}$ and $C_N = A_N + B_N$. This is the perfect quantum analogue of the classical formula (7).
1.4. The semiclassical limit. The relationship (10) between BPP matrices and BPP observables is a useful tool in the analysis of the point process (4) only insofar as the moment method can be adapted to the setting of quantum random matrices.

The generalization of the moment method from random Hermitian matrices to BPP matrices turns out to be a highly non-trivial undertaking, and its complete development constitutes the technical core of the present paper. While it is true that the matrix elements \( \{(A_N)_{ij}\} \) and \( \{(B_N)_{ij}\} \) form two families of commuting (indeed, classically independent) quantum random variables, the members of these families do not commute amongst themselves: in general,

\[
[(A_N)_{ij}, (A_N)_{kl}] \neq 0 \quad \text{and} \quad [(B_N)_{ij}, (B_N)_{kl}] \neq 0.
\]

Instead, the matrix elements of \( A_N, B_N \) are governed by the commutation relations

\[
\begin{align*}
[(A_N)_{ij}, (A_N)_{kl}] &= \hbar_N (\delta_{jk} (A_N)_{il} - \delta_{li} (A_N)_{kj})\\
[(B_N)_{ij}, (B_N)_{kl}] &= \hbar_N (\delta_{jk} (B_N)_{il} - \delta_{li} (B_N)_{kj}),
\end{align*}
\]

which are inherited from the defining relations of \( U(gl_N(\mathbb{C})) \). This non-commutativity adds a new layer of complexity to the moment method, and working with mixed moments in the entries of \( A_N \) and \( B_N \) is vastly more complicated than working with mixed moments in the entries of \( X_N \) and \( Y_N \).

Daunting as this may seem, the situation simplifies considerably if we exploit a parameter which we have so far ignored: the lattice constant \( \hbar_N \). A glance at the commutation relations (11) shows that, if \( \hbar_N \) tends to zero as \( N \) increases, we find ourselves in the semiclassical limit: as \( \hbar_N \) decays, the matrix elements of each BPP matrix tend toward classical (commutative) behaviour, while the pair \( A_N, B_N \) retains its quantum (noncommutative) aspect. It is thus reasonable to hope that, in the semiclassical limit, computations with mixed moments of entries of independent BPP matrices become structurally similar to computations with mixed moments of entries independent random Hermitian matrices. If so, via the identity (10), one may accurately say that BPP matrices provide an effective matrix model with which to analyze the discrete particle ensemble (4).

1.5. Mean values and asymptotic freeness. Let us consider the 1-point functions of the BPP observables, i.e. the mean values

\[
\langle \mathcal{P}_d^{(N)} \rangle = \mathbb{E} \text{tr}(C_N^d), \quad d \in \mathbb{N}^*.
\]
We have
\[ \langle \mathcal{P}_d^{(N)} \rangle = E \text{tr}((A_N + B_N)^d) = \sum_{p,q:d \to N, |p|+|q|=d} E \text{tr}(A_N^{p(1)}B_N^{q(1)} \cdots A_N^{p(d)}B_N^{q(d)}) , \]
so that the computation of the 1-point functions \( \langle \mathcal{P}_d^{(N)} \rangle \) amounts to the computation of the joint distribution of \( A_N, B_N \), viewed as quantum random variables living in the noncommutative probability space \((\text{Mat}_N(A_N), E \text{tr})\), where \( \text{Mat}_N(A_N) \) is the algebra of \( N \times N \) matrices over \( A_N \) and \( E \text{tr} \) is the expected normalized matrix trace.

The joint distribution of \( A_N, B_N \) in the semiclassical limit was first addressed by Biane in [Bia95], who proved that \( A_N, B_N \) are asymptotically free, in the sense of Voiculescu’s free probability theory, provided \( h_N \) decays superpolynomially in \( N \), i.e. \( h_N = o(N^{-k}) \) as \( N \to \infty \) for any \( k \in \mathbb{N}^* \). This result was improved by Collins and Śniady [CS09], who showed that it holds assuming only superlinear decay of the semiclassical parameter, \( h_N = o(N^{-1}) \). In this superlinear regime, the rapid decay of the semiclassical parameter results in the suppression of quantum effects well before the large \( N \) limit. Recently, Bufetov and Gorin [BG15, Conjecture 1.8] have conjectured that \( A_N, B_N \) remain asymptotically free when \( h_N \) decays only linearly in \( N \). This regime is much more delicate, and quantum effects persist into the large \( N \) limit.

In this paper, we prove an optimal result: BPP matrices are asymptotically free in the semiclassical limit \( h_N \to 0, N \to \infty \), with no assumptions on the decay rate of \( h_N \). This subsumes the previous results of Biane and Collins-Śniady, and proves and generalizes the conjecture of Bufetov and Gorin.

**Theorem 1.1.** Suppose that \( h_N = o(1) \) as \( N \to \infty \), and the data \( (2) \) is such that, for each fixed \( k \in \mathbb{N}^* \),
\[ E \text{tr}(A_N^k) = O(1) \quad \text{and} \quad E \text{tr}(B_N^k) = O(1) \]
as \( N \to \infty \). Then, the quantum random matrices \( A_N, B_N \) are asymptotically free with respect to the state \( E \text{tr} \).

As an immediate corollary of Theorem 1.1 we get that the 1-point functions of the random point process \( (4) \) are, in the large \( N \) limit, completely and explicitly determined by the universal formulas of free probability theory in any and all semiclassical regimes. This allows for a complete understanding of the expected semiclassical asymptotic behaviour of \( \text{GL}_N(\mathbb{C}) \) tensor products via the powerful and honed tools of free probability.
Let us also remark that Theorem 1.1 can be turned around: it provides a “quantum alternative” to the usual practice of modelling abstract free random variables using large (classical) random matrices. In this setup, one also has the additional parameter $\hbar_N$, whose decay rate may be chosen at will. This extra degree of freedom could potentially be useful from the point of view of free probability theory.

1.6. Proof strategy. The proof of Theorem 1.1 occupies Sections 2 and 3 below. Our proof strategy is as follows. Fix a positive integer $d \in \mathbb{N}^*$ and a pair of functions $p, q : [d] \to \mathbb{N}$, and let

\[ \tau_N = \mathbb{E} \text{tr}(A_N^{p(1)} B_N^{q(1)} \ldots A_N^{p(d)} B_N^{q(d)}) \]

be the corresponding mixed moment of the BPP matrices $A_N, B_N$. Building on techniques pioneered by Biane in his second groundbreaking paper on asymptotic representation theory [Bia98] (which in fact focused on symmetric groups), we demonstrate that $\tau_N$ decomposes as

\[ \tau_N = \text{Classical}_N + \hbar \text{Quantum}_N, \]

where $\text{Classical}_N$ and $\text{Quantum}_N$ are polynomial functions of the pure moments

\[ \mathbb{E} \text{tr}(A_N), \ldots, \mathbb{E} \text{tr}(A_N^{p|}), \quad \mathbb{E} \text{tr}(B_N), \ldots, \mathbb{E} \text{tr}(B_N^{q|}) \]

The classical part of $\tau_N$ is independent of the Planck constant $\hbar_N$ — its form coincides exactly with the resolution of the classical random matrix mixed moment

\[ \mathbb{E} \text{tr}(X_N^{p(1)} Y_N^{q(1)} \ldots X_N^{p(d)} Y_N^{q(d)}) \]

as a polynomial in the pure moments

\[ \mathbb{E} \text{tr}(X_N), \ldots, \mathbb{E} \text{tr}(X_N^{p|}), \quad \mathbb{E} \text{tr}(Y_N), \ldots, \mathbb{E} \text{tr}(Y_N^{q|}). \]

The quantum part of $\tau_N$, which is a polynomial in the pure moments (14) as well as the parameters $\hbar_N$ and $N$, is present because of the noncommutativity of the entries of BPP matrices, which are governed by the commutation relations (11). In order to move past previous works and free our semiclassical analysis from contrived assumptions on the decay rate of the Planck constant, we must establish unconditional control on the growth of the quantum part. Refining the combinatorial analysis from [Bia98], we demonstrate that the quantum part of the decomposition (13) remains bounded as $N \to \infty$ even when $\hbar_N = \hbar$ is fixed. Thus, when $\hbar_N$ varies with $N$, the classical/quantum decomposition (13) becomes

\[ \tau_N = \text{Classical}_N + O(\hbar_N). \]
It follows that $\tau_N$ agrees with its classical component up to an error determined precisely by the order of magnitude of the semiclassical parameter — in particular, we have

$$\tau_N = \text{Classical}_N + o(1)$$

whenever $\hbar_N = o(1)$ as $N \to \infty$.

The negligibility of the quantum part of $\tau_N$ in the semiclassical limit identifies the classical part as the source of freeness. Going beyond Biane’s computations in [Bia95, Bia98], which relied on techniques of Xu [Xu97] for the computation of polynomial integrals on unitary groups, we use the modern techniques of Weingarten Calculus [Col03, CS06, Nov10] to show that, for any $N \geq d$, the classical part admits an absolutely convergent series expansion of the form

$$\text{Classical}_N = \sum_{k=0}^{\infty} \frac{e_k(N)}{N^{2k}},$$

where each $e_k(N)$ is a polynomial in the pure moments whose coefficients are universal integers enumerating certain special “monotone” paths in the Cayley graph of the symmetric group $S_d$, as generated by the conjugacy class of transpositions. This is a version of the topological expansion familiar from the context of classical random matrix theory, and the leading term $e_0(N)$ is exactly the free probability limit. In particular, our proof of Theorem 1.1 does not rely on prior knowledge that the classical random matrices $X_N, Y_N$ are asymptotically free; rather, it says that any proof which works for $X_N, Y_N$ also works for $A_N, B_N$ in the semiclassical limit.

1.7. Higher correlators and higher order freeness. Theorem 1.1, which shows that the expected asymptotic behaviour of $\text{GL}_N(\mathbb{C})$ tensor products is governed by free probability in any semiclassical regime, should be viewed as the main result of this paper. However, our analysis extends beyond the level of 1-point functions to higher correlators and their connected counterparts, i.e. cumulants. In the classical case, higher connected correlators are governed by the theory of higher order freeness developed in [MSS07], and the techniques used in the proof of Theorem 1.1 may be pushed to show that fluctuations of BPP matrices are also governed by this theory in the semiclassical limit.

In Section 4, we explain in detail how this works for the the covariance

$$\langle \varphi_{d_1}^{(N)} | \varphi_{d_2}^{(N)} \rangle_c = \langle \varphi_{d_1}^{(N)} | \varphi_{d_2}^{(N)} \rangle - \langle \varphi_{d_1}^{(N)} \rangle \langle \varphi_{d_2}^{(N)} \rangle$$

$$= \mathbb{E}[\text{tr}(C_{d_1}^N) \text{tr}(C_{d_1}^d)] - \mathbb{E}[\text{tr}(C_{d_1}^N)] \mathbb{E}[\text{tr}(C_{d_1}^d)].$$

In particular, we establish the following concentration result.
Theorem 1.2. Under the hypotheses of Theorem 1.1, the variance of $\overline{\mathcal{W}}_d^{(N)}$ tends to zero as $N \to \infty$.

Theorems 1.1 and 1.2 together yield, via Chebychev’s inequality, a global Law of Large Numbers for the observables $\overline{\mathcal{W}}_d^{(N)}$ of the discrete particle ensemble (4): whenever $\overline{\mathcal{W}}_d^{(N)}$ converges in expectation to a limit $\gamma_d$, we have $\overline{\mathcal{W}}_d^{(N)} \to \gamma_d$ in probability. This uniformly recovers and generalizes results of Collins-Śniady [CS09] and Bufetov-Gorin [BG15], who worked in the particular semiclassical regimes $\hbar N = o(N^{-1})$ and $\hbar N \sim N^{-1}$, respectively. The LLN obtained here is unconditional; it does not depend on the decay rate of $\hbar N$.

2. Computations at the Planck Scale

In this section, we view the BPP matrices $A_N, B_N$ as quantum random variables living in the noncommutative probability space $(\text{Mat}_N(A_N), \mathbb{E} \text{ tr})$. From this point of view, the joint distribution of $A_N, B_N$ is the data set of all mixed moments

$$\mathbb{E} \text{ tr} \left( A_N^{p(1)} B_N^{q(1)} \cdots A_N^{p(d)} B_N^{q(d)} \right), \quad d \in \mathbb{N}^*, \quad p, q: [d] \to \mathbb{N}. $$

We fix a particular but arbitrary choice of the discrete parameters $d, p, q$, and let $\tau_N$ denote the corresponding mixed moment. We analyze $\tau_N$ at the Planck scale, $\hbar N = \hbar$ fixed, where quantum effects hold full sway. In particular, we derive the classical/quantum decomposition of $\tau_N$ announced above as equation (13). The results of this section are non-asymptotic, i.e. they hold for any $N \in \mathbb{N}^*$.

2.1. Unitary invariance. Our starting point is the following observation of Biane: unitary invariance survives quantization. More precisely, we have the following distributional symmetry of $A_N$ and $B_N$.

Proposition 2.1 ([Bia98 Section 9.2]). Let $U(N)$ denote the group of $N \times N$ complex unitary matrices, and define a function $f_N : U(N) \to \mathbb{C}$ by

$$f_N(U) := \mathbb{E} \text{ tr} \left( U A_N^{p(1)} U^{-1} B_N^{q(1)} \cdots U A_N^{p(d)} U^{-1} B_N^{q(d)} \right).$$

Then, $f_N$ is constant, being equal to $\tau_N$ for all $U \in U(N)$.

As a consequence of Proposition 2.1, we have

$$\tau_N = \int_{U(N)} f_N(U) dU,$$
where the integration is against the unit-mass Haar measure on \( U(N) \). Expanding the trace, this averaging invariance gives us the following representation:

\[
\tau_N = \frac{1}{N} \sum_{r: \{[d]\to [N]\}} \int_{U(N)} \mathbb{E} \left[ U_{r(1)r(2)} \left( A^p_N \right)_{r(2)r(3)} U^{-1}_{r(3)r(4)} \left( B^q_N \right)_{r(4)r(5)} \cdots \right] \\
\cdots U_{r(4d-3)r(4d-2)} \left( A^p_N \right)_{r(4d-2)r(4d-1)} U^{-1}_{r(4d-1)r(4d)} \left( B^q_N \right)_{r(4d)r(1)} dU.
\]

Let us reparameterize the summation index \( r: \{[d]\to [N]\} \) by the quadruple of functions \( i, j, i', j': \{[d]\to [N]\} \) defined by

\[
(r(1), r(2), r(3), r(4), \ldots, r(4d-3), r(4d-2), r(4d-1), r(4d)) = (i(1), j(1), j'(1), i(1), \ldots, i(d), j(d), j'(d), i'(d)).
\]

Then, using the classical independence of the families of (quantum) random variables \( \{(A_N)_{ij}\} \) and \( \{(B_N)_{ij}\} \) in \( (A_N, \mathbb{E}) \), the above becomes

\[
\tau_N = \frac{1}{N} \sum_{i, j, i', j': \{[d]\to [N]\}} I_N(i, j, i', j') \mathbb{E} \left[ \prod_{k=1}^{d} \left( A^p_N \right)_{j(k)j'(k)} \left( B^q_N \right)_{i'(k)i\gamma(k)} \right] \\
= \frac{1}{N} \sum_{i, j, i', j': \{[d]\to [N]\}} I_N(i, j, i', j') \mathbb{E} \left[ \prod_{k=1}^{d} \left( A^p_N \right)_{j(k)j'(k)} \right] \mathbb{E} \left[ \prod_{k=1}^{d} \left( B^q_N \right)_{i'(k)i\gamma(k)} \right],
\]

where \( \gamma = (1 2 \ldots d) \) is the full forward cycle in the symmetric group \( \mathfrak{S}_d \), and

\[
I_N(i, j, i', j') := \int_{U(N)} \prod_{k=1}^{d} U_{i(k)j(k)} U_{j'(k)i'(k)} dU.
\]

2.2. The Weingarten function. Matrix integrals of the form \ref{eq:18} have a long history in mathematical physics; they appear in contexts ranging from lattice gauge theory to quantum chromodynamics and string theory, see e.g. \cite{BB96, BDW77, GT93, Sam80, Xu97}. In the context of free probability and random matrices, these integrals were treated by Collins \cite{Col03} and Collins–Śniady \cite{CS06}, who proved that

\[
I_N(i, j, i', j') = \sum_{(\alpha, \beta) \in \mathfrak{S}_d^2} \delta_{\gamma, i\alpha} \delta_{\gamma, j\beta} W_{g_N}(\alpha, \beta),
\]

where

\[
W_{g_N}: \mathfrak{S}_d^2 \to \mathbb{Q}
\]
is a special function on pairs of permutations which they named the \textit{Wein-
garten function}.

There are now several descriptions of the \textit{Weingarten function} available; in
this paper, we will use a series expansion of $W_N^\alpha$ obtained by Novak \cite{Novak2010} and Matsumoto-Novak \cite{MatsumotoNovak2013}, which is explained in Section 3. For now, plugging \eqref{eq:19} into our calculation above eliminates the indices $i',j'$ and produces the formula

\begin{equation}
\tau_N = \frac{1}{N} \sum_{(\alpha,\beta) \in S_2^d} W_N^\alpha(\alpha,\beta) \mathbb{E}[S_\alpha] \mathbb{E}[S_{\beta^{-1}}], \tag{21}
\end{equation}

where

\begin{equation}
S_\alpha := \sum_{i: [d] \to [N]} \prod_{k=1}^d \left( A_N^{(k)} \right)_{i(k)i^{\alpha(k)}} \tag{22}
\end{equation}

and

\begin{equation}
S_{\beta^{-1}} := \sum_{i: [d] \to [N]} \prod_{k=1}^d \left( B_N^{q(k)} \right)_{i(k)i^{\beta^{-1}(k)}}. \tag{23}
\end{equation}

Our goal now is to express the operators \eqref{eq:22} and \eqref{eq:23} in terms of the operators

\[ \text{tr}(A_N), \text{tr}(A_N^2), \ldots \quad \text{and} \quad \text{tr}(B_N), \text{tr}(B_N^2), \ldots, \]

a task which is non-trivial due to the fact that the matrix elements of $A_N$ and $B_N$ do not commute. It is advantageous to lift this problem to the universal enveloping algebra $\mathcal{U}(\mathfrak{gl}_N)$.

2.3. \textbf{Casimirs and Biasimirs.} Let $Z_N$ be the $N \times N$ matrix over $\mathcal{U}(\mathfrak{gl}_N)$ with elements $(Z_N)_{ij} = \hbar e_{ij}$. This matrix was introduced by Perelomov and Popov \cite{PerelomovPopov1967}, who studied traces of its powers,

\[ C_k := \text{Tr} Z_N^k = \sum_{i: [k] \to [N]} (Z_N)^{i(1)i(2)} \cdots (Z_N)^{i(k)i(1)}, \quad k \in \mathbb{N}^*, \]

which they called \textit{higher Casimirs}, see also \cite{Zelikin1973}. This nomenclature stems from the fact that, up to a multiplicative factor of $\hbar^2$, the element $C_2$ coincides with the usual Casimir element which resides in the center $Z_N$ of $\mathcal{U}(\mathfrak{gl}_N)$. The following Theorem summarizes the main properties of higher Casimirs.

\textbf{Theorem 2.2} (\cite{PerelomovPopov1967, Zelikin1973}). \textit{The higher Casimirs generate $Z_N$ as a polynomial ring,}

\[ Z_N = \mathbb{C}[C_1, C_2, C_3, \ldots]. \]
Moreover, if \((X, \rho)\) is the irreducible representation of \(\text{GL}_N(\mathbb{C})\) indexed by the particle configuration \(c_1 > \cdots > c_N\) on \(\mathbb{Z}[\hbar]\), the image of \(C_k\) in this representation is the scalar operator

\[ \rho(C_k) = \varphi_k(c_1, \ldots, c_N)I_X \]

with eigenvalue

\[ \varphi_k(c_1, \ldots, c_N) = \sum_{i=1}^{N} \prod_{j \neq i} \left( 1 - \frac{\hbar}{c_i - c_j} \right) c_i^k. \]

Note that traces of powers of our quantum random matrices \(A_N\) and \(B_N\), which are operators acting in \(V_N \otimes W_N\), are essentially images of higher Casimirs in irreducible representations; more precisely, we have

\[ \begin{align*}
\text{Tr}(A^k_N) &= \rho_N(C_k) \otimes I_{W_N} \\
\text{Tr}(B^l_N) &= I_{V_N} \otimes \sigma_N(C_l).
\end{align*} \tag{24} \]

In particular, by Theorem 2.2, these traces are the following scalar operators,

\[ \begin{align*}
\text{Tr}(A^k_N) &= \varphi_k(a^{(N)}_1, \ldots, a^{(N)}_N)I_{V_N} \otimes I_{W_N} \\
\text{Tr}(B^l_N) &= \varphi_k(b^{(N)}_1, \ldots, b^{(N)}_N)I_{V_N} \otimes I_{W_N}.
\end{align*} \tag{25} \]

We conclude that, for any \(k, l \in \mathbb{N}^*\), the operators \(\text{Tr}(A^k_N)\) and \(\text{Tr}(A^l_N)\) are classically independent quantum random variables in \((A_N, \mathbb{E})\) with known distributions, and similarly for the operators \(\text{Tr}(B^k_N), \text{Tr}(B^l_N)\).

In order to understand the operators \(S_\alpha, S_{\beta^{-1}\gamma}\) which appear in our formula (21) for \(\tau_N\), we must understand certain elements of \(\mathcal{U}(\mathfrak{gl}_N)\) which further generalize higher Casimirs. More precisely, we have that

\[ \begin{align*}
S_\alpha &= \rho_N(C^{(\mathfrak{gl})}_\alpha) \otimes I_{W_N} \\
S_{\beta^{-1}\gamma} &= I_{V_N} \otimes \sigma_N(C^{(\mathfrak{gl})}_{\beta^{-1}\gamma}),
\end{align*} \tag{26} \]

where, for any permutation \(\pi \in \mathfrak{S}_d\) and function \(r: [d] \to \mathbb{N}\), we define

\[ C^{(r)}_\pi := \sum_{i: [d] \to [N]} (Z^{r(1)}_N)^{i(1)i\pi(1)} \cdots (Z^{r(d)}_N)^{i(d)i\pi(d)}. \tag{27} \]

Elements in \(\mathcal{U}(\mathfrak{gl}_N)\) of the form (27) were first considered by Biane in [Bia98], and we shall refer to them as Biasimirs, a portmanteau of “Biane” and “Casimir”. Indeed, if \(\pi = \gamma\) is the full forward cycle in \(\mathfrak{S}_d\), then \(C^{(r)}_\pi\) reduces to the higher Casimir \(C_{[r]}\).

Let us look at some examples of Biasimirs. As an easy example, take \(d = 5\) and \(\pi \in \mathfrak{S}_5\) to be the permutation \(\pi = (1 \ 2 \ 3)(4 \ 5)\). Then, for any
More generally, whenever \( \pi \in S_d \) is a canonical permutation, i.e. a permutation of the form
\[
\pi = (1 \ 2 \ \ldots \ n_1)(n_1 + 1 \ n_1 + 2 \ \ldots \ n_1 + n_2) \cdots ,
\]
for some composition \((n_1, n_2, \ldots)\) of \(d\), the corresponding Biasimir will be a simple monomial function of Casimirs. To be precise, if \( \pi = \gamma_1 \gamma_2 \cdots \gamma_k \) is the disjoint cycle decomposition of a canonical permutation \( \pi \), then
\[
C^{(r)}_\pi = \prod_{j=k}^\ell C_{\sum_{i \in \gamma_j} r(i)}.
\]
Biasimirs corresponding to non-canonical permutations are more complicated functions of higher Casimirs. For example, consider the Biasimir of degree \(d = 3\) corresponding to the non-canonical permutation \( \pi = (1 \ 3 \ 2) \) and some general power function \( r \),
\[
C^{(r)}_\pi = \sum_{i: [3] \to [N]} (Z^{(1)}_N)_{i(1)i(3)} (Z^{(2)}_N)_{i(2)i(1)} (Z^{(3)}_N)_{i(3)i(2)}.
\]
This is not the higher Casimir \( C_{(r(1)+r(2)+r(3))} \), because the factors in each term of the sum are in the wrong order. However, we can sort the letters in each summand using the commutation relations
\[
[(Z^{(n)}_N)_{ij}, (Z^{(n)}_N)_{kl}] = \hbar N \left[ \delta_{jk} (Z^{m+n-1}_N)_{il} - \delta_{il} (Z^{m+n-1}_N)_{kj} \right],
\]
which are easily deduced from the defining relations of \( U(gl_N) \). Carrying this out and summing over all \( i: [3] \to [N] \), we obtain
\[
C^{(r)}_\pi = \text{Tr} \ Z^{(1)}_N Z^{(2)}_N Z^{(3)}_N + \hbar \text{Tr} \ Z^{(1)}_N \text{Tr} Z^{(2)\ast}_N Z^{(3)\ast}_N + \hbar N \text{Tr} \ Z^{(1)}_N Z^{(2)\ast}_N Z^{(3)\ast}_N - h N C_{(r(1)+r(2)+r(3)-1)}.
\]
In general, we have the following polynomial representation of Biasimirs in terms of higher Casimirs, which is a slightly refined version of [Bia98, Lemma 8.4.1].

**Proposition 2.3.** For any permutation \( \pi \in S_d \), and any function \( r: [d] \to \mathbb{N} \), there exist unique polynomials \( P^{(r)}_\pi \) and \( Q^{(r)}_\pi \) in \(|r|\) and \(|r| + 2\) variables, respectively, such that
\[
C^{(r)}_\pi = P^{(r)}_\pi (C_1, \ldots, C_{|r|}) + h Q^{(r)}_\pi (\hbar, N, C_1, \ldots, C_{|r|})
\]
holds for all \( N \in \mathbb{N}^* \).
We refer to the polynomials $P^{(r)}_\pi$ and $Q^{(r)}_\pi$ as the “classical” and “quantum” components of the Biasimir $C^{(r)}_\pi$. The classical component is simple, being given by the right hand side of formula (28) above; the quantum component is more complicated. Returning to our previous example, where $\pi \in S_3$ is the cyclic permutation $\pi = (1 \ 3 \ 2)$, we have

$$P^{(r)}_\pi(C_1, \ldots, C_{|\pi|}) = C_{r(1)+r(2)+r(3)},$$

$$Q^{(r)}_\pi(h_N, N, C_1, \ldots, C_{|\pi|}) = C_{r(1)}C_{r(2)+r(3)-1} - NC_{r(1)+r(2)+r(3)-1}$$

for the classical and quantum components of the Biasimir $C^{(r)}_\pi$.

Let us view the quantum component $Q^{(r)}_\pi$ of a given Biasimir $C^{(r)}_\pi$ as an element of the polynomial ring $\mathbb{Z}[\hbar][N, C_1, \ldots, C_{|\pi|}]$. On this polynomial ring we impose the grading in which each variable $N, C_1, \ldots, C_{|\pi|}$ has degree one. Let $\text{cyc}(\pi)$ denote the number of factors in the decomposition of $\pi$ into disjoint cyclic permutations, and let $\text{aex}(\pi)$ denote the number of antiexceedances of the permutation $\pi$, that is the number of indices $i \in [d]$ such that $\pi(i) \leq i$. The following Proposition is a small refinement of [Bia98, Proposition 8.5].

**Proposition 2.4.** The degree of the classical component $P^{(r)}_\pi$ is $\text{cyc}(\pi)$. The degree of the quantum component $Q^{(r)}_\pi$ is at most $\text{aex}(\pi)$.

### 2.4. Classical/Quantum decomposition.

We are now ready to obtain the decomposition (13) of $\tau_N$ into classical and quantum parts. Let us return to the formula (21) for $\tau_N$, and consider a particular term in the sum corresponding to the pair $(\alpha, \beta) \in S_d^2$.

First, by Proposition 2.3, we have that

$$S_\alpha = P^{(p)}_\alpha(\text{Tr} \ A_N, \ldots, \text{Tr} \ A_{N}^{[p]}) + hQ^{(p)}_\alpha(h, N, \text{Tr} \ A_N, \ldots, \text{Tr} \ A_{N}^{[p]}).$$

Let us rewrite this in terms of normalized traces. Put

$$\overline{P}^A_\alpha := \frac{1}{N^{\text{cyc}(\alpha)}}P^{(p)}_\alpha(\text{Tr} \ A_N, \ldots, \text{Tr} \ A_{N}^{[p]}),$$

$$\overline{Q}^A_\alpha := \frac{1}{N^{\text{aex}(\alpha)}}Q^{(p)}_\alpha(h, N, \text{Tr} \ A_N, \ldots, \text{Tr} \ A_{N}^{[p]})$$

From (28), $\overline{P}^A_\alpha$ is an explicit polynomial in the operators

$$\text{tr} \ A_N, \text{tr} \ A_{N}^{2}, \ldots, \text{tr} \ A_{N}^{[p]},$$

while Proposition 2.4 implies that $\overline{Q}^A_\alpha$ is a polynomial in the numbers $h, N^{-1}$ and the operators

$$\text{tr} \ A_N, \text{tr} \ A_{N}^{2}, \ldots, \text{tr} \ A_{N}^{[p]}.$$
Now we apply the expectation $\mathbb{E}$ to both sides of this identity in $\mathcal{A}_N$ to get an identity in $\mathbb{C}$. Because traces of powers of $A_N$ are classically independent, we have

$$\mathbb{E} P^A_{\alpha} (N) := \mathbb{E} \mathbb{E}^A_{\alpha} = \frac{1}{N^{\text{cyc}(\alpha)}} P^{(p)}_{\alpha} \left( \mathbb{E} \text{Tr} A_N, \ldots, \mathbb{E} \text{Tr} A_N^{[p]} \right),$$

$$\mathbb{E} Q^A_{\alpha} (N) := \mathbb{E} \mathbb{Q}^A_{\alpha} = \frac{1}{N^{\text{aex}(\alpha)}} Q^{(p)}_{\alpha} \left( h, N, \mathbb{E} \text{Tr} A_N, \ldots, \mathbb{E} \text{Tr} A_N^{[p]} \right);$$

the first of these is a polynomial in the numbers

$$\mathbb{E} \text{tr} A_N, \mathbb{E} \text{tr} A_N^2, \ldots, \mathbb{E} \text{tr} A_N^{[p]},$$

while the second is a polynomial in the numbers

$$h, N^{-1}, \mathbb{E} \text{tr} A_N, \mathbb{E} \text{tr} A_N^2, \ldots, \mathbb{E} \text{tr} A_N^{[p]}.$$

We conclude that

$$\mathbb{E} [S_{\alpha}] = N^{\text{cyc}(\alpha)} P^A_{\alpha} (N) + h N^{\text{aex}(\alpha)} Q^A_{\alpha} (N).$$

Second, we have that

$$S_{\beta-1, \gamma} = P^{(q)}_{\beta-1, \gamma} (\text{Tr} B_N, \ldots, \text{Tr} B_N^{[q]}) + h Q^{(q)}_{\beta-1, \gamma} (h, N, \text{Tr} B_N, \ldots, \text{Tr} B_N^{[q]}).$$

Once again, let us rewrite this in terms of normalized traces. Put

$$\mathbb{E} P^B_{\beta-1, \gamma} := \frac{1}{N^{\text{cyc}(\beta-1, \gamma)}} P^{(q)}_{\beta-1, \gamma} \left( \text{Tr} B_N, \ldots, \text{Tr} B_N^{[q]} \right),$$

$$\mathbb{E} Q^B_{\beta-1, \gamma} := \frac{1}{N^{\text{aex}(\beta-1, \gamma)}} Q^{(q)}_{\beta-1, \gamma} \left( h, N, \text{Tr} B_N, \ldots, \text{Tr} B_N^{[q]} \right).$$

From (28), $P^B_{\beta-1, \gamma}$ is an explicit polynomial in the operators

$$\text{tr} B_N, \text{tr} B_N^2, \ldots, \text{tr} B_N^{[q]},$$

while Proposition 2.4 implies that $Q^B_{\beta-1, \gamma}$ is a polynomial in the numbers $h, N^{-1}$ and the operators

$$\text{tr} B_N, \text{tr} B_N^2, \ldots, \text{tr} B_N^{[q]}.$$

We thus have

$$S_{\beta-1, \gamma} = N^{\text{cyc}(\beta-1, \gamma)} P^B_{\beta-1, \gamma} + h N^{\text{aex}(\beta-1, \gamma)} Q^B_{\beta-1, \gamma}.$$

Once again, we apply $\mathbb{E}$ to both sides of this identity in $\mathcal{A}_N$ to get an identity in $\mathbb{C}$. As above, we declare

$$\mathbb{E} P^B_{\beta-1, \gamma} (N) := \mathbb{E} P^B_{\beta-1, \gamma} = \frac{1}{N^{\text{cyc}(\beta-1, \gamma)}} P^{(q)}_{\beta-1, \gamma} \left( \mathbb{E} \text{Tr} B_N, \ldots, \mathbb{E} \text{Tr} B_N^{[q]} \right),$$

$$\mathbb{E} Q^B_{\beta-1, \gamma} (N) := \mathbb{E} Q^B_{\beta-1, \gamma} = \frac{1}{N^{\text{aex}(\beta-1, \gamma)}} Q^{(q)}_{\beta-1, \gamma} \left( h, N, \mathbb{E} \text{Tr} B_N, \ldots, \mathbb{E} \text{Tr} B_N^{[q]} \right).$$
The first of these is a polynomial in the numbers
\[ \mathbb{E} \text{tr} B_N, \mathbb{E} \text{tr} B_N^2, \ldots, \mathbb{E} \text{tr} B_N^{|q|}, \]
while the second is a polynomial in the numbers
\[ h, N^{-1}, \mathbb{E} \text{tr} B_N, \mathbb{E} \text{tr} B_N^2, \ldots, \mathbb{E} \text{tr} B_N^{|q|}. \]

We conclude that
\[ \mathbb{E} \left[ S_{\beta^{-1}\gamma} \right] = N^{\text{cyc}(\beta^{-1}\gamma)} P_\beta^{(\beta^{-1}\gamma)}(N) + h N^{\text{aex}(\beta^{-1}\gamma)} Q_\beta^{(\beta^{-1}\gamma)}(N). \]

Putting these two calculations together, we compute the \((\alpha, \beta)\) term of \(\tau_N\) as
\[ Wg_N(\alpha, \beta) \mathbb{E}[S_\alpha] \mathbb{E}[S_{\beta^{-1}\gamma}] = \]
\[ Wg_N(\alpha, \beta) \left( N^{\text{cyc}(\alpha)} P_\alpha^{(\alpha)}(N) + h N^{\text{aex}(\alpha)} Q_\alpha^{(\alpha)}(N) \right) \times \]
\[ \times \left( N^{\text{cyc}(\beta^{-1}\gamma)} P_\beta^{(\beta^{-1}\gamma)}(N) + h N^{\text{aex}(\beta^{-1}\gamma)} Q_\beta^{(\beta^{-1}\gamma)}(N) \right). \]

Expanding the brackets and summing \((\alpha, \beta)\) over \(S_d^2\), we arrive at the classical/quantum decomposition of the mixed moment \(\tau_N\).

**Theorem 2.5.** We have
\[ \tau_N = \text{Classical}_N + h N \text{ Quantum}_N, \]
where
\[ \text{Classical}_N = \frac{1}{N} \sum_{(\alpha, \beta) \in S_d^2} N^{\text{cyc}(\alpha) + \text{cyc}(\beta^{-1}\gamma)} Wg_N(\alpha, \beta) P_\alpha^{(\alpha)}(N) P_\beta^{(\beta^{-1}\gamma)}(N) \]
and
\[ \text{Quantum}_N = \frac{1}{N} \sum_{(\alpha, \beta) \in S_d^2} \left( N^{\text{cyc}(\alpha) + \text{aex}(\beta^{-1}\gamma)} Wg_N(\alpha, \beta) P_\alpha^{(\alpha)}(N) Q_\beta^{(\beta^{-1}\gamma)}(N) \right) \]
\[ + N^{\text{aex}(\alpha) + \text{aex}(\beta^{-1}\gamma)} Wg_N(\alpha, \beta) Q_\alpha^{(\alpha)}(N) P_\beta^{(\beta^{-1}\gamma)}(N) \]
\[ + h N^{\text{aex}(\alpha) + \text{aex}(\beta^{-1}\gamma)} Wg_N(\alpha, \beta) Q_\alpha^{(\alpha)}(N) Q_\beta^{(\beta^{-1}\gamma)}(N) \right). \]

**3. ASYMPTOTIC COMPUTATIONS**

In this Section, we apply the exact results obtained in Section 2 to analyze the asymptotic behaviour of the mixed moment \(\tau_N\) in the limit where \(N \to \infty\) and \(h N \to 0\). We adopt the hypotheses of Theorem 1.1, which is to say that we henceforth assume the expectations
\[ \mathbb{E} \text{tr}(A_N^k) \quad \text{and} \quad \mathbb{E} \text{tr}(B_N^k) \]
remain bounded as \(N \to \infty\) for each fixed \(k \in \mathbb{N}^*\).
3.1. The Weingarten function. A key component of our asymptotic analysis will be an absolutely convergent series expansion for the Weingarten function which renders its asymptotic behaviour transparent.

In order to state this expansion, let us identify the symmetric group $\mathfrak{S}_d$ with its right Cayley graph, as generated by the conjugacy class of transpositions. We denote by $|\cdot|$ the corresponding word norm, so that $|\alpha^{-1} \beta|$ is the graph theory distance from $\alpha$ to $\beta$, i.e. the length of a geodesic path in the Cayley graph joining these two permutations. Equip the Cayley graph with the Biane–Stanley edge labelling, in which each edge corresponding to the transposition $(s\, t)$ is marked by $t$, the larger of the two elements interchanged. This edge labelling was introduced in the context of enumerative combinatorics by Stanley [Sta97] and Biane [Bia02] as a means to relate parking functions and noncrossing partitions. Figure 1 shows $\mathfrak{S}_4$ with the Biane-Stanley labelling, where 2-edges are drawn in blue, 3-edges in yellow, and 4-edges in red.

A walk on $\mathfrak{S}_d$ is said to be monotone if the labels of the edges it traverses form a weakly increasing sequence. The fundamental fact we need [Nov10, MN13] is that the Weingarten function expands as a generating function for monotone walks: we have

$$(30) \quad Wg_N(\alpha, \beta) = \frac{1}{N^d} \sum_{r=0}^{\infty} (-1)^r \frac{\tilde{W}^r(\alpha, \beta)}{N^r},$$

where $\tilde{W}^r(\alpha, \beta)$ is the number of $r$-step monotone walks on $\mathfrak{S}_d$ which begin at the permutation $\alpha$ and end at the permutation $\beta$. This series is absolutely convergent provided $N \geq d$, but divergent if $N < d$ (this divergence is a related to the De Wit–’t Hooft anomalies in $U(N)$ lattice gauge theory, see e.g. [BDW77, Mor09, Sam80]).

Since $\tilde{W}^r(\alpha, \beta) = \tilde{W}^r(\text{id}, \alpha^{-1} \beta)$, and since every permutation is either even or odd, the number $\tilde{W}^r(\alpha, \beta)$ is nonzero if and only if $r = |\alpha^{-1} \beta| + 2g$ for some $g \in \mathbb{N}$. We may thus rewrite (30) as

$$(31) \quad Wg_N(\alpha, \beta) = \frac{(-1)^{|\alpha^{-1} \beta|}}{N^{d+|\alpha^{-1} \beta|}} \sum_{g=0}^{\infty} \frac{\tilde{W}_g(\alpha, \beta)}{N^{2g}},$$

where $\tilde{W}_g(\alpha, \beta) := \tilde{W}^{|\alpha^{-1} \beta|+2g}(\alpha, \beta)$. The formulas (19) and (31) may be effectively combined to yield a sort of Feynman calculus for unitary matrix integrals, in which the role of Feynman diagrams is played by monotone walks on symmetric groups, see e.g. of [GGPN16b, GGPN16a].

3.2. Quantum asymptotics. We now show that the quantum part of $\tau_N$ can be controlled even for $\hbar_N$ fixed. In order to do this, we introduce a new
permutation statistic defined by
\begin{equation}
\text{defect}(\pi) = \text{aex}(\pi) - \text{cyc}(\pi), \quad \pi \in S_d.
\end{equation}
Moreover, for any $k \in \mathbb{N}^*$ and $(\pi_1, \ldots, \pi_k) \in \mathcal{G}_d^k$, the quantity
\begin{equation}
\text{genus}(\pi_1, \ldots, \pi_k) := \frac{|\pi_1| + \cdots + |\pi_k| - |\pi_1 \cdots \pi_k|}{2}
\end{equation}
is a nonnegative integer; we refer to it as the genus of the $k$-tuple $(\pi_1, \ldots, \pi_k)$.

**Lemma 3.1.** For any $\pi \in \mathcal{G}_d$, we have
\[\text{defect}(\pi) \geq 0.\]
Moreover, for any $(\alpha, \beta) \in \mathcal{G}_d^2$, we have
\[\text{defect}(\alpha) + \text{defect}(\beta^{-1}\gamma) \leq 2 \text{genus}(\alpha, \alpha^{-1}\beta, \beta^{-1}\gamma).\]

**Proof.** The first part is obvious, since each cycle of a permutation gives at least one contribution to the number of antiexceedances.

Our proof of the second part will be based on revisiting the work of Biane. First, we must present a small erratum to his work. In [Bia98], at the top of Page 173, Biane states that
\begin{equation}
-d - 2 - |\alpha^{-1}\beta| + \text{aex}(\alpha) + \text{aex}(\alpha) + \text{aex}(\gamma\epsilon\beta^{-1}) \leq 0
\end{equation}
for any transposition \( \epsilon \in \mathcal{S}_d \) (note that Biane’s notation is different from ours: his \( q_2 \) is our \( d \), his \( \sigma \) is our \( \alpha \), his \( \tau \) is our \( \beta \), and he denotes by \( W \) the product \( \gamma \epsilon \) of the long cycle and a transposition). This statement of Biane is incorrect and the correct statement is

\[
-d - 2 - |\alpha^{-1}\beta| + \text{aex}(\alpha) + \text{aex}(\gamma\epsilon\beta^{-1}) \leq 0.
\]

Also, in the last step of the proof of this inequality Biane claims that he uses part (1) of [Bia98, Lemma 8.2] — actually, the right tool is part (2) of the same lemma.

We are now ready to establish the inequality which is our aim. Just like Biane, we show that for arbitrary permutations \( \pi, \phi \in \mathcal{S}_d \)

\[
\text{aex}(\gamma\phi^{-1}) = d + 1 - \text{aex}(\phi) \leq d + 1 + |\pi^{-1}\phi| - \text{aex}(\pi).
\]

Setting \( \phi := \alpha^{-1}Z, \pi := \beta^{-1}\gamma \) we obtain our target inequality. \( \square \)

**Theorem 3.2.** For any fixed \( h_N = h \), Quantum \( N \) is \( O(1) \) as \( N \to \infty \).

**Proof.** By Theorem 2.5, the quantum part of \( \tau_N \) may be written as

\[
\text{Quantum}_N = \sum_{(\alpha,\beta)\in \mathcal{S}_d^2} N^{\text{cyc}(\alpha)+\text{cyc}(\beta^{-1}\gamma)-1} Wg_{N}(\alpha,\beta) R_{(\alpha,\beta)}(N)
\]

where

\[
R_{(\alpha,\beta)}(N) = N^{\text{defect}(\beta^{-1}\gamma)} P_A^A(N)Q_B^B(\beta^{-1}\gamma)(N)
\]

\[
+ N^{\text{defect}(\alpha)} Q_A^A(N)P_B^B(\beta^{-1}\gamma)(N)
\]

\[
+ h N^{\text{defect}(\alpha)+\text{defect}(\beta^{-1}\gamma)}(N)Q_A^A(N)Q_B^B(\beta^{-1}\gamma)(N).
\]

We will show that each term in the sum (36) is \( O(1) \).

By the first part of Lemma 3.1, nonnegativity of the defect statistic, we have

\[
R_{(\alpha,\beta)}(N) = O\left(N^{\text{defect}(\alpha)+\text{defect}(\beta^{-1}\gamma)}\right)
\]

for each \( (\alpha, \beta) \in \mathcal{S}_d^2 \).

Now, let us consider the order of the factor

\[
N^{\text{cyc}(\alpha)+\text{cyc}(\beta^{-1}\gamma)-1} Wg_{N}(\alpha,\beta).
\]

Invoking the expansion (31), for any \( N \geq d \) we have

\[
N^{\text{cyc}(\alpha)+\text{cyc}(\beta^{-1}\gamma)-1} Wg_{N}(\alpha,\beta) = N^{\text{cyc}(\alpha)+\text{cyc}(\beta^{-1}\gamma)-1} \frac{(-1)^{|\alpha^{-1}\beta|}}{N^d+|\alpha^{-1}\beta|} \sum_{g=0}^{\infty} \frac{\tilde{W}_g(\alpha,\beta)}{N^{2g}}
\]

\[
= N^{-|\alpha|-|\alpha^{-1}\beta|-|\beta^{-1}\gamma|+|\gamma|} \sum_{g=0}^{\infty} \frac{\tilde{W}_g(\alpha,\beta)}{N^{2g}}
\]

\[
= O(N^{-2\text{genus}(\alpha,\alpha^{-1}\beta,\beta^{-1}\gamma)}).
\]
We conclude that each term of $\text{Quantum}_N$ is of order

$$N^{\cyc(\alpha) + \cyc(\beta^{-1}) - 1} \text{Wg}_N(\alpha, \beta)R_{(\alpha, \beta)}(N) = O\left( N^{\text{defect}(\alpha) + \text{defect}(\beta^{-1}) - 2 \text{genus}(\alpha, \alpha^{-1}, \beta, \beta^{-1})} \right),$$

and hence is $O(1)$ by the second part of Lemma 3.1.

3.3. Classical asymptotics. We now deal with the asymptotics of the classical part of $\tau_N$.

**Theorem 3.3.** Under the hypotheses of Theorem 1.1, the classical part of $\tau_N$ admits, for each $N \geq d$, the absolutely convergent series expansion

$$\text{Classical}_N = \sum_{k=0}^{\infty} \frac{e_k(N)}{N^{2k}},$$

where

$$e_k(N) = \sum_{(g,h) \in \mathbb{N}^2} \sum_{g+h=k \text{ genus}(\alpha, \alpha^{-1}, \beta, \beta^{-1}, \gamma) = h} (-1)^{|\alpha^{-1} \beta|} \bar{W}_g(\alpha, \beta) P^A_\alpha(N) P^B_{\beta^{-1} \gamma}(N).$$

**Proof.** According to Theorem 2.5 and the expansion (31), we have

$$\text{Classical}_N = \sum_{(\alpha, \beta) \in \mathbb{S}_d^2} N^{\cyc(\alpha) + \cyc(\beta^{-1}) - 1} \text{Wg}_N(\alpha, \beta) P^A_\alpha(N) P^B_{\beta^{-1} \gamma}(N) \sum_{g=0}^{\infty} \frac{\bar{W}_g(\alpha, \beta)}{N^{2g}}$$

$$= \sum_{g,h=0}^{\infty} \frac{1}{N^{2(g+h)}} \sum_{(\alpha, \beta) \in \mathbb{S}_d^2 \text{ genus}(\alpha, \alpha^{-1}, \beta, \beta^{-1}, \gamma) = h} (-1)^{|\alpha^{-1} \beta|} \bar{W}_g(\alpha, \beta) P^A_\alpha(N) P^B_{\beta^{-1} \gamma}(N)$$

$$= \sum_{k=0}^{\infty} \sum_{g,h \geq 0} \frac{1}{N^k} \sum_{g+h=k \text{ genus}(\alpha, \alpha^{-1}, \beta, \beta^{-1}, \gamma) = h} (-1)^{|\alpha^{-1} \beta|} \bar{W}_g(\alpha, \beta) P^A_\alpha(N) P^B_{\beta^{-1} \gamma}(N).$$

3.4. Semiclassical asymptotics and freeness. Combining Theorems 3.2 and 3.3, we obtain the following corollary.

**Corollary 3.4.** For any sequence $h_N$, we have

$$\tau_N = \text{Classical}_N + O(h_N)$$

as $N \to \infty$. In particular, if $h_N = o(N^{-2l})$ as $N \to \infty$, then

$$\tau_N = \text{Classical}_N + o(N^{-2l}) = \sum_{k=0}^{l} \frac{e_k(N)}{N^{2k}} + o(N^{-2l}).$$
The $l = 0$ case of Corollary 3.4 yields
\[ \tau_N = \text{Classical}_N + o(1) = e_0(N) + o(1) \]
as $N \to \infty$, with
\[ e_0(N) = \sum_{(\alpha,\beta) \in \mathbb{S}_d^2 \text{ genus}(\alpha,\alpha-1,\beta,\beta-1,\gamma)=0} (-1)^{|\alpha-1,\beta|} \tilde{W}_0(\alpha, \beta) P^A_\alpha(N) P^B_{\beta-1,\gamma}(N). \]

Theorem 1.1 is proved.

4. Covariance of traces of BPP matrices

4.1. Covariance setup. Let $C_N = A_N + B_N$. Let $d_1, d_2 \in \mathbb{N}^*$ be a pair of positive integers. Then, $\text{tr}(C^{d_1}_N)$ and $\text{tr}(C^{d_2}_N)$ are random variables in the noncommutative probability space $(\mathcal{A}_N, \mathbb{E})$. Our goal is to compute the covariance of these two quantum random variables,

\[ \text{cov}\left( \text{tr}(C^{d_1}_N), \text{tr}(C^{d_2}_N) \right) = \mathbb{E}\left[ \text{tr}(C^{d_1}_N) \text{tr}(C^{d_2}_N) \right] - \mathbb{E}\left[ \text{tr}(C^{d_1}_N) \right] \mathbb{E}\left[ \text{tr}(C^{d_2}_N) \right]. \]

Let us start with the second term, i.e. the product of expected traces. For the first factor in this product, we have

\[ \mathbb{E} \text{tr}(C^{d_1}_N) = \mathbb{E}\left[ \text{tr}\left( (A_N + B_N)^{d_1} \right) \right] = \sum_{p_1, q_1 : [d_1] \rightarrow \mathbb{N}} \mathbb{E} \text{tr}\left( A^{p_1(1)}_N B^{q_1(1)}_N \cdots A^{p_1(d_1)}_N B^{q_1(d_1)}_N \right). \]

Similarly, for the second factor we have

\[ \mathbb{E} \text{tr}(C^{d_2}_N) = \mathbb{E}\left[ \text{tr}\left( (A_N + B_N)^{d_2} \right) \right] = \sum_{p_2, q_2 : [d_2] \rightarrow \mathbb{N}} \mathbb{E} \text{tr}\left( A^{p_2(1)}_N B^{q_2(1)}_N \cdots A^{p_2(d_2)}_N B^{q_2(d_2)}_N \right). \]

Thus, the second term of the covariance is

\[ \mathbb{E}\text{tr}(C^{d_1}_N)\mathbb{E}\text{tr}(C^{d_2}_N) = \sum_{p_1, q_1 : [d_1] \rightarrow \mathbb{N}} \sum_{p_2, q_2 : [d_2] \rightarrow \mathbb{N}} \mathbb{E} \text{tr}\left( A^{p_1(1)}_N B^{q_1(1)}_N \cdots A^{p_1(d_1)}_N B^{q_1(d_1)}_N \right) \mathbb{E} \text{tr}\left( A^{p_2(1)}_N B^{q_2(1)}_N \cdots A^{p_2(d_2)}_N B^{q_2(d_2)}_N \right). \]
The first term of the covariance is
\[
\mathbb{E}[\text{tr}(C_{N}^{d_1}) \text{tr}(C_{N}^{d_2})] = \mathbb{E}[\text{tr}((A_N + B_N)^{d_1}) \text{tr}((A_N + B_N)^{d_2})]
\]
\[
= \sum_{p_1,q_1:[d_1] \to \mathbb{N}} \sum_{p_2,q_2:[d_2] \to \mathbb{N}} \mathbb{E} \left[ \text{tr}(A_N^{p_1(1)} B_N^{q_1(1)} \ldots A_N^{p_1(d_1)} B_N^{q_1(d_1)}) \text{tr}(A_N^{p_2(1)} B_N^{q_2(1)} \ldots A_N^{p_2(d_2)} B_N^{q_2(d_2)}) \right].
\]

We conclude that the covariance of \(\text{tr}(C_{N}^{d_1})\) and \(\text{tr}(C_{N}^{d_2})\) is gotten by summing the difference
\[
\mathbb{E} \left[ \text{tr}(A_N^{p_1(1)} B_N^{q_1(1)} \ldots A_N^{p_1(d_1)} B_N^{q_1(d_1)}) \text{tr}(A_N^{p_2(1)} B_N^{q_2(1)} \ldots A_N^{p_2(d_2)} B_N^{q_2(d_2)}) \right] - \mathbb{E} \text{tr}(A_N^{p_1(1)} B_N^{q_1(1)} \ldots A_N^{p_1(d_1)} B_N^{q_1(d_1)}) \mathbb{E} \text{tr}(A_N^{p_2(1)} B_N^{q_2(1)} \ldots A_N^{p_2(d_2)} B_N^{q_2(d_2)})
\]
over all quadruples of functions \(p_1, q_1: [d_1] \to \mathbb{N}\) and \(p_2, q_2: [d_2] \to \mathbb{N}\) such that \(|p_1| + |q_1| = d_1\) and \(|p_2| + |q_2| = d_2\). Let us fix such a quadruple corresponding to a particular term in this massive sum, and set
\[
\tau_{12}^{(N)} = \mathbb{E} \left[ \text{tr}(A_N^{p_1(1)} B_N^{q_1(1)} \ldots A_N^{p_1(d_1)} B_N^{q_1(d_1)}) \text{tr}(A_N^{p_2(1)} B_N^{q_2(1)} \ldots A_N^{p_2(d_2)} B_N^{q_2(d_2)}) \right],
\]
\[
\tau_1^{(N)} = \mathbb{E} \text{tr}(A_N^{p_1(1)} B_N^{q_1(1)} \ldots A_N^{p_1(d_1)} B_N^{q_1(d_1)})
\]
\[
\tau_2^{(N)} = \mathbb{E} \text{tr}(A_N^{p_2(1)} B_N^{q_2(1)} \ldots A_N^{p_2(d_2)} B_N^{q_2(d_2)}).
\]

We are going to estimate the difference
\[
(37) \quad \tau_{12}^{(N)} - \tau_1^{(N)} \tau_2^{(N)}
\]
in the small \(h_N\), large \(N\) limit. We already know how to deal with the second term; we thus proceed to analyze the first term.

4.2. More computations at the Planck scale. Let us rewrite \(\tau_{12}^{(N)}\) as follows. Put \(d = d_1 + d_2\), and define functions \(p, q: [d] \to \mathbb{N}\) by
\[
p|_{[d_1]} = p_1, \quad p|_{[d_1+1, d]} = p_2
\]
\[
q|_{[d_1]} = q_1, \quad q|_{[d_1+1, d]} = q_2.
\]
We then have
\[
\tau_{12}^{(N)} = \mathbb{E} \left[ \text{tr}(A_N^{p(1)} B_N^{q(1)} \ldots A_N^{p(d_1)} B_N^{q(d_1)}) \text{tr}(A_N^{p(d_1+1)} B_N^{q(d_1+1)} \ldots A_N^{p(d_1+d_2)} B_N^{q(d_1+d_2)}) \right].
\]
It will be convenient to affect the same notational change for the quantities \(\tau_1^{(N)}\) and \(\tau_2^{(N)}\), that is we write
\[
\tau_1^{(N)} = \mathbb{E} \text{tr}(A_N^{p(1)} B_N^{q(1)} \ldots A_N^{p(d_1)} B_N^{q(d_1)})
\]
\[
\tau_2^{(N)} = \mathbb{E} \text{tr}(A_N^{p(d_1+1)} B_N^{q(d_1+1)} \ldots A_N^{p(d_1+d_2)} B_N^{q(d_1+d_2)}).
We now analyze $\tau_{12}^{(N)}$ following the same steps as in Sections 2 and 3.

4.2.1. Unitary invariance.

**Proposition 4.1.** Define a function $f : U(N) \to \mathbb{C}$ by

$$f_N(U) := \mathbb{E} \left[ \text{tr}(U A_N^{(p(1))} U^{-1} B_N^{q(1)} \ldots U A_N^{(p(d_1))} U^{-1} B_N^{q(d_1)}) \text{tr}(U A_N^{(p(d_1 + d_2))} U^{-1} B_N^{q(d_1 + d_2)} \ldots U A_N^{(p(d_1 + d_2))} U^{-1} B_N^{q(d_1 + d_2)}) \right].$$

Then, $f_N$ is constant, being equal to $\tau_{12}^{(N)}$ for all $U \in U(N)$.

As a consequence of this invariance, we have

$$\tau_{12}^{(N)} = \int_{U(N)} f_N(U) dU.$$

We want to use this in exactly the same way as we did in our mean value computation.

Expanding the first trace yields the sum

$$\frac{1}{N} \sum_{r_1 : [4d_1] \to [N]} U_{r_1(1)r_1(2)} \left( A_N^{(p(1))} \right)_{r_1(2)r_1(3)} U^{-1}_{r_1(3)r_1(4)} \left( B_N^{q(1)} \right)_{r_1(4)r_1(5)} \ldots$$

$$\ldots U_{r_1(4d_1 - 3)r_1(4d_1 - 2)} \left( A_N^{(p(d_1))} \right)_{r_1(4d_1 - 2)r_1(4d_1 - 1)} U^{-1}_{r_1(4d_1 - 1)r_1(4d_1)} \left( B_N^{q(d_1)} \right)_{r_1(4d_1)r_1(1)}.$$

Expanding the second trace yields the sum

$$\frac{1}{N} \sum_{r_2 : [4d_2] \to [N]} U_{r_2(1)r_2(2)} \left( A_N^{(p(d_1 + d_2))} \right)_{r_2(2)r_2(3)} U^{-1}_{r_2(3)r_2(4)} \left( B_N^{q(d_1 + d_2)} \right)_{r_2(4)r_2(5)} \ldots$$

$$\ldots U_{r_2(4d_2 - 3)r_2(4d_2 - 2)} \left( A_N^{(p(d_1 + d_2))} \right)_{r_2(4d_2 - 2)r_2(4d_2 - 1)} U^{-1}_{r_2(4d_2 - 1)r_2(4d_2)} \left( B_N^{q(d_1 + d_2)} \right)_{r_2(4d_2)r_2(1)}.$$

For the first trace, let us reparameterize the summation index $r_1 : [4d_1] \to [N]$ by a quadruple of functions $i_1, j_1, i'_1, j'_1 : [d_1] \to [N]$ according to

$$(r_1(1), r_1(2), r_1(3), r_1(4), \ldots, r_1(4d_1 - 3), r_1(4d_1 - 2), r_1(4d_1 - 1), r_1(4d_1))$$

$$= (i_1(1), j_1(1), i'_1(1), j'_1(1), \ldots, i_1(d_1), j_1(d_1), i'_1(d_1), j'_1(d_1)).$$

Then, the above expansion of the first trace becomes

$$\frac{1}{N} \sum_{i_1, j_1, i'_1, j'_1 : [d_1] \to [N]} L_N(i_1, j_1, i'_1, j'_1) \prod_{k=1}^{d_1} \left( A_N^{(p(k))} \right)_{j_1(k)j'_1(k)} \left( B_N^{q(k)} \right)_{i'_1(k)i_1(k)}$$

$$\Rightarrow \frac{1}{N} \sum_{i_1, j_1, i'_1, j'_1 : [d_1] \to [N]} L_N(i_1, j_1, i'_1, j'_1) \prod_{k=1}^{d_1} \left( A_N^{(p(k))} \right)_{j_1(k)j'_1(k)} \prod_{k=1}^{d_1} \left( B_N^{q(k)} \right)_{i'_1(k)i_1(k)}.$$
where
\[ L_N(i_1, j_1, i'_1, j'_1) = \prod_{k=1}^{d_1} U_{i_1(k)j_1(k)} U_{i'_1(k)j'_1(k)}, \]
and \( \gamma_1 \) is the cycle \((1 2 \ldots d_1)\) in \( S_{d_1+d_2} \). Note that we have used the fact that the matrix elements of \( A_N \) commute with those of \( B_N \).

Similarly, if we reparametrize the summation index \( r_2: [4d_2] \to [N] \) by a quadruple of functions \( i_2, j_2, i'_2, j'_2: [d_2] \to [N] \) according to
\[
(\begin{array}{c}
r_2(1), r_2(2), r_2(3), r_2(4), \ldots, r_2(4d_2 - 3), r_2(4d_2 - 2), r_2(4d_2 - 1), r_2(4d_2)
\end{array})
= (i_2(d_1 + 1), j_2(d_1 + 1), j'_2(d_1 + 1), i'_2(d_1 + 1), \ldots, i_2(d), j_2(d), j'_2(d), i'_2(d)),
\]
the expansion of the second trace takes the form
\[
\frac{1}{N} \sum_{i_2, j_2, i'_2, j'_2: [d_1+1, d_1+d_2] \to [N]} L_N(i_2, j_2, i'_2, j'_2) \prod_{k=d_1+1}^{d_1+d_2} \left( A^p_N \right)_{j_2(k)j'_2(k)} \left( B^q_N \right)_{i'_2(k)i_2(k)}
\]
and \( \gamma_2' \) is the cycle \((d_1 + 1 d_1 + 2 \ldots d_1 + d_2)\) in \( S_{d_1+d_2} \).

We now smash the expansions of the two traces together to get the huge compound expansion
\[
\frac{1}{N^2} \sum_{i_1,i_2,i'_1,i'_2,j_1,j'_1,j'_2} L_N(i_1, j_1, i'_1, j'_1) L_N(i_2, j_2, i'_2, j'_2)
\times \prod_{k=1}^{d_1} \left( A^p_N \right)_{j_1(k)j'_1(k)} \prod_{k=d_1+1}^{d_1+d_2} \left( A^p_N \right)_{j_2(k)j'_2(k)} \prod_{k=1}^{d_1} \left( B^q_N \right)_{i'_1(k)i_1(k)} \prod_{k=d_1+1}^{d_1+d_2} \left( B^q_N \right)_{i'_2(k)i_2(k)}
\]
where
\[
L_N(i, j, i', j') = \prod_{k=1}^{d_1+d_2} U_{i(k)j(k)} U_{i'(k)j'(k)}.
\]
Thus, we obtain the following representation of $\tau_{12}^{(N)}$:

$$
\tau_{12}^{(N)} = \frac{1}{N^2} \sum_{i,j,i',j': [d_1+d_2] \to [N]} I_N(i,j,i',j') \mathbb{E}\left[ \prod_{k=1}^{d_1+d_2} \left( A_N^{(k)} \right)^{i(k)j'(k)} \right] \mathbb{E}\left[ \prod_{k=1}^{d_1+d_2} \left( B_N^{(k)} \right)^{i'(k)i\gamma_1\gamma_2'(k)} \right],
$$

where

$$I_N(i,j,i',j') = \int_{U(N)} L_N(i,j,i',j') dU.$$

Plugging (19) into our calculation above eliminates the indices $i',j'$ and produces the formula

$$
\tau_{12}^{(N)} = \frac{1}{N^2} \sum_{(\alpha,\beta) \in \Theta^2_{d_1+d_2}} Wg_N(\alpha,\beta) \mathbb{E}[S_\alpha] \mathbb{E}[S_{\beta^{-1}\gamma_1\gamma_2'}],
$$

where

$$S_\alpha := \sum_{i: [d_1+d_2] \to [N]} \prod_{k=1}^{d_1+d_2} \left( A_N^{(k)} \right)^{i(k)i\alpha(k)}$$

and

$$S_{\beta^{-1}\gamma_1\gamma_2'} := \sum_{i: [d_1+d_2] \to [N]} \prod_{k=1}^{d_1+d_2} \left( B_N^{(k)} \right)^{i(k)i\beta^{-1}\gamma_1\gamma_2'(k)}.$$

Our goal now is to express the operators $S_\alpha$ and $S_{\beta^{-1}\gamma_1\gamma_2'}$ in terms of the operators

$$\text{tr } A_N, \text{tr } A_N^2, \ldots \text{ and } \text{tr } B_N, \text{tr } B_N^2, \ldots.$$

Just like in the mean value computation, we lift the problem to the universal enveloping algebra and use Biasimirs.

4.2.2. Biasimirs again. The operators $S_\alpha$ and $S_{\beta^{-1}\gamma_1\gamma_2'}$ are, up to tensoring with an identity operator, images of Biasimirs in irreducible representations:

$$S_\alpha = \rho_N(C_\alpha^{(p)}) \otimes I_{W_N}$$

$$S_{\beta^{-1}\gamma_1\gamma_2'} = I_{V_N} \otimes \sigma_N(C_{\beta^{-1}\gamma_1\gamma_2'}^{(q)}).$$

Each of the Biasimirs $C_\alpha^{(p)}$ and $C_{\beta^{-1}\gamma_1\gamma_2'}^{(q)}$ has its own classical/quantum decomposition:

$$C_\alpha^{(p)} = \psi_\alpha^{(p)}(C_1, \ldots, C_{|p|}) + h_N Q_\alpha^{(p)}(h_N, N, C_1, \ldots, C_{|p|})$$

$$C_{\beta^{-1}\gamma_1\gamma_2'}^{(q)} = \psi_{\beta^{-1}\gamma_1\gamma_2'}^{(q)}(C_1, \ldots, C_{|q|}) + h_N Q_{\beta^{-1}\gamma_1\gamma_2'}^{(q)}(h_N, N, C_1, \ldots, C_{|q|}).$$

Now we come back to the operators $S_\alpha$ and $S_{\beta^{-1}\gamma_1\gamma_2'}$. First, we have that

$$S_\alpha = \psi_\alpha^{(p)}(\text{Tr } A_N, \ldots, \text{Tr } A_N^{[p]}) + h_N Q_\alpha^{(p)}(h_N, N, \text{Tr } A_N, \ldots, \text{Tr } A_N^{[p]}).$$
Let us rewrite this in terms of normalized traces. Put
\[ \mathbf{P}_\alpha^A := \frac{1}{N_{\text{cyc}}(\alpha)} \varphi_\alpha^{(p)} \left( \text{Tr} A_N, \ldots, \text{Tr} A_N^{[p]} \right), \]
\[ \mathbf{Q}_\alpha^A := \frac{1}{N_{\text{aex}}(\alpha)} Q_\alpha^{(p)} \left( h_N, N, \text{Tr} A_N, \ldots, \text{Tr} A_N^{[p]} \right) \]

\( \mathbf{P}_\alpha^A \) is an explicit polynomial in the operators
\[ \text{tr} A_N, \text{tr} A_N^2, \ldots, \text{tr} A_N^{[p]}, \]
while \( \mathbf{Q}_\alpha^A \) is a polynomial in the numbers \( h_N, N^{-1} \) and the operators
\[ \text{tr} A_N, \text{tr} A_N^2, \ldots, \text{tr} A_N^{[p]} \]

We thus have
\[ S_\alpha = N_{\text{cyc}}(\alpha) \mathbf{P}_\alpha^A + h_N N_{\text{aex}}(\alpha) \mathbf{Q}_\alpha^A. \]

Now we want to apply the expectation \( \mathbb{E} \) to both sides of this identity in \( \mathcal{A}_N \) to get an identity in \( \mathbb{C} \). We set
\[ \mathbf{P}_\alpha^A(N) := \mathbb{E}\mathbf{P}_\alpha^A = \frac{1}{N_{\text{cyc}}(\alpha)} \varphi_\alpha^{(p)} \left( \mathbb{E} \text{Tr} A_N, \ldots, \mathbb{E} \text{Tr} A_N^{[p]} \right), \]
\[ \mathbf{Q}_\alpha^A(N) := \mathbb{E}\mathbf{Q}_\alpha^A = \frac{1}{N_{\text{aex}}(\alpha)} Q_\alpha^{(p)} \left( h_N, N, \mathbb{E} \text{Tr} A_N, \ldots, \mathbb{E} \text{Tr} A_N^{[p]} \right) ; \]
the first of these is a polynomial in the numbers
\[ \mathbb{E} \text{tr} A_N, \mathbb{E} \text{tr} A_N^2, \ldots, \mathbb{E} \text{tr} A_N^{[p]}, \]
while the second is a polynomial in the numbers
\[ h_N, N^{-1}, \mathbb{E} \text{tr} A_N, \mathbb{E} \text{tr} A_N^2, \ldots, \mathbb{E} \text{tr} A_N^{[p]} \]

We conclude that
\[ \mathbb{E}[S_\alpha] = N_{\text{cyc}}(\alpha) \mathbf{P}_\alpha^A(N) + h_N N_{\text{aex}}(\alpha) \mathbf{Q}_\alpha^A(N). \]

Second, we have that
\[ S_{\beta-1\gamma_2^{\beta}} = \varphi_{\beta-1\gamma_2^{\beta}}^{(q)} \left( \text{Tr} B_N, \ldots, \text{Tr} B_N^{[q]} \right) + h_N Q_{\beta-1\gamma_2^{\beta}}^{(q)} \left( h_N, N, \text{Tr} B_N, \ldots, \text{Tr} B_N^{[q]} \right). \]

Once again, let us rewrite this in terms of normalized traces. Put
\[ \mathbf{P}_{\beta-1\gamma_2^{\beta}}^B := \frac{1}{N_{\text{cyc}}(\beta-1\gamma_2^{\beta})} \varphi_{\beta-1\gamma_2^{\beta}}^{(q)} \left( \text{Tr} B_N, \ldots, \text{Tr} B_N^{[q]} \right), \]
\[ \mathbf{Q}_{\beta-1\gamma_2^{\beta}}^B := \frac{1}{N_{\text{aex}}(\beta-1\gamma_2^{\beta})} Q_{\beta-1\gamma_2^{\beta}}^{(q)} \left( h_N, N, \text{Tr} B_N, \ldots, \text{Tr} B_N^{[q]} \right) . \]

\( \mathbf{P}_{\beta-1\gamma_2^{\beta}}^B \) is an explicit polynomial in the operators
\[ \text{tr} B_N, \text{tr} B_N^2, \ldots, \text{tr} B_N^{[q]}, \]
while \( \overline{Q}_{\beta^{-1}\gamma\gamma_2} \) is a polynomial in the numbers \( \hbar N, N^{-1} \) and the operators

\[
\text{tr} B_N, \text{tr} B_N^2, \ldots, \text{tr} B_N^{|q|}.
\]

We thus have

\[
S_{\beta^{-1}\gamma\gamma_2} = N^{\text{cyc}(\beta^{-1}\gamma\gamma_2)} \overline{P}_{\beta^{-1}\gamma\gamma_2}^B + \hbar N N^{\text{aex}(\beta^{-1}\gamma\gamma_2')} \overline{Q}_{\beta^{-1}\gamma\gamma_2}^B.
\]

We apply \( \mathbb{E} \) to both sides of this identity in \( \mathcal{A}_N \) to get an identity in \( \mathbb{C} \). As above, we declare

\[
\overline{P}_{\beta^{-1}\gamma\gamma_2}(N) := \mathbb{E} \overline{P}_{\beta^{-1}\gamma\gamma_2}^B = \frac{1}{N^{\text{cyc}(\beta^{-1}\gamma\gamma_2)}} \rho^{(q)}_{\beta^{-1}\gamma\gamma_2} \left( \mathbb{E} \text{tr} B_N, \ldots, \mathbb{E} \text{tr} B_N^{|q|} \right),
\]

\[
\overline{Q}_{\beta^{-1}\gamma\gamma_2}(N) := \mathbb{E} \overline{Q}_{\beta^{-1}\gamma\gamma_2}^B = \frac{1}{N^{\text{aex}(\beta^{-1}\gamma\gamma_2)}} \mathcal{Q}^{(q)}_{\beta^{-1}\gamma\gamma_2} \left( \hbar N, N, \mathbb{E} \text{tr} B_N, \ldots, \mathbb{E} \text{tr} B_N^{|q|} \right).
\]

The first of these is a polynomial in the numbers

\[
\mathbb{E} \text{tr} B_N, \mathbb{E} \text{tr} B_N^2, \ldots, \mathbb{E} \text{tr} B_N^{|q|},
\]

while the second is a polynomial in the numbers

\[
\hbar N, N^{-1}, \mathbb{E} \text{tr} B_N, \mathbb{E} \text{tr} B_N^2, \ldots, \mathbb{E} \text{tr} B_N^{|q|}.
\]

We conclude that

\[
\mathbb{E}[S_{\beta^{-1}\gamma\gamma_2}] = N^{\text{cyc}(\beta^{-1}\gamma\gamma_2)} \overline{P}_{\beta^{-1}\gamma\gamma_2}^B(N) + \hbar N N^{\text{aex}(\beta^{-1}\gamma\gamma_2')} \overline{Q}_{\beta^{-1}\gamma\gamma_2}^B(N).
\]

4.2.3. Classical/Quantum Decomposition of \( \tau_{12}^{(N)} \). Putting these two calculations together, we compute the \((\alpha, \beta)\) term of \( \tau_{12}^{(N)} \) as

\[
W_{\mathcal{G}N}(\alpha, \beta) \mathbb{E}[S_{\alpha}] \mathbb{E}[S_{\beta^{-1}\gamma\gamma_2}] = \mathbb{E}[S_{\alpha}](N) + \hbar N N^{\text{aex}(\alpha)} \overline{Q}_{\alpha}^A(N) \times \left( N^{\text{cyc}(\beta^{-1}\gamma\gamma_2')} \overline{P}_{\beta^{-1}\gamma\gamma_2}^B(N) + \hbar N N^{\text{aex}(\beta^{-1}\gamma\gamma_2')} \overline{Q}_{\beta^{-1}\gamma\gamma_2}^B(N) \right).
\]

Expanding the brackets and summing \((\alpha, \beta)\) over \( \mathcal{G}_d^2 \), we arrive at the classical/quantum decomposition of the mixed moment \( \tau_N \).

**Theorem 4.2.** We have

\[
\tau_{12}^{(N)} = \text{Classical}_{12}^{(N)} + \hbar N \text{Quantum}_{12}^{(N)},
\]

where

\[
\text{Classical}_{12}^{(N)} = \frac{1}{N^2} \sum_{(\alpha, \beta) \in \mathcal{G}_d^2} N^{\text{cyc}(\alpha)+\text{cyc}(\beta^{-1}\gamma\gamma_2)} W_{\mathcal{G}N}(\alpha, \beta) \overline{P}_{\alpha}^A(N) \overline{P}_{\beta^{-1}\gamma\gamma_2}^B(N).
\]
and

\[
\text{Quantum}_{12}^{(N)} = \frac{1}{N^2} \sum_{(\alpha, \beta) \in S_d^2} \left( N^{\text{cyc}(\alpha) + \text{cyc}(\beta^{-1} \gamma_1 \gamma'_2)} \operatorname{Wg}_N(\alpha, \beta) \mathbf{P}_\alpha \mathbf{Q}_\beta - N^{\text{aex}(\alpha) + \text{cyc}(\beta^{-1} \gamma_1 \gamma'_2)} \operatorname{Wg}_N(\alpha, \beta) \mathbf{P}_\alpha \mathbf{Q}_\beta - N^{\text{aex}(\alpha) + \text{aex}(\beta^{-1} \gamma_1 \gamma'_2)} \operatorname{Wg}_N(\alpha, \beta) \mathbf{P}_\alpha \mathbf{Q}_\beta \right).
\]

4.3. More asymptotics. We now apply the above exact computations to obtain the semiclassical asymptotics of \( \tau_{12}^{(N)} \).

4.3.1. Classical asymptotics.

**Theorem 4.3.** For each \( N \geq d \), the classical part of \( \tau_{12}^{(N)} \) admits, for each \( N \geq d \), an absolutely convergent series expansion of the form

\[
\text{Classical}_{12}^{(N)} = \sum_{k=0}^{\infty} e_k^{(12)}(N) N^{-2k},
\]

the coefficients of which are given by

\[
e_k^{(12)}(N) := \sum_{(g, h) \in \mathbb{N}^2} \sum_{(\alpha, \beta) \in S_d^2} (-1)^{\lvert \alpha^{-1} \beta \rvert} \operatorname{Wg}_g(\alpha, \beta) \mathbf{P}_\alpha \mathbf{Q}_\beta.
\]

**Proof.** We have

\[
N^{\text{cyc}(\alpha) + \text{cyc}(\beta^{-1} \gamma_1 \gamma'_2)} \operatorname{Wg}_N(\alpha, \beta) = N^{\text{cyc}(\alpha) + \text{cyc}(\beta^{-1} \gamma_1 \gamma'_2) - 2(\lvert \alpha^{-1} \beta \rvert - \lvert \alpha^{-1} \beta^{-1} \gamma_1 \gamma'_2 \rvert)} \sum_{g=0}^{\infty} \frac{\operatorname{Wg}_g(\alpha, \beta)}{N^{2g}}
\]

\[
= N^{-\lvert \alpha^{-1} \beta \rvert - \lvert \beta^{-1} \gamma_1 \gamma'_2 \rvert} \sum_{g=0}^{\infty} \frac{\operatorname{Wg}_g(\alpha, \beta)}{N^{2g}}
\]

\[
= (-1)^{\lvert \alpha^{-1} \beta \rvert} N^{-2 \text{genus}(\alpha, \alpha^{-1} \beta, \beta^{-1} \gamma_1 \gamma'_2)} \sum_{g=0}^{\infty} \frac{\operatorname{Wg}_g(\alpha, \beta)}{N^{2g}}.
\]

\[\square\]

4.3.2. Quantum asymptotics.

**Lemma 4.4.** For any \((\alpha, \beta) \in S_d^2\), we have

\[
defect(\alpha) + defect(\beta^{-1} \gamma_1 \gamma'_2) \leq 2 \text{genus}(\alpha, \alpha^{-1} \beta, \beta^{-1} \gamma_1 \gamma'_2).
\]

**Theorem 4.5.** For any fixed \( h_N \), the quantum part \( \text{Quantum}_{12}^{(N)} \) of \( \tau_{12}^{(N)} \) is \( O(1) \) as \( N \to \infty \).
Proof. The quantum part of $\tau_{12}^{(N)}$ may be written as

$$\text{Quantum}_N = \sum_{(\alpha, \beta) \in \mathbb{S}_d^2} N^{\text{cyc}(\alpha)+\text{cyc}(\beta^{-1}\gamma_1\gamma_2')} - 2 \ W_{g,N}(\alpha, \beta) R_{(\alpha, \beta)}(N)$$

where

$$R_{(\alpha, \beta)}(N) = N^\text{defect}(\beta^{-1}\gamma_1\gamma_2') P_{\alpha}^A(N) Q_{\beta^{-1}\gamma_1\gamma_2'}^B(N) + N^\text{defect}(\alpha) Q_{\alpha}^A(N) P_{\beta^{-1}\gamma_1\gamma_2'}^B(N) + \hbar N^\text{defect}(\alpha)+\text{defect}(\beta^{-1}\gamma_1\gamma_2')(N) Q_{\alpha}^A(N) Q_{\beta^{-1}\gamma_1\gamma_2'}^B(N).$$

We will show that each term in the sum is $O(1)$.

By the first part of Lemma 3.1, nonnegativity of the defect statistic, we have

$$R_{(\alpha, \beta)}(N) = O\left(N^\text{defect}(\alpha)+\text{defect}(\beta^{-1}\gamma_1\gamma_2')\right)$$

for each $(\alpha, \beta) \in \mathbb{S}_d^2$. Moreover, from the classical asymptotics calculation above, we know that

$$N^{\text{cyc}(\alpha)+\text{cyc}(\beta^{-1}\gamma_1\gamma_2')} - 2 \ W_{g,N}(\alpha, \beta) = O(N^{-2\text{genus}(\alpha, \alpha^{-1}\beta, \beta^{-1}\gamma_1\gamma_2')}).$$

Thus, the order of the $(\alpha, \beta)$ term in the sum is

$$O\left(N^\text{defect}(\alpha)+\text{defect}(\beta^{-1}\gamma_1\gamma_2')-2\text{genus}(\alpha, \alpha^{-1}\beta, \beta^{-1}\gamma_1\gamma_2')\right).$$

By Lemma 4.4,

$$\text{defect}(\alpha) + \text{defect}(\beta^{-1}\gamma_1\gamma_2') - 2 \text{genus}(\alpha, \alpha^{-1}\beta, \beta^{-1}\gamma_1\gamma_2')$$

is nonpositive. \qed

4.3.3. Semiclassical asymptotics.

Corollary 4.6. For any sequence $\hbar_N$, we have

$$\tau_{12}^{(N)} = \text{Classical}_{12}^{(N)} + O(\hbar_N).$$

In particular, if $\hbar_N = o(N^{-2l})$, then

$$\tau_{12}^{(N)} = \sum_{k=0}^l \frac{e_k^{(12)}(N)}{N^{2k}} + o\left(\frac{1}{N^{2l}}\right).$$
REFERENCES


APPENDIX A. BPP Matrices and Geometric Quantization

As indicated above, BPP matrices quantize independent unitarily invariant random Hermitian matrices with deterministic eigenvalues. This statement falls under the broad umbrella of geometric quantization, in the sense of Kirillov and Kostant, see e.g. [Kir04]. In this section, we give a self-contained, physically motivated treatment of this quantization, specific to our setting. The Reader who is not interested in physical arguments may skip this section entirely.

A.1. Toy example. We begin by considering a toy example: a physical system consisting of a single stationary particle with an angular momentum. For an alternative (but related) exposition of this example see the work of Kuperberg [Kup02]. We will use the corresponding symmetry group $\text{Spin}(3) \cong \text{SU}(2)$ as a starting point for exploration of the unitary group $\text{U}(N)$ and related algebraic and probabilistic objects.

The traditional way to view the angular momentum in Newtonian mechanics is to regard it as a vector $\vec{J} = (J_x, J_y, J_z) \in \mathbb{R}^3$. However, for our purposes it will be more convenient to view the angular momentum as a functional on the Lie algebra of the special orthogonal group $\text{SO}(3)$, that is as an element of $\left(\mathfrak{so}(3)\right)^*$. This functional $J$ is defined as follows. For a given $x \in \mathfrak{so}(3)$ we denote by $J(x)$ Noether’s invariant corresponding to the one-dimensional Lie group $\mathbb{R} \ni t \mapsto e^{tx} \in \text{SO}(3)$ of rotations. Since the map $x \mapsto J(x)$ is linear, it defines an element of the dual space.

From a conceptual point of view, regarding angular momentum as an element of $\left(\mathfrak{so}(3)\right)^*$ is advantageous; for example it scales nicely to other choices of the dimension of the physical space than 3. Unfortunately, the mathematical vocabulary concerning this dual space is rather limited, and
hence it will be convenient to have a more concrete alternative available. For this reason in the following we shall describe the dual of $\mathfrak{so}(3) \cong \mathfrak{su}(2)$ in more detail.

A.2. The dual space. In greater generality, we are interested in the dual of the Lie algebra $\mathfrak{su}(N)$ of traceless antihermitian matrices, as well as the dual of the Lie algebra $\mathfrak{u}(N)$ of general antihermitian matrices.

Each of these Lie algebras can be equipped with the symmetric, non-degenerate, bilinear form

\begin{equation}
\langle x, y \rangle = \text{Tr} \, x^T y.
\end{equation}

In this way $(\mathfrak{su}(N))^* \cong \mathfrak{su}(N)$ and $(\mathfrak{u}(N))^* \cong \mathfrak{u}(N)$. Thanks to these isomorphisms, it makes sense to speak about the eigenvalues of elements of the dual spaces $(\mathfrak{su}(N))^*$ and $(\mathfrak{u}(N))^*$.

In the latter case, this isomorphism takes the following more concrete form. Since the complexification $\mathfrak{u}(N) \otimes \mathbb{C} = \mathfrak{gl}(N) = \text{Mat}_N(\mathbb{C})$ has a matrix structure, it follows that $\mathfrak{u}(N)^* \otimes \mathbb{C} \cong \mathfrak{u}(N) \otimes \mathbb{C} = \text{Mat}_N(\mathbb{C})$ can be also identified with matrices. More specifically, a functional $x \in \mathfrak{u}(N)^* \otimes \mathbb{C}$ corresponds to the matrix

\begin{equation}
\begin{bmatrix}
x(e_{11}) & \cdots & x(e_{N1}) \\
\vdots & \ddots & \vdots \\
x(e_{1N}) & \cdots & x(e_{NN})
\end{bmatrix} = \sum_{k,l} x(e_{kl}) \, e_{kl} \in \text{Mat}_N(\mathbb{C}),
\end{equation}

where $e_{kl} \in \text{Mat}_N(\mathbb{C}) = \mathfrak{u}(N) \otimes \mathbb{C}$ are the standard matrix units. Indeed, the above matrix defines via (39) a functional which on a matrix unit $e_{ij}$ takes the same value as the functional $x$.

Note the subtlety in the formulation of (40): since $e_{kl}$ is not an antihermitian matrix, for $x \in \mathfrak{u}(N)^*$ the quantity $x(e_{kl})$ might be not well defined. Nevertheless, $x(e_{kl})$ may be defined thanks to the observation that $e_{kl} \in \text{Mat}_N(\mathbb{C}) = \mathfrak{u}(N) \otimes \mathbb{C}$ belongs to the complexification of antihermitian matrices, thus we may extend the domain of $x$ by linearity as follows:

\[
x(e_{kl}) = x \left( \frac{e_{kl} + e_{lk}}{2i} + \frac{e_{kl} - e_{lk}}{2} \right) = i x \left( \frac{e_{kl} + e_{lk}}{2i} \right) + x \left( \frac{e_{kl} - e_{lk}}{2} \right).
\]

A.3. Back to the angular momentum. Suppose that for some physical Newtonian system its angular momentum — viewed as a vector $\vec{J} \in \mathbb{R}^3$ — is random, with the uniform distribution on the sphere with radius $|J|$. One can show that this corresponds to $J$ being a random element of the dual space $(\mathfrak{su}(2))^*$, uniformly random on the manifold of antihermitian matrices with specified eigenvalues $\pm i \, |J|$.
In other words, under the isomorphism from Appendix A.2 the distribution of the angular momentum coincides with the distribution of the random matrix

\[ U \begin{bmatrix} i \langle J \rangle \\ -i \langle J \rangle \end{bmatrix} U^{-1}, \]

where \( U \in SU(2) \) is a random matrix from the special unitary group, distributed according to the Haar measure. We now describe a quantum analogue of this probability distribution.

A.4. Angular momentum in quantum mechanics. We consider the following quantum analogue of the Newtonian system considered above: a quantum particle with fixed spin \( j \hbar \), where \( j \in \{0, 1/2, 1, 3/2, \ldots \} \) and \( \hbar \) denotes the Planck constant. Such a particle is described by a Hilbert space \( V \), this space being the appropriate unitary representation \( \alpha : \text{Spin}(3) \to \text{GL}(V) \). The Lie group \( \text{Spin}(3) \cong SU(2) \) is the universal cover of the group \( \text{SO}(3) \) describing rotations of the physical space. To be more specific, \( \alpha \) is the irreducible representation of the Lie group \( SU(2) \) with the dimension \( 2j + 1 \in \{1, 2, \ldots \} \).

In order to sustain the concordance with the Newtonian situation discussed above, the angular momentum should be a functional

\[ J : \mathfrak{so}(3) \to \text{End} V \]

which to an element of the Lie algebra \( x \in \mathfrak{so}(3) \) associates the infinitesimal \textit{hermitian} generator of the action of the one-parameter Lie group \( \mathbb{R} \ni t \mapsto e^{tx} \in \text{Spin}(3) \) on its representation \( V \), i.e.

\[ \alpha(e^{tx}) = e^{-it \frac{J(x)}{\hbar}}. \]

The choice of normalization on the right hand side comes from the notations used in quantum mechanics. Clearly, this means that (up to a scalar multiple) the angular momentum

\[ -i \frac{J}{\hbar} = \alpha : \mathfrak{so}(3) \to \text{End} V \]

is a representation of the Lie algebra \( \mathfrak{so}(3) = \mathfrak{su}(2) \). If \( \text{End} V \) is viewed as an algebra of noncommutative random variables,

\[ -i \frac{J}{\hbar} = \alpha \in (\mathfrak{so}(3))^* \otimes \text{End} V \]

becomes a \textit{quantum random element of the dual space} \( (\mathfrak{so}(3))^* = (\mathfrak{su}(2))^* \).

Just as before we assume that we have no further information about the particle; in other words, the quantum system is in the maximally mixed state and thus the algebra \( \text{End} V \) of noncommutative random variables is
equipped with the state $\text{tr}_V$. Just as before, it is convenient to have a concrete matrix representation from Appendix A.2 for the elements of the dual space $(\text{su}(2))^\ast$. We shall discuss this concrete representation now.

A.5. **The dual space.** Consider a slightly more general situation in which $\alpha: \mathfrak{u}(N) \rightarrow \text{End} \ V$ is a representation of the Lie algebra $\mathfrak{u}(N)$.

Equation (40) shows that $-i\frac{J}{\hbar} = \alpha$ can be identified with the matrix

$$
-\frac{i J}{\hbar} = \alpha = \begin{bmatrix}
\alpha(e_{11}) & \cdots & \alpha(e_{N1}) \\
\vdots & \ddots & \vdots \\
\alpha(e_{1N}) & \cdots & \alpha(e_{NN})
\end{bmatrix},
$$

(42)

A.6. **Conclusion.** The above considerations show that from a physicist’s point of view, for $N = 2$ the $2 \times 2$ matrix (42) is a natural quantization of the random matrix (41) which describes the angular momentum in Newtonian mechanics.

It is time to detach from the physical toy example related to the group $\text{Spin}(3) \cong \text{SU}(2)$ and consider the general situation treated in this article. The classical object which we considered in this section was a random element of $(\mathfrak{u}(N))^\ast$ (or, a random antihermitian matrix), sampled uniformly from the elements with specified spectrum. Its quantization is a BPP matrix: a quantum random element of $(\mathfrak{u}(N))^\ast$ which corresponds to a specified irreducible representation of $\text{U}(N)$.

A.7. **Choice of the matrix structure on $\mathfrak{u}(N)^\ast$.** Unlike in the case of the Lie algebra $\mathfrak{u}(N)$, there is no canonical choice of matrix structure on the dual $\mathfrak{u}(N)^\ast$. In Appendix A.2 this structure was chosen based on the bilinear form $\langle A, B \rangle = \text{Tr} \ A^\ast B$. One can argue however, that the bilinear form $\langle A, B \rangle = \frac{1}{2} \text{Tr} \ AB$ would be equally natural. With respect to this new convention, the representation $\alpha$ viewed as a matrix becomes

$$
\begin{bmatrix}
\alpha(e_{11}) & \cdots & \alpha(e_{1N}) \\
\vdots & \ddots & \vdots \\
\alpha(e_{N1}) & \cdots & \alpha(e_{NN})
\end{bmatrix},
$$

(43)

which is not a BPP matrix. The matrices (42) and (43) differ only by transposition with respect to the first factor of the tensor product $\text{Mat}_N(\mathbb{C}) \otimes \text{End} \ V$, an operation known as partial transposition. The minor advantage of the notation (42) is that it coincides with the notation of Želobenko [Žel73] who calculated the spectral measure of BPP matrices.

There are, however, no serious advantages of one notation over the other, since the calculation of the spectral measure of (43) can be done by the analogous methods to those of Želobenko [Žel73]. The only difference is that
instead of considering the tensor product with the canonical representation, one should consider the tensor product with the contragradient one.

DEPARTMENT OF MATHEMATICS, GRADUATE SCHOOL OF SCIENCE, KYOTO UNIVERSITY, KYOTO 606-8502, JAPAN
CNRS, FRANCE
E-mail address: collins@math.kyoto-u.ac.jp

DEPARTMENT OF MATHEMATICS, UC SAN DIEGO, 9500 GILMAN DRIVE, LA JOLLA CA 92093-0112, USA
E-mail address: jinovak@ucsd.edu

INSTITUTE OF MATHEMATICS, POLISH ACADEMY OF SCIENCES, UL. ŚNIADECKICH 8, 00-956 WARSZAWA, POLAND
E-mail address: psniady@impan.pl