

Heat Kernel Empirical Laws on \mathbb{U}_N and \mathbb{GL}_N

Todd Kemp*
Department of Mathematics
University of California, San Diego
La Jolla, CA 92093-0112
tkemp@math.ucsd.edu

January 23, 2015

Abstract

This paper studies the empirical laws of eigenvalues and singular values for random matrices drawn from the heat kernel measures on the unitary groups \mathbb{U}_N and the general linear groups \mathbb{GL}_N , for $N \in \mathbb{N}$. It establishes the strongest known convergence results for the empirical eigenvalues in the \mathbb{U}_N case, and the first known almost sure convergence results for the eigenvalues and singular values in the \mathbb{GL}_N case. The limit noncommutative distribution associated to the heat kernel measure on \mathbb{GL}_N is identified as the projection of a flow on an infinite dimensional polynomial space. These results are then strengthened from variance estimates to L^p estimates for even integers p .

Contents

1	Introduction	2
1.1	Main Theorems	2
1.2	Discussion	5
2	Background	6
2.1	Heat Kernels on \mathbb{U}_N and \mathbb{GL}_N	6
2.2	The Heat Kernel on \mathbb{U} , Sobolev Spaces, and Gevrey Classes	8
2.3	Functional Calculus and Empirical Laws	10
2.4	Noncommutative Distributions	11
2.5	Free Probability	14
3	Intertwining Operators and Concentration of Measure	16
3.1	The Action of $\Delta_{\mathbb{U}_N}$ and $A_{s,t}^N$ on Trace Polynomials	16
3.2	Concentration of Heat Kernel Measure	20
3.3	Estimates on the Constants $C(s, t, P)$	21
4	Convergence of Empirical Distributions	25
4.1	Empirical Eigenvalues on \mathbb{U}_N	25
4.2	Empirical Noncommutative Distribution on \mathbb{GL}_N	30
4.3	Empirical Eigenvalues on \mathbb{GL}_N	33
4.4	Empirical Singular Values on \mathbb{GL}_N	36

*Supported by NSF CAREER Award DMS-1254807

5	L^p Convergence	38
5.1	Strong Convergence and Noncommutative L^p -norms	38
5.2	Almost Sure L^p Convergence	39

1 Introduction

This paper is concerned with the empirical eigenvalue laws associated to heat kernels on the unitary groups and the general linear groups. Let \mathbb{M}_N denote $N \times N$ complex matrices, let $\mathbb{U}_N = \{U \in \mathbb{M}_N : UU^* = I_N\}$ be the unitary group, and $\mathbb{GL}_N \subset \mathbb{M}_N$ the general linear group of invertible $N \times N$ complex matrices. The unitary group \mathbb{U}_N is a real Lie group, and \mathbb{GL}_N is its complexification. These Lie groups possess natural Laplace operators $\Delta_{\mathbb{U}_N}$ and $\Delta_{\mathbb{GL}_N}$; cf. Definition 2.2 below. The *heat kernel* ρ_t^N is the fundamental solution to the heat equation $\partial_t \rho_t^N = \frac{1}{2} \Delta_{\mathbb{U}_N} \rho_t^N$ on \mathbb{U}_N ; similarly the heat kernel μ_t^N is the fundamental solution to the heat equation $\partial_t \mu_t^N = \frac{1}{2} \Delta_{\mathbb{GL}_N} \mu_t^N$ on \mathbb{GL}_N . They are strictly positive smooth probability densities with respect to the (right) Haar measures, and so we identify each density with its measure when convenient. In fact, we will consider a two-parameter heat kernel $\mu_{s,t}^N$ on \mathbb{GL}_N , where $s, t > 0$ and $s > t/2$, which interpolates between ρ_s^N when $t = 0$ and $\mu_{t/2}^N$ when $s = t$; cf. Definition 2.2.

To fix notation, for $N \in \mathbb{N}$ and $s, t > 0$ with $s > t/2$, we set

$$U_t^N \text{ is a random unitary matrix with joint law of entries } \rho_t^N, \text{ and}$$

$$Z_{s,t}^N \text{ is a random invertible matrix with joint law of entries } \mu_{s,t}^N.$$

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space from which all the random matrices $\{U_t^N, Z_{s,t}^N; N \in \mathbb{N}, s, t > 0, s > t/2\}$ are sampled. As usual, for $F \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, denote $\mathbb{E}(F) = \int_{\Omega} F d\mathbb{P}$.

1.1 Main Theorems

We are interested in the **empirical eigenvalue laws** of these matrices. For $Z \in \mathbb{M}_N$ denote by $\Lambda(Z)$ the unordered list of eigenvalues of Z , counted with multiplicities. The empirical eigenvalue laws are the following random discrete measures on \mathbb{C} :

$$\tilde{\nu}_t^N = \frac{1}{N} \sum_{\lambda \in \Lambda(U_t^N)} \delta_{\lambda} \quad \text{and} \quad \tilde{\phi}_{s,t}^N = \frac{1}{N} \sum_{\lambda \in \Lambda(Z_{s,t}^N)} \delta_{\lambda}. \tag{1.1}$$

To describe the limit behavior of these random measures, we introduce the following one-parameter family of probability measures.

Theorem / Definition 1.1. *For each $t \in \mathbb{R}$, there exists a unique probability measure ν_t on $\mathbb{C}^* = \mathbb{C} \setminus \{0\}$ with the following properties. For $t > 0$, ν_t is supported in the unit circle \mathbb{U} ; for $t < 0$, ν_t is supported in $\mathbb{R}_+ = (0, \infty)$; and $\nu_0 = \delta_1$. In all cases, ν_t is determined by its moments: $\nu_0(t) \equiv 1$ and, for $n \in \mathbb{Z} \setminus \{0\}$,*

$$\nu_n(t) \equiv \int_{\mathbb{C}^*} u^n \nu_t(du) = e^{-\frac{|n|}{2}t} \sum_{k=0}^{|n|-1} \frac{(-t)^k}{k!} |n|^{k-1} \binom{|n|}{k+1}. \tag{1.2}$$

For all $t \neq 0$, ν_t possesses a continuous density ϱ_t with connected, compact support; ϱ_t is strictly positive in a neighborhood of 1 (in \mathbb{U} for $t > 0$, in \mathbb{R}_+ for $t < 0$), and real analytic on the set where it is positive; cf. [10] for the $t > 0$ case, and [40] for the $t < 0$ case. Section 2.5 has further discussion of the measures ν_t and their relevance to free probability theory.

For $t > 0$, ν_t was identified as $\lim_{N \rightarrow \infty} \mathbb{E}(\tilde{\nu}_t^N)$ in [9], and independently in [32]. In the latter case, the convergence was proved to be weakly almost sure for polynomial test functions. Our first main theorem weakens the regularity conditions requires for the almost sure convergence.

Theorem 1.2. *For $t > 0$ and $N \in \mathbb{N}$, let $\tilde{\nu}_t^N$ and ν_t be the measures in (1.1) and Definition 1.1. Then $\tilde{\nu}_t^N$ converges to ν_t weakly in probability:*

$$\mathbb{P} \left(\left| \int_{\mathbb{U}} f d\tilde{\nu}_t^N - \int_{\mathbb{U}} f d\nu_t \right| > \epsilon \right) \rightarrow 0, \quad \epsilon > 0, \quad f \in C(\mathbb{U}). \quad (1.3)$$

Moreover, if $1 < p < \frac{3}{2}$ and f is in the Sobolev space $H_p(\mathbb{U})$ (cf. Definition 2.7), then the convergence is almost sure, and

$$\text{Var} \left(\int_{\mathbb{U}} f d\tilde{\nu}_t^N \right) \leq \frac{C(t, p)}{N^{2p-1}} \|f\|_{H_p(\mathbb{U})}^2 \quad (1.4)$$

for some constant $C(t, p) < \infty$ that depends continuously on t and p . Finally, if $f \in H_p(\mathbb{U})$ with $p > \frac{3}{2}$, then f is Lipschitz on \mathbb{U} , and

$$\text{Var} \left(\int_{\mathbb{U}} f d\tilde{\nu}_t^N \right) \leq \frac{2t}{N^2} \|f\|_{\text{Lip}(\mathbb{U})}. \quad (1.5)$$

See (4.2) for the definition of the Lipschitz norm on \mathbb{U} .

By taking $f \in C(\mathbb{U})$ close to the indicator function of any given arc, (1.3) and (1.4) show that the density of eigenvalues of U_t^N converges, in a fairly strong sense, to ν_t . We prove Theorem 1.2 (on page 29) incorporating some estimates from [30] with a Fourier cut-off argument. Note: in [30], the (Gaussian) fluctuations of the empirical integrals $\int_{\mathbb{U}} f d\tilde{\nu}_t^N$ are computed: they are on the scale of the Sobolev space $H_{1/2}(\mathbb{U})$ as $t \rightarrow \infty$. We conjecture that the $O(1/N^{2p-1})$ in (1.4) can be improved to $O(1/N^2)$, and that therefore the a.s. convergence holds for $f \in H^p(\mathbb{U})$ for any $p > \frac{1}{2}$. At the end of Section 4.2, we discuss how tighter bounds on the constants from Section 3.3 would lead to this minimal-regularity conjecture.

As most matrices in \mathbb{GL}_N are not normal, there are limits to what we can say about the empirical law $\tilde{\phi}_{s,t}^N$. The following is a natural analogue of Theorem 1.2 in this context.

Theorem 1.3. *Let $s, t > 0$ with $s > t/2$ and $N \in \mathbb{N}$, and let $\tilde{\phi}_{s,t}^N$ be the empirical eigenvalue law of (1.1). Let $f(z) = \sum_n a_n z^n$ be in the ultra-analytic Gevrey class $G_\sigma(\mathbb{C}^*)$ (meaning $\|f\|_{G_\sigma}^2 \equiv \sum_n |a_n|^2 e^{2\sigma n^2} < \infty$; cf. Definition 2.8) for some $\sigma > s$. Then*

$$\left| \mathbb{E} \left(\int_{\mathbb{C}^*} f d\tilde{\phi}_{s,t}^N \right) - \int_{\mathbb{C}^*} f d\nu_{s-t} \right| \leq \frac{C_1(s)}{N^2} \|f\|_{G_\sigma}, \quad \text{and} \quad (1.6)$$

$$\text{Var} \left(\int_{\mathbb{C}^*} f d\tilde{\phi}_{s,t}^N \right) \leq \frac{C_2(s)}{N^2} \|f\|_{G_\sigma}^2, \quad (1.7)$$

for some constants $C_1(s), C_2(s) < \infty$ that depend continuously on s (and are independent of t).

To be clear, the class $G_\sigma(\mathbb{C}^*)$ of test functions is not rich enough to approximate indicator functions of disks, and so Theorem 1.3 does not necessarily imply that the density of eigenvalues converges to ν_{s-t} . The proof of Theorem 1.5 is on page 35.

We also consider the convergence of the density of **singular values** of $Z_{s,t}^N$; i.e. the square roots of the eigenvalues of the positive-definite matrix $Z_{s,t}^N (Z_{s,t}^N)^*$.

Definition 1.4. *Let $\mathbb{M}_N^{>0}$ denote the set of positive definite $N \times N$ matrices. The map $\Phi: \mathbb{GL}_N \rightarrow \mathbb{M}_N^{>0}$ given by $\Phi(Z) = ZZ^*$ is a smooth surjection. Let $\tilde{\eta}_{s,t}^N$ be the empirical eigenvalue law of $\Phi(Z_{s,t}^N)$.*

Theorem 1.5. For $s, t > 0$ with $s > t/2$ and $N \in \mathbb{N}$, the empirical eigenvalue law $\tilde{\eta}_{s,t}^N$ of Definition 1.4 converges ultra-analytically almost surely to ν_{-2t} : if f is in the Gevrey class $G_\sigma(\mathbb{C}^*)$ for some $\sigma > 4s$, then

$$\left| \mathbb{E} \left(\int_0^\infty f d\tilde{\eta}_{s,t}^N \right) - \int_0^\infty f d\nu_{-2t} \right| \leq \frac{C_1(4s)}{N^2} \|f\|_{G_\sigma}, \quad \text{and} \quad (1.8)$$

$$\text{Var} \left(\int_0^\infty f d\tilde{\eta}_{s,t}^N \right) \leq \frac{C_2(4s)}{N^2} \|f\|_{G_\sigma}^2, \quad (1.9)$$

where the constants $C_1(\cdot)$ and $C_2(\cdot)$ are the same ones given in Theorem 1.3.

The proof of Theorem 1.5 is on page 38. It is likely that (1.9) holds for much less regular test functions, as in Theorem 1.2.

Remark 1.6. Equation (1.8), in the special case of polynomial test functions, was stated without proof at the end of [9]. Note there is a scaling disagreement: Biane stated that the moments are those of ν_{-t} . This is because the free stochastic process he uses to define the limit Brownian motion on \mathbb{GL}_N is scaled differently from ours: instead of (2.36), his process is defined by $d\Gamma_t = \frac{1}{2} dz_t \Gamma_t$. This amounts to using (a right-invariant version of) $-\frac{1}{2}$ of the generator we use, thus accounting for the discrepancy.

In [9], it was alluded that the computation of (1.8) follows from combinatorial representation-theoretic tools like used earlier in that paper. Our present approach is more geometric. In fact, we give a unified approach to Theorems 1.2, 1.3 and 1.5, which applies to the much more general context of the noncommutative distribution of $Z_{s,t}^N$; cf. Section 2.4.

Theorem 1.7. Let $s, t > 0$ with $s > t/2$, and let $\tilde{\varphi}_{s,t}^N$ denote the empirical noncommutative distribution of $Z_{s,t}^N$; cf. Definition 2.14. There exists a noncommutative distribution $\varphi_{s,t}$ (cf. Definition 2.12) such that $\tilde{\varphi}_{s,t}^N \rightarrow \varphi_{s,t}$ weakly almost surely: for each noncommutative Laurent polynomial $f \in \mathbb{C}\langle A, A^{-1}, A^*, A^{-*} \rangle$,

$$\left| \mathbb{E}[\tilde{\varphi}_{s,t}^N(f)] - \varphi_{s,t}(f) \right| \leq \frac{C_1(s, t, f)}{N^2}, \quad \text{and} \quad (1.10)$$

$$\text{Var}[\tilde{\varphi}_{s,t}^N(f)] \leq \frac{C_2(s, t, f)}{N^2}, \quad (1.11)$$

for some constants $C_1(s, t, f), C_2(s, t, f) < \infty$ that depend continuously on s, t .

Let $\text{tr}(Z) = \frac{1}{N} \text{Tr}(Z)$ denote the normalized trace on \mathbb{M}_N . Theorem 1.7 asserts that all of the random trace moments $\text{tr}((Z_{s,t}^N)^{\varepsilon_1} \cdots (Z_{s,t}^N)^{\varepsilon_n})$ (for $\varepsilon_1, \dots, \varepsilon_n \in \{\pm 1, \pm *\}$) converge almost surely to their means. In fact, our techniques show the stronger claim that all *products* of such trace moments also have $O(1/N^2)$ -variance, hence also describing the fluctuations of these random variables. The proof of Theorem 1.7 is on page 31.

Remark 1.8. Restricting all test functions to (Laurent) polynomials, Theorem 1.2 is the special case $(s, t) \mapsto (t, 0)$ of Theorem 1.7; and Theorems 1.3 and 1.5 are achieved by taking f to depend only on Z in the first case, and only on ZZ^* in the second.

The essential idea behind the above concentration results can be described succinctly in the unitary case as follows. Since the solution $h(t, \cdot)$ to the heat equation $\partial_t h = \frac{1}{2} \Delta_{\mathbb{U}_N} h$ with initial condition $h(0, U) = f(U)$ is given by convolution against the heat kernel (cf. [24]),

$$h(t, U) = \int_{\mathbb{U}_N} f(UV) \rho_t^N(dV), \quad (1.12)$$

evaluating this convolution at the identity shows $h(t, I_N)$ is the integral of f against the heat kernel ρ_t^N . But $h(t, \cdot)$ may also be represented in terms of the *heat semigroup*, $h(t, \cdot) = e^{\frac{t}{2} \Delta_{\mathbb{U}_N}} f$; thus we have

$$\int_{\mathbb{U}_N} f d\rho_t^N = \left(e^{\frac{t}{2} \Delta_{\mathbb{U}_N}} f \right) (I_N). \quad (1.13)$$

In fact, (1.13) determines the measure ρ_t^N when taken over all $f \in C(\mathbb{U}_N)$; we take it as the definition of ρ_t^N in (2.7) below. Now, as explained below in Section 3 following [16, Theorem 1.18], on a sufficiently rich space of functions, $\Delta_{\mathbb{U}_N}$ has a decomposition

$$\Delta_{\mathbb{U}_N} = D_N + \frac{1}{N^2} L_N \quad (1.14)$$

where D_N and L_N are first- and second-order differential operators, both uniformly bounded in N ; they are given explicitly as intertwining operators in Theorem 3.6. In fact, D_N has a limit as $N \rightarrow \infty$, which we can think of as the *generator* of free unitary Brownian motion; cf. [9] and Section 2.5. Hence, in the limit as $N \rightarrow \infty$, the heat operator $e^{\frac{t}{2}\Delta_{\mathbb{U}_N}}$ behaves as the flow of a vector field; i.e. it is an algebra homomorphism, which shows that variances vanish in the limit. The same idea holds in the $\mathbb{G}\mathbb{L}_N$ -case as well, in the much larger context of the “test-functions” (noncommutative polynomials) of noncommutative distributions; cf. Definition 2.15.

These same ideas allow us to prove a stronger form of convergence of these empirical distributions.

Theorem 1.9. *Fix $s, t > 0$ with $s > t/2$. Let (\mathcal{A}, τ) be a noncommutative probability space (Definition 2.10) that contains the almost sure weak limits u_t and $z_{s,t}$ of U_t^N and $Z_{s,t}^N$; cf. Theorem 1.7. Then, for any noncommutative polynomial $f \in \mathbb{C}\langle A, A^* \rangle$, and any even integer $p \geq 2$,*

$$\begin{aligned} \|f(U_t^N, (U_t^N)^*)\|_{L^p(\mathbb{M}_N, \text{tr})} &\rightarrow \|f(u_t, u_t^*)\|_{L^p(\mathcal{A}, \tau)} \text{ a.s. as } N \rightarrow \infty, \text{ and} \\ \|f(Z_{s,t}^N, (Z_{s,t}^N)^*)\|_{L^p(\mathbb{M}_N, \text{tr})} &\rightarrow \|f(z_{s,t}, z_{s,t}^*)\|_{L^p(\mathcal{A}, \tau)} \text{ a.s. as } N \rightarrow \infty. \end{aligned}$$

Section 5 is devoted to Theorem 1.9, where the noncommutative L^p -norms are defined and discussed.

1.2 Discussion

The problems discussed above are natural extensions of now well-known theorems in random matrix theory. Let us be slightly more general for the moment. Let ρ^N be a probability measure on \mathbb{M}_N , and let A_N be a random matrix with ρ^N as its joint law of entries. Denote

$$\tilde{\nu}^N = \frac{1}{N} \sum_{\lambda \in \Lambda(A_N)} \delta_\lambda \quad (1.15)$$

the empirical eigenvalue law of A_N . If the support of ρ^N is contained in the normal matrices $\mathbb{M}_N^{\text{nor}}$, then *empirical integrals* against measurable test functions $f: \mathbb{C} \rightarrow \mathbb{C}$ can be computed by

$$\int_{\mathbb{C}} f d\tilde{\nu}^N = \text{tr} \circ f_N, \quad (1.16)$$

where the function $f_N: \mathbb{M}_N^{\text{nor}} \rightarrow \mathbb{M}_N^{\text{nor}}$ is given by *measurable functional calculus*; cf. Section 2.3 below. In particular, (1.16) will often be used to compute expectations against continuous functions:

$$\mathbb{E} \left(\int_{\mathbb{C}} f d\tilde{\nu}^N \right) = \int_{\mathbb{M}_N^{\text{nor}}} (\text{tr} \circ f_N) d\rho^N, \quad f \in C_c(\mathbb{C}). \quad (1.17)$$

The most well-known example of such a normal (in fact Hermitian) empirical eigenvalue law comes from Wigner’s semicircle law; cf. [37, 38, 39]. In the original Gaussian case, ρ^N is supported on Hermitian matrices, with

$$\rho^N(dX) = c_N e^{-N\text{Tr}(X^2)} dX \quad (1.18)$$

where dX denotes the Lebesgue measure on Hermitian matrices (coordinatized by the real and imaginary parts of the upper-triangular entries), and c_N is a normalization constant. This measure is known as the GUE_N

or *Gaussian Unitary Ensemble*; it is equivalently described by insisting that the upper-triangular entries of the Hermitian random matrix X are i.i.d. normal random variables of variance $1/N$. Wigner proved that, in this case, the empirical eigenvalue law converges weakly in expectation to the *semicircle law* $\zeta(dx) = \frac{1}{2\pi} \sqrt{(4-x^2)_+} dx$. That is to say: Wigner proved that the quantities in (1.17) converge to the relevant integrals against $d\zeta$. It was shown later [2, 3, 4] that this convergence is *weakly almost sure*, in the sense that the random variables $\int f d\nu^N$ converge to their means almost surely.

Remark 1.10. Having realized all requisite random matrices (of all sizes $N \in \mathbb{N}$) over a single probability space $(\Omega, \mathcal{F}, \mathbb{P})$, proving almost sure convergence amounts to showing that the variances tend to 0 summably-fast (by Chebyshev's inequality and the Borel-Cantelli lemma).

Much of the modern theory of random matrices is concerned with generalizations of Wigner's example in one of two ways: either to other measures ρ^N on Hermitian matrices that make the upper-triangular entries i.i.d., or to measures with densities generalizing the form of (1.18), for example by replacing $\text{Tr}(X^2)$ with a different (sufficiently convex) potential. A great deal is understood in both these arenas about the empirical laws and many other statistics of the random eigenvalues; the interested reader should consult [1].

Another well-studied example is the Haar measure $\rho^N = \text{Haar}(\mathbb{U}_N)$ on the the unitary group \mathbb{U}_N . Unitary matrices are normal, and so (1.16) characterizes the empirical eigenvalue laws; in this case, they are known (cf. [14]) to converge weakly almost surely to the uniform probability measure on \mathbb{U} . In both this case and the Wigner ensembles described above, stronger convergence results are known, such as in Theorem 1.9 above.

Remark 1.11. If, instead of \mathbb{U}_N , we take the additive Lie group of Hermitian matrices, the heat kernel is precisely the Gaussian measure (1.18), where N is replaced by N/t on the right-hand-side. The space of Hermitian matrices can be identified as $i\mathfrak{u}_N$, where $\mathfrak{u}_N = \{X \in \mathbb{M}_N: X^* = -X\}$ is the Lie algebra of \mathbb{U}_N ; thus, the GUE_N is the Lie algebra version of the heat kernel on \mathbb{U}_N . As $t \rightarrow \infty$, the heat kernel measure ρ_t^N on \mathbb{U}_N converges to the Haar measure. In this sense, the heat kernel measures considered in the present paper fit into a larger scheme of well-studied random matrix ensembles.

The support of the heat kernel measures $\mu_{s,t}^N$ on \mathbb{GL}_N consists largely of non-normal matrices, and so measurable functional calculus is not available. It is for this reason that our analysis is restricted to holomorphic test functions in this case. Nevertheless, the results presented in Theorems 1.3 – 1.7 are new; in particular, the existence of the noncommutative distribution $\varphi_{s,t}$ in Theorem 1.7 was part of a conjecture posed in [9]. The full conjecture deals with the limit of the stochastic process $t \mapsto Z_{t,t}^N$, the Brownian motion on \mathbb{GL}_N which, for each fixed t , has distribution $\mu_{t/2}^N$. In the present paper, we deal only with a single $t > 0$, with all theorems proved with bounds that are uniform for t in compact intervals. Identifying the limit distribution of the full stochastic process will be addressed in the separate paper [26].

2 Background

In this section, we give concise discussions of the necessary constructs for this paper: heat kernel analysis on the groups \mathbb{U}_N and \mathbb{GL}_N ; regularity of test functions (Sobolev spaces and Gevrey classes); measurable functional calculus on \mathbb{U}_N and holomorphic functional calculus on \mathbb{GL}_N ; and noncommutative probability theory (in particular free probability and free multiplicative convolution). For general reference, readers are directed to the monograph [33] for heat kernel analysis on Lie groups, and the lecture notes [31] for a thorough treatment of noncommutative and free probability.

2.1 Heat Kernels on \mathbb{U}_N and \mathbb{GL}_N

Let $G \subset \mathbb{M}_N$ be a matrix Lie group, with Lie algebra $\text{Lie}(G)$; relevant to this paper are \mathbb{U}_N with $\text{Lie}(\mathbb{U}_N) = \mathfrak{u}_N = \{X \in \mathbb{M}_N: X^* = -X\}$, and \mathbb{GL}_N with $\text{Lie}(\mathbb{GL}_N) = \mathfrak{gl}_N = \mathbb{M}_N$. Note that $\mathfrak{gl}_N = \mathfrak{u}_N \oplus i\mathfrak{u}_N$. Hence, if β_N is a basis for \mathfrak{u}_N as a real vector space, then β_N is also a basis for \mathfrak{gl}_N as a complex vector space.

We will use the following (scaled) Hilbert-Schmidt inner product on \mathfrak{gl}_N :

$$\langle \xi, \zeta \rangle_N \equiv N \text{Tr}(\xi \zeta^*) = N^2 \text{tr}(\xi \zeta^*), \quad \xi, \zeta \in \mathfrak{gl}_N. \quad (2.1)$$

Restricted to \mathfrak{u}_N , this inner product is $\text{Ad}_{\mathbb{U}_N}$ -invariant, and real valued:

$$\langle X, Y \rangle_N = -N \text{Tr}(XY), \quad X, Y \in \mathfrak{u}_N. \quad (2.2)$$

The scaling chosen here is consistent with the scaling in (1.18); as we will see in the following, it is the unique scaling that leads to limit distributions as $N \rightarrow \infty$.

Definition 2.1. Let G be a Lie group and $\xi \in \text{Lie}(G)$. Then the exponential $e^{t\xi}$ is in G for $t \in \mathbb{R}$. The **left-invariant vector field** or **derivative** associated to ξ is the operator ∂_ξ on $C^\infty(G)$ defined by

$$(\partial_\xi f)(g) = \left. \frac{d}{dt} \right|_{t=0} f(ge^{t\xi}). \quad (2.3)$$

Definition 2.2. Let β_N be an orthonormal basis (with respect to (2.2)) for \mathfrak{u}_N . The **Laplace operator** on $C^\infty(\mathbb{U}_N)$ is

$$\Delta_{\mathbb{U}_N} = \sum_{X \in \beta_N} \partial_X^2. \quad (2.4)$$

The **Laplace operator** on $C^\infty(\mathbb{GL}_N)$ is

$$\Delta_{\mathbb{GL}_N} = \sum_{X \in \beta_N} (\partial_X^2 + \partial_{iX}^2). \quad (2.5)$$

More generally, for $s, t \in \mathbb{R}$, define the operators $A_{s,t}^N$ on $C^\infty(\mathbb{GL}_N)$ by

$$A_{s,t}^N = \left(s - \frac{t}{2} \right) \sum_{X \in \beta_N} \partial_X^2 + \frac{t}{2} \sum_{X \in \beta_N} \partial_{iX}^2. \quad (2.6)$$

A routine calculation shows that these definitions do not depend on the particular orthonormal basis used.

Remark 2.3. (1) The operator $\Delta_{\mathbb{U}_N}$ is the Casimir element in the universal enveloping algebra $\mathcal{U}(\mathfrak{u}_N)$. Since the inner product (2.2) is Ad -invariant, $\Delta_{\mathbb{U}_N}$ commutes with the left- and right-actions of \mathbb{U}_N on $C^\infty(\mathbb{U}_N)$; i.e. it is bi-invariant. It is equal to the Laplace-Beltrami operator on \mathbb{U}_N associated to the bi-invariant Riemannian metric induced by (2.2).

(2) The non-semisimple Lie group \mathbb{GL}_N possesses no Ad -invariant inner product. Eq. (2.5) matches the Laplace-Beltrami operator on \mathbb{GL}_N associated to the left-invariant Riemannian metric induced by (2.1).

(3) The interpolating operator $A_{s,t}^N$ is negative-definite when $s, t > 0$ and $s > t/2$; in this regime, it is essentially self-adjoint on $L^2(\mathbb{GL}_N)$ equipped with any right Haar measure; cf. [15, 23]. In the special case $s = t$, $A_{t,t}^N = \frac{t}{2} \Delta_{\mathbb{GL}_N}$. Note also that $t \Delta_{\mathbb{U}_N} = A_{t,0}^N|_{C^\infty(\mathbb{U}_N)}$.

Definition 2.4. For $t > 0$, the **heat kernel measure** ρ_t^N on \mathbb{U}_N is the unique probability measure which satisfies

$$\mathbb{E}_{\rho_t^N}(f) \equiv \int_{\mathbb{U}_N} f d\rho_t^N = \left(e^{\frac{t}{2} \Delta_{\mathbb{U}_N}} f \right) (I_N), \quad f \in C(\mathbb{U}_N). \quad (2.7)$$

Additionally, for $s > t/2$, the **heat kernel measure** $\mu_{s,t}^N$ on \mathbb{GL}_N is the unique probability measure which satisfies

$$\mathbb{E}_{\mu_{s,t}^N}(f) \equiv \int_{\mathbb{GL}_N} f d\mu_{s,t}^N = \left(e^{\frac{1}{2} A_{s,t}^N} f \right) (I_N), \quad f \in C_c(\mathbb{GL}_N). \quad (2.8)$$

In particular, the standard heat kernel measure on \mathbb{GL}_N is $\mu_{t/2}^N = \mu_{t,t}^N$; cf. Remark 2.3(3).

Remark 2.5. (1) The operators $e^{\frac{t}{2}\Delta_{\mathbb{U}_N}}$ and $e^{\frac{1}{2}A_{s,t}}$ can be made sense of with PDE methods (since $\Delta_{\mathbb{U}_N}$ and $A_{s,t}^N$ are elliptic) or functional analytic methods (since they are essentially self-adjoint). In most of our applications, the test functions f will be polynomials in the entries of the matrix argument, and the operators can be interpreted via the power series expansion of \exp .

(2) Eq. (2.8) holds, a priori, only for compactly-supported continuous test functions. In fact, it holds much more generally; in particular, it holds for any function f that is polynomial in the matrix entries. This follows from Langland's Theorem [33, Theorem 2.1 (p. 152)]; see also [16, Appendix A].

(3) More generally, for $s, t > 0$ and $s > t/2$, there is a strictly-positive smooth heat kernel function

$$h_{s,t}^N: \mathbb{GL}_N \times \mathbb{GL}_N \rightarrow \mathbb{R}_+$$

such that, for $f: \mathbb{GL}_N \rightarrow \mathbb{C}$ of sufficiently slow growth (as in (2) above),

$$\left(e^{\frac{1}{2}A_{s,t}^N} f\right)(Z) = \int h_{s,t}^N(Z, W) f(W) dW$$

where dW denotes the right-Haar measure on \mathbb{GL}_N . Thus, the density of $\mu_{s,t}^N$ is equal to $h_{s,t}^N(I_N, \cdot)$; cf. [15, 23]. Since $h_{s,t}^N$ is real-valued, for any f in the domain of $e^{\frac{1}{2}A_{s,t}^N}$ it follows that

$$\overline{e^{\frac{1}{2}A_{s,t}^N} f} = e^{\frac{1}{2}A_{s,t}^N} \bar{f},$$

where $\bar{f}(Z) = \overline{f(Z)}$ is the complex conjugate. Setting $t = 0$ shows that the same property holds for the heat operator $e^{\frac{s}{2}\Delta_{\mathbb{U}_N}}$. This will be useful in the proof of Lemma 3.11 below.

Remark 2.6. Had we taken the usual (unscaled) Hilbert-Schmidt inner product $(X, Y) = -\text{Tr}(XY)$ in (2.2), the resulting heat kernel measure on \mathbb{U}_N would have been ρ_{Nt}^N . This is the approach taken in [29, 30], and instead the heat kernel is evaluated at time t/N to compensate. In that sense, our limiting concentration results can be interpreted as statements about the heat kernel in a neighborhood of $t = 0$.

2.2 The Heat Kernel on \mathbb{U} , Sobolev Spaces, and Gevrey Classes

If $f \in L^2(\mathbb{U})$, its Fourier expansion is given by

$$f = \sum_{n \in \mathbb{Z}} \hat{f}(n) \chi_n, \quad \hat{f}(n) = \langle f, \chi_n \rangle_{L^2(\mathbb{U})} = \int_{\mathbb{U}} f(u) u^{-n} du,$$

where $\chi_n(u) = u^n$ for $u \in \mathbb{U}$ and $n \in \mathbb{Z}$, and du denotes the normalized Haar measure on \mathbb{U} .

Definition 2.7. For $p > 0$, the **Sobolev space** $H_p(\mathbb{U})$ is defined by

$$H_p(\mathbb{U}) = \left\{ f \in L^2(\mathbb{U}) : \|f\|_{H_p}^2 \equiv \sum_{n \in \mathbb{Z}} (1 + n^2)^p |\hat{f}(n)|^2 < \infty \right\}. \quad (2.9)$$

Note that $H_0(\mathbb{U}) = L^2(\mathbb{U})$. The definition makes sense even for $p < 0$, where the elements are no longer L^2 -functions but rather distributions. If $k, \ell \geq 1$ are integers, and $\ell \geq p \geq k + \frac{1}{2}$, then $C^\ell \subset H_p(\mathbb{U}) \subset C^k(\mathbb{U})$; it follows that $H_\infty(\mathbb{U}) \equiv \bigcap_{p \geq 0} H_p(\mathbb{U}) = C^\infty(\mathbb{U})$. For $\frac{1}{2} < p \leq \frac{3}{2}$, functions in $H_p(\mathbb{U})$ are Hölder continuous of any modulus $< p - \frac{1}{2}$, but generically not smoother. For $p \leq \frac{1}{2}$, $H_p(\mathbb{U})$ functions are generally not continuous. These are standard Sobolev imbedding theorems (that hold for smooth manifolds); for reference, see [18, Chapter 5.6] and [35, Chapter 3.2].

It is elementary to describe the heat semigroup on $\mathbb{U} = \mathbb{U}_1$ in terms of Fourier expansions. Indeed,

$$(\Delta_{\mathbb{U}_1} f)(u) = -\frac{\partial^2}{\partial u^2} f(u) \quad (2.10)$$

(Here $u = e^{i\theta}$; (2.10) is more commonly written as $(\Delta_{\mathbb{U}_1} f)(e^{i\theta}) = \frac{\partial^2}{\partial \theta^2} f(e^{i\theta})$ in PDE textbooks.) Hence, the characters χ_n are eigenfunctions $\Delta_{\mathbb{U}_1} \chi_n = -n^2 \chi_n$, and so

$$\Delta_{\mathbb{U}_1} \chi_n = e^{-\frac{t}{2} n^2} \chi_n, \quad n \in \mathbb{Z}, t \in \mathbb{R}. \quad (2.11)$$

It follows that the heat semigroup is completely described on $L^2(\mathbb{U})$ as a Fourier multiplier

$$e^{\frac{t}{2} \Delta_{\mathbb{U}_1}} f = \sum_{n \in \mathbb{Z}} e^{-\frac{t}{2} n^2} \hat{f}(n) \chi_n. \quad (2.12)$$

Let $f \in L^2(\mathbb{U})$, and for $t > 0$ let $f_t = e^{\frac{t}{2} \Delta_{\mathbb{U}_1}} f$. Then (2.12) shows that $\hat{f}_t(n) = e^{-\frac{t}{2} n^2} \hat{f}(n)$. In particular, this means that

$$\sum_{n \in \mathbb{Z}} e^{tn^2} |\hat{f}_t(n)|^2 = \sum_{n \in \mathbb{Z}} |\hat{f}(n)|^2 = \|f\|_{L^2(\mathbb{U})}^2 < \infty. \quad (2.13)$$

It follows that $f_t \in H_\infty(\mathbb{U}) = C^\infty(\mathbb{U})$. It is, in fact, *ultra-analytic*.

Definition 2.8. Let $\sigma > 0$. The **Gevrey class** $G_\sigma(\mathbb{U})$ consists of those $f \in L^2(\mathbb{U})$ such that

$$\|f\|_{G_\sigma}^2 \equiv \sum_{n \in \mathbb{Z}} e^{2\sigma n^2} |\hat{f}(n)|^2 < \infty. \quad (2.14)$$

More generally, the Gevrey class $G_\sigma^{s,p}(\mathbb{U})$ consists of those $f \in L^2(\mathbb{U})$ for which

$$\|f\|_{G_\sigma^{s,p}}^2 \equiv \sum_{n \in \mathbb{Z}} (1 + n^2)^p e^{2\sigma |n|^{1/s}} |\hat{f}(n)|^2 < \infty,$$

so that $G_\sigma(\mathbb{U})$ is the $s = 1/2$, $p = 0$ case of $G_\sigma^{s,p}(\mathbb{U})$.

These spaces arise naturally in the analysis of some non-linear parabolic PDEs, cf. [19, 20, 28]. The superexponent s is usually taken to be 1, in which case $G_\sigma^{1,p}$ is a Hilbert space of real analytic functions. For $s > 1$, Gevrey functions in $G_\sigma^{s,p}$ are C^∞ but generally not analytic, and when $s = \infty$ we recover the Sobolev spaces; thus the two-parameter family $G_\sigma^{s,p}$ interpolates between C^∞ functions and analytic functions for $s \geq 1$.

In the regime $0 < s < 1$ such functions are called **ultra-analytic**. Indeed, if $f \in G_\sigma(\mathbb{U})$ for some $\sigma > 0$, then f has a unique analytic continuation to a holomorphic function on \mathbb{C}^* given by the convergent Laurent series $f(z) = \sum_{n=-\infty}^{\infty} \hat{f}(n) z^n$. (The holomorphic $n \geq 0$ sum converges uniformly on \mathbb{C} and the principal part $n < 0$ converges uniformly on \mathbb{C}^* due to the fast decay of the coefficients.) We therefore refer to the set of such holomorphic functions as

$$G_\sigma(\mathbb{C}^*) = \{f \in \text{Hol}(\mathbb{C}^*) : f|_{\mathbb{U}} \in G_\sigma(\mathbb{U})\} = \left\{ f(z) = \sum_{n \in \mathbb{Z}} a_n z^n : \|f\|_{G_\sigma}^2 \equiv \sum_{n \in \mathbb{Z}} e^{2\sigma n^2} |a_n|^2 < \infty \right\}. \quad (2.15)$$

Note, as shown in (2.13), the Gevrey class G_σ characterizes the domain of the backwards heat flow:

$$G_\sigma(\mathbb{U}) = \left\{ f \in L^2(\mathbb{U}) : e^{-\frac{t}{2} \Delta_{\mathbb{U}_1}} f \text{ exists in } L^2(\mathbb{U}) \text{ for small time } 0 \leq t \leq 2\sigma \right\}. \quad (2.16)$$

2.3 Functional Calculus and Empirical Laws

For a normal matrix $X \in \mathbb{M}_N^{\text{nor}}$, the spectral theorem asserts that there are mutually orthogonal projection operators $\{\Pi_\lambda^X : \lambda \in \Lambda(X)\} \subset \text{End}(\mathbb{C}^N)$ so that

$$X = \sum_{\lambda \in \Lambda(X)} \lambda \Pi_\lambda^X.$$

For any measurable function $f : \mathbb{C} \rightarrow \mathbb{C}$, define $f_N : \mathbb{M}_N^{\text{nor}} \rightarrow \mathbb{M}_N^{\text{nor}}$ by

$$f_N(X) = \sum_{\lambda \in \Lambda(X)} f(\lambda) \Pi_\lambda^X. \quad (2.17)$$

That is: if $X = U\Lambda U^*$ is any unitary diagonalization of X , then $f_N(X) = Uf(\Lambda)U^*$ where $[f(\Lambda)]_{jj} = f([\Lambda]_{jj})$ for $1 \leq j \leq N$. The map $f \mapsto f_N$ is called **measurable functional calculus**. We adhere to the notation we used in [16]; in [10], f_N was denoted θ_f^N .

Let ρ^N be a probability measure supported in $\mathbb{M}_N^{\text{nor}}$. The linear functional

$$C_c(\mathbb{C}) \ni f \mapsto \int_{\mathbb{M}_N^{\text{nor}}} \text{tr}(f_N(X)) \rho^N(dX)$$

is easily verified to be positive; also, if $\overline{\mathbb{D}}_r$ is the disk of radius $r > 0$, then

$$\int_{\mathbb{M}_N^{\text{nor}}} \text{tr}([\mathbb{1}_{\overline{\mathbb{D}}_r}]_N(X)) \rho^N(dX) \rightarrow 1 \quad \text{as } n \rightarrow \infty.$$

Hence, by the Riesz Representation Theorem [34, Theorem 2.14], there is a unique Borel probability measure ν^N on \mathbb{C} such that

$$\int_{\mathbb{C}} f d\nu^N = \int_{\mathbb{M}_N^{\text{nor}}} (\text{tr} \circ f_N) d\rho^N, \quad f \in C_c(\mathbb{C}). \quad (2.18)$$

Comparing to (1.17), this Riesz measure ν^N is the mean of the empirical law $\tilde{\nu}^N$ (1.15). In particular, if ν is a (deterministic) measure such that $\tilde{\nu}^N \rightharpoonup \nu$ weakly in probability, then we must have $\nu^N \rightharpoonup \nu$ weakly.

Remark 2.9. In the special case that $\text{supp}(\rho^N)$ is compact, the Weierstrass approximation theorem shows that (2.18) is equivalent to equating the moments of ν^N with the *trace moments* of ρ^N :

$$\int_{\mathbb{C}} x^n \bar{x}^m \nu^N(dx) = \int_{\mathbb{M}_N^{\text{nor}}} \text{tr}(X^n (X^*)^m) \rho^N(dX). \quad (2.19)$$

In our first case of interest where ρ_t^N is the heat kernel on the compact group \mathbb{U}_N , this amounts to defining ν_t^N by its integrals against *Laurent polynomials*; cf. Section 2.4.

If $\text{supp}(\rho^N)$ is not contained in $\mathbb{M}_N^{\text{nor}}$, measurable functional calculus is not available. Instead, we can consider *holomorphic* test functions. In the case of interest (the heat kernel $\mu_{s,t}^N$ on \mathbb{GL}_N), all empirical eigenvalues are in \mathbb{C}^* , so we take $f \in \text{Hol}(\mathbb{C}^*)$; for simplicity, we assume the Laurent series $f(z) = \sum_{n=-\infty}^{\infty} a_n z^n$ converges on all of \mathbb{C}^* . (This is not necessary, but it simplifies matters and suffices for our purposes.) Then the series

$$f_N(Z) \equiv \sum_{n=-\infty}^{\infty} a_n Z^n, \quad (2.20)$$

where we interpret the $n = 0$ term as $a_0 I_N$, converges for any $Z \in \mathbb{GL}_N$. The map $f \mapsto f_N$ is called **holomorphic functional calculus**. We use the same notation as for functional calculus, and this is consistent: if

Z is normal and f is holomorphic as above, then the Laurent series (2.20) coincides with the functional calculus map of (2.17).

Since there are no non-constant positive holomorphic functions, no integration formula like (2.18) can be used to define an “expected empirical eigenvalue law” in this case. There may or may not exist such a measure ν^N on \mathbb{C} ; if it does exist, it will not be uniquely determined by (2.18). In general, there is just too much information in the trace (noncommutative) moments of a non-normally supported measure ρ^N to be captured by a single measure on \mathbb{C} . Instead, we need the notion of a *noncommutative distribution*.

2.4 Noncommutative Distributions

Definition 2.10. Let \mathcal{A} be a unital complex $*$ -algebra. A **tracial state** $\tau: \mathcal{A} \rightarrow \mathbb{C}$ is a linear functional that is unital ($\tau(1) = 1$), tracial ($\tau(ab) = \tau(ba)$ for $a, b \in \mathcal{A}$), and positive semidefinite ($\tau(aa^*) \geq 0$ for all $a \in \mathcal{A}$). If, in addition, $\tau(aa^*) \neq 0$ for $a \neq 0$, τ is called **faithful**. The pair (\mathcal{A}, τ) is called a (faithful, tracial) **noncommutative probability space**. If \mathcal{A} is a C^* -algebra, we refer to (\mathcal{A}, τ) as a C^* -probability space; if \mathcal{A}^* is a W^* -algebra (i.e. von Neumann algebra), we refer to (\mathcal{A}, τ) as a W^* -probability space.

If (Ω, \mathcal{F}) is a probability space and \mathbb{P} is a probability measure on (Ω, \mathcal{F}) , the expectation $\mathbb{E} = \int \cdot d\mathbb{P}$ is a faithful tracial state on the algebra $L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ of complex-valued random variables (where $F^* = \overline{F}$); thus the *probability space* terminology. Truly noncommutative examples are afforded by \mathbb{M}_N equipped with tr , which is a faithful tracial state. It is these examples that will be most relevant to us.

In the example $L^\infty(\Omega, \mathcal{F}, \mathbb{P})$, any random variable $F \in L^\infty$ has a probability distribution μ_F (on \mathbb{C} if the random variables are \mathbb{C} -valued), which is the push-forward $\mu_F(B) = (F_*(\mathbb{P}))(B) = \mathbb{P}(F^{-1}(B))$ for Borel sets $B \subseteq \mathbb{C}$. In terms of the expectation, this can be written as

$$\int f d\mu_F = \mathbb{E}(f(F)), \quad f \in C_c(\mathbb{C}). \quad (2.21)$$

If (\mathcal{A}, τ) is a W^* -probability space, any measurable $f: \mathbb{C} \rightarrow \mathbb{C}$ induces (by the spectral theorem) a function $f_{\mathcal{A}}: \mathcal{A}^{\text{nor}} \rightarrow \mathcal{A}^{\text{nor}}$ as in (2.17); here \mathcal{A}^{nor} refers to the normal operators in \mathcal{A} . The map $f \mapsto f_{\mathcal{A}}$ is the *measurable functional calculus*. We then define the **distribution** μ_a of $a \in \mathcal{A}^{\text{nor}}$ to be the unique Borel probability measure on \mathbb{C} mimicking (2.21):

$$\int_{\mathbb{C}} f d\mu_a = \tau(f_{\mathcal{A}}(a)), \quad f \in C_c(\mathbb{C}). \quad (2.22)$$

Indeed, (2.22) determines μ_a for $f \in C(\sigma(a))$, as the spectrum $\sigma(a)$ is compact (since $a \in \mathcal{A}$ is a bounded operator). Therefore, as in (2.19), in (2.22) we need only use test functions of the form $f(x) = x^n \bar{x}^m$, $n, m \in \mathbb{N}$, so that $f_{\mathcal{A}}(a) = a^n (a^*)^m$. Hence, in this case, μ_a is equivalently determined by all moments, through the formula

$$\int_{\mathbb{C}} x^n \bar{x}^m \mu_a(dx) = \tau(a^n (a^*)^m), \quad n, m \in \mathbb{N}. \quad (2.23)$$

Remark 2.11. In the special case $(\mathcal{A}, \tau) = (\mathbb{M}_N, \text{tr})$, the distribution of a normal matrix is precisely its empirical eigenvalue law; cf. (1.16).

If a is a non-normal operator in (\mathcal{A}, τ) , it may or may not be the case that there is a measure μ_a on \mathbb{C} satisfying (2.23). Even if there is, these moments do not determine all other moments $\tau(a^{n_1} (a^*)^{m_1} \dots a^{n_k} (a^*)^{m_k})$. We therefore *define* this collection of moments to be the noncommutative distribution of a . In the spirit of the Riesz theorem identifying measures as linear functionals, this can be formulated as follows.

Definition 2.12. Let $\mathbb{C}\langle A, A^{-1}, A^*, A^{-*} \rangle$ denote the algebra of **noncommutative Laurent polynomials** in two variables A and A^* ; in other words, $\mathbb{C}\langle A, A^{-1}, A^*, A^{-*} \rangle \cong \mathbb{C}\mathbb{F}_2$ is the complex group algebra of the free group

on two generators A, A^* . Let $\mathbb{C}\langle A, A^* \rangle$ denote the subalgebra of **noncommutative polynomials** in two variables A, A^* ; in other words, $\mathbb{C}\langle A, A^* \rangle \cong \mathbb{C}\mathbb{F}_2^+$ is the group algebra over the free semigroup \mathbb{F}_2^+ generated by A, A^* .

If (\mathcal{A}, τ) is a noncommutative probability space and $a \in \mathcal{A}$, the **noncommutative distribution** of a is the linear functional $\varphi_a: \mathbb{C}\langle A, A^* \rangle \rightarrow \mathbb{C}$ defined by

$$\varphi_a(f) = \tau[f(a, a^*)], \quad f \in \mathbb{C}\langle A, A^* \rangle \quad (2.24)$$

for any element $f = f(A, A^*)$. If a is invertible in \mathcal{A} , then φ_a extends uniquely to a linear functional on $\mathbb{C}\langle A, A^{-1}, A^*, A^{-*} \rangle$ by (2.24).

Notation 2.13. For $n \in \mathbb{N}$, let \mathcal{E}_n denote the set of all n -tuples $\varepsilon \in \{\pm 1, \pm *\}^n$, and let \mathcal{E}_n^+ be the subset $\{1, *\}^n$. ($\mathcal{E}_0 = \emptyset$.) For $\varepsilon \in \mathcal{E}_n$, denote $|\varepsilon| = n$. Set $\mathcal{E} = \bigcup_n \mathcal{E}_n$, and $\mathcal{E}^+ = \bigcup_n \mathcal{E}_n^+$.

Given a $*$ -algebra \mathcal{A} , for $a \in \mathcal{A}$ and $\varepsilon \in \mathcal{E}^+$, denote $a^\varepsilon = a^{\varepsilon_1} a^{\varepsilon_2} \cdots a^{\varepsilon_n}$ where $n = |\varepsilon|$. Then $\mathbb{C}\langle A, A^* \rangle$ can be described explicitly as

$$\mathbb{C}\langle A, A^* \rangle = \text{span}_{\mathbb{C}} \{A^\varepsilon : \varepsilon \in \mathcal{E}^+\}.$$

The vectors A^ε form a basis for this \mathbb{C} -space. The algebra structure is given by concatenation in \mathcal{E}^+ : $A^\varepsilon \cdot A^\delta = A^{\varepsilon\delta}$ where, if $\varepsilon \in \mathcal{E}_n^+$ and $\delta \in \mathcal{E}_m^+$, then $\varepsilon\delta = (\varepsilon_1, \dots, \varepsilon_n, \delta_1, \dots, \delta_m) \in \mathcal{E}_{n+m}^+$.

The algebra $\mathbb{C}\langle A, A^{-1}, A^*, A^{-*} \rangle$ is similarly equal to the \mathbb{C} -span of A^ε for $\varepsilon \in \mathcal{E}$, with product defined by concatenation; but in this case these words are no longer linearly independent (for example $A^* A A^{-1} = A^*$). A basis for $\mathbb{C}\langle A, A^{-1}, A^*, A^{-*} \rangle$ consists of reduced words A^ε in the sense of free groups.

Thus, the noncommutative distribution of $a \in (\mathcal{A}, \tau)$ can equivalently be described as the linear functional $\varphi_a: \mathbb{C}\langle A, A^* \rangle \rightarrow \mathbb{C}$ defined by

$$\varphi_a(A^\varepsilon) = \tau(a^\varepsilon), \quad \varepsilon \in \mathcal{E}^+. \quad (2.25)$$

If a is invertible in \mathcal{A} , this extends by the same formula to a linear functional on $\mathbb{C}\langle A, A^{-1}, A^*, A^{-*} \rangle$ (due to the universal property of free groups).

If a is normal, then for any $\varepsilon \in \mathcal{E}^+$, $a^\varepsilon = a^n (a^*)^m$ where n is the number of 1s and m is the number of *s in ε . Hence, in this case, φ_a is completely determined by the measure μ_a of (2.23). Thus φ_a generalizes the classical notion of distribution of a random variable.

We will work largely with the noncommutative probability spaces $(\mathbb{M}_N, \text{tr})$, often with randomness involved.

Definition 2.14. Let ρ^N be a probability measure on \mathbb{M}_N , such that all polynomial functions of the matrix entries are in $L^1(\rho^N)$; this condition holds for the heat kernel measures $\mu_{s,t}^N$ on $\mathbb{G}\mathbb{L}_N$ by Remark 2.5(2). The associated **empirical noncommutative distribution** $\tilde{\varphi}^N$ is defined to be the $\text{Hom}(\mathbb{C}\langle A, A^* \rangle; \mathbb{C})$ -valued random variable on the probability space (\mathbb{M}_N, ρ^N) given by

$$\tilde{\varphi}^N(Z) = \varphi_Z \text{ with respect to the noncommutative probability space } (\mathbb{M}_N, \text{tr}). \quad (2.26)$$

That is: $(\tilde{\varphi}^N(Z))(A^\varepsilon) = \text{tr}(Z^\varepsilon)$ for $\varepsilon \in \mathcal{E}^+$. If ρ^N is supported on $\mathbb{G}\mathbb{L}_N$, then $\tilde{\varphi}^N$ extends to a random linear functional on $\mathbb{C}\langle A, A^{-1}, A^*, A^{-*} \rangle$. The expectation $\mathbb{E}(\tilde{\varphi}^N)$ is defined to be the linear functional on $\mathbb{C}\langle A, A^* \rangle$ given by

$$\mathbb{E}(\tilde{\varphi}^N)(f) = \int_{\mathbb{M}_N} \varphi_Z(f) \rho^N(dZ), \quad f \in \mathbb{C}\langle A, A^* \rangle. \quad (2.27)$$

Equations (2.26) and (2.27) are natural generalization of (1.16) and (1.17). The polynomial-integrability condition we placed on ρ^N guarantees that (2.27) is a well-defined linear functional; moreover, $\mathbb{E}(\tilde{\varphi}^N)$ is the noncommutative distribution of some random variable. Indeed, we can construct this random variable in the algebra $\mathbb{C}\langle A, A^* \rangle$ itself. Define the linear functional τ_{ρ^N} on $\mathbb{C}\langle A, A^* \rangle$ to verify (2.27):

$$\tau_{\rho^N}(f) = \int_{\mathbb{M}_N} \varphi_Z(f) \rho^N(dZ) = \int_{\mathbb{M}_N} \text{tr}[f(Z, Z^*)] \rho^N(dZ).$$

The linear functional τ_{ρ^N} is easily verified to be a tracial state, so $(\mathcal{A}, \tau_{\rho^N})$ is a noncommutative probability space; cf. Definition 2.10. It is faithful provided $\text{supp}(\rho^N)$ is infinite. Let $a \in \mathbb{C}\langle A, A^* \rangle$ denote the coordinate random variable $a(A, A^*) = A$; then its noncommutative distribution φ_a with respect to $(\mathbb{C}\langle A, A^* \rangle, \tau_{\rho^N})$ is, by (2.25) and (2.27),

$$\varphi_a(A^\varepsilon) = \tau_{\rho^N}(a(A)^\varepsilon) = \tau_{\rho^N}(A^\varepsilon) = \int_{\mathbb{M}_N} \text{tr}(Z^\varepsilon) \rho^N(dZ) = \mathbb{E}(\tilde{\varphi}^N)(A^\varepsilon), \quad \varepsilon \in \mathcal{E}^+.$$

Thus, $\mathbb{E}(\tilde{\varphi}^N)$ defines a (deterministic) noncommutative distribution which we call the **mean** of $\tilde{\varphi}^N$.

Definition 2.15. Let φ^N be a sequence of noncommutative distributions; that is, there are noncommutative probability spaces (\mathcal{A}_N, τ_N) with some distinguished elements $a_N \in \mathcal{A}_N$ so that $\varphi^N = \varphi_{a_N}$ over \mathcal{A}_N . We say that φ^N **converges weakly** (or **converges in distribution**) if there is a noncommutative distribution φ so that $\varphi^N(f) \rightarrow \varphi(f)$ for all $f \in \mathbb{C}\langle A, A^* \rangle$. That is: there exists a noncommutative probability space (\mathcal{A}, τ) with a distinguished element $a \in \mathcal{A}$ so that $\varphi = \varphi_a$, such that $\varphi_{a_N}(f) \rightarrow \varphi_a(f)$ for all $f \in \mathbb{C}\langle A, A^* \rangle$.

Thus, Theorem 1.7 asserts that, in the case $\rho^N = \mu_{s,t}^N$, the mean empirical noncommutative distribution $\varphi_{s,t}^N = \mathbb{E}(\tilde{\varphi}_{s,t}^N)$ converges weakly, and moreover the empirical distribution converges weakly almost surely to the limit. As these distributions are supported on invertible operators, the weak convergence statements hold on the larger class of “test functions” $f \in \mathbb{C}\langle A, A^{-1}, A^*, A^{-*} \rangle$.

We now introduce extensions of $\mathbb{C}\langle A, A^* \rangle$ and $\mathbb{C}\langle A, A^{-1}, A^*, A^{-*} \rangle$ that deserve to be called the **universal enveloping algebras** of these spaces. The reader is also directed to [16, Section 3.4].

Notation 2.16. With \mathcal{E} and \mathcal{E}^+ as in Notation 2.13, define

$$\mathcal{P} = \mathbb{C}[\{v_\varepsilon\}_{\varepsilon \in \mathcal{E}}] \quad \text{and} \quad \mathcal{P}^+ = \mathbb{C}[\{v_\varepsilon\}_{\varepsilon \in \mathcal{E}^+}] \subset \mathcal{P}, \quad (2.28)$$

the spaces of polynomials in the (commuting) indeterminates v_ε . Elements of these spaces are generally denoted P, Q, R ; when emphasizing their variables, we write $P(\mathbf{v}) = P(\{v_\varepsilon\})$. For shorthand, we denote

$$v_k = v_{\varepsilon(k)}, \quad k \in \mathbb{Z} \setminus \{0\}, \quad (2.29)$$

where $\varepsilon(k) = (\overbrace{1, \dots, 1}^k)$ for $k > 0$ and $\varepsilon(k) = (\overbrace{-1, \dots, -1}^{|k|})$ for $k < 0$. Set $v_0 \equiv 1$. Define the subalgebra $\mathcal{HP} \subset \mathcal{P}$ as follows:

$$\mathcal{HP} = \mathbb{C}[\{v_k\}_{k \in \mathbb{Z} \setminus \{0\}}]. \quad (2.30)$$

Remark 2.17. In [16], \mathcal{P} was referred to as \mathcal{W} , while \mathcal{HP} was simply denoted $\mathbb{C}[\mathbf{v}]$.

We may naturally identify $\mathbb{C}\langle A, A^* \rangle$ as a linear subspace of \mathcal{P}^+ , via the linear map

$$\Upsilon: \mathbb{C}\langle A, A^* \rangle \rightarrow \mathcal{P}^+ \quad \text{defined by} \quad \Upsilon(A^\varepsilon) = v_\varepsilon, \quad \varepsilon \in \mathcal{E}^+. \quad (2.31)$$

This is a complex vector space isomorphism from $\mathbb{C}\langle A, A^* \rangle$ onto $\text{span}_{\mathbb{C}}\{v_\varepsilon: \varepsilon \in \mathcal{E}^+\}$, the space of linear polynomials in \mathcal{P}^+ . A similar identification could be made for $\mathbb{C}\langle A, A^{-1}, A^*, A^{-*} \rangle$ in \mathcal{P} , although for the inclusion to be well-defined and one-to-one we must restrict $\varepsilon \in \mathcal{E}$ to *reduced* words in the sense of \mathbb{F}_2 ; then $\Upsilon(\mathbb{C}\langle A, A^{-1}, A^*, A^{-*} \rangle)$ is a strict subset of the linear polynomials in \mathcal{P} . Thus, if φ is a linear functional on $\mathbb{C}\langle A, A^* \rangle$, it extends uniquely to a *homomorphism* $\mathcal{P}^+ \rightarrow \mathbb{C}$; in this sense, \mathcal{P}^+ is the universal enveloping algebra of $\mathbb{C}\langle A, A^* \rangle$. This will be useful in Section 3.1, and so we record this new role for φ_a in the following notation.

Notation 2.18. Let (\mathcal{A}, τ) be a noncommutative probability space. Let $\varepsilon \in \mathcal{E}^+$, and define $V_\varepsilon: \mathcal{A} \rightarrow \mathbb{C}$ by $V_\varepsilon = \varphi_{(\cdot)}(A^\varepsilon)$:

$$V_\varepsilon(a) = \tau(a^\varepsilon) = \tau(a^{\varepsilon_1} a^{\varepsilon_2} \cdots a^{\varepsilon_n}), \quad (2.32)$$

where $n = |\varepsilon|$. Let \mathcal{A}^{inv} denote the group of invertible elements in \mathcal{A} . Then $V_\varepsilon: \mathcal{A}^{\text{inv}} \rightarrow \mathbb{C}$ is well-defined for any $\varepsilon \in \mathcal{E}$ by (2.32), setting $a^{+*} \equiv a^*$ and $a^{-*} \equiv (a^*)^{-1} = (a^{-1})^*$.

Remark 2.19. Strictly speaking, we should denote $V_\varepsilon = V_\varepsilon^{(\mathcal{A}, \tau)}$ since this symbol represents different functions on different noncommutative probability spaces. We will usually suppress this indexing, which will always be clear from context.

2.5 Free Probability

Definition 2.20. Let (\mathcal{A}, τ) be a noncommutative probability space. Unital $*$ -subalgebras $\mathcal{A}_1, \dots, \mathcal{A}_m \subset \mathcal{A}$ are called **free** with respect to τ if, given any $n \in \mathbb{N}$ and $k_1, \dots, k_n \in \{1, \dots, m\}$ such that $k_{j-1} \neq k_j$ for $1 < j \leq n$, and any elements $a_j \in \mathcal{A}_{k_j}$ with $\tau(a_j) = 0$ for $1 \leq j \leq n$, it follows that $\tau(a_1 \cdots a_n) = 0$. Random variables a_1, \dots, a_m are said to be **freely independent** if the unital $*$ -algebras $\mathcal{A}_j = \langle a_j, a_j^* \rangle \subset \mathcal{A}$ they generate are free.

Free independence is a $*$ -moment factorization property. By centering $a_j - \tau(a_j)1_{\mathcal{A}} \in \mathcal{A}_j$, the freeness rule allows (inductively) any moment $\tau(a_{k_1}^{\varepsilon_1} \cdots a_{k_n}^{\varepsilon_n})$ to be decomposed as a polynomial in moments $\tau(a_j^{\varepsilon_j})$ in the variables separately. In terms of Definition 2.12 (which can be extended naturally to the multivariate case, see [31, Lecture 4]), if a_1, \dots, a_m are freely independent then their joint noncommutative distribution $\varphi_{a_1, \dots, a_n}$ is determined (computationally effectively) by the individual noncommutative distributions $\varphi_{a_1}, \dots, \varphi_{a_m}$.

If \mathcal{A} is a W^* -algebra and $a \in \mathcal{A}$ is normal, then φ_a is completely described by a compactly-supported measure μ_a on \mathbb{C} ; cf. (2.22). Thus, if $u, v \in \mathcal{A}$ are freely independent unitary operators, uv is also unitary, and the distributions μ_u, μ_v , and μ_{uv} are supported on \mathbb{U} . Since μ_{uv} is determined by $\varphi_{u,v}$ which, by freeness, is determined by μ_u and μ_v , there is a well-defined operation, **free multiplicative convolution** \boxtimes , on probability measures on \mathbb{U} such that $\mu_{uv} = \mu_u \boxtimes \mu_v$. Similarly, if $x, y \in \mathcal{A}$ are positive definite, the distribution φ_{xy} of their product is determined by the measures μ_x and μ_y supported in \mathbb{R}_+ . Although xy is not necessarily normal, it is easy to check that it has the same moments as the positive definite operators $\sqrt{xy}\sqrt{x}$ and $\sqrt{yx}\sqrt{y}$. So if we define $x \odot y = \sqrt{xy}\sqrt{x}$, then there is a well-defined operation \boxtimes on probability measures on \mathbb{R}_+ such that $\mu_{x \odot y} = \mu_x \boxtimes \mu_y$; this is also called free multiplicative convolution. In both frameworks, it can be described succinctly in terms of the **Σ -transform**.

Definition 2.21. Let μ be a probability measure on \mathbb{C} . Define the function

$$\psi_\mu(z) = \int_{\mathbb{C}} \frac{\zeta z}{1 - \zeta z} \mu(d\zeta), \quad \frac{1}{z} \notin \text{supp } \mu,$$

which is analytic on its domain. If μ is supported in \mathbb{U} , it is customary to restrict ψ_μ to the unit disk \mathbb{D} ; if μ is supported in \mathbb{R} , it is customary to restrict ψ_μ to the upper half-plane \mathbb{C}_+ . Define $\eta_\mu(z) = \psi_\mu(z)/(1 + \psi_\mu(z))$. This function is one-to-one on a neighborhood of 0 if $\text{supp } \mu \subset \mathbb{U}$ (and the first moment of μ is non-zero); it is one-to-one on the right-half plane $i\mathbb{C}_+$ if $\text{supp } \mu \subset \mathbb{R}_+$; cf. [8]. The **Σ -transform** Σ_μ is the analytic function

$$\Sigma_\mu(z) = \frac{\eta_\mu^{-1}(z)}{z}, \quad (2.33)$$

for z in a neighborhood of 0 in the \mathbb{U} -case and for $z \in \eta_\mu(i\mathbb{C}_+)$ in the \mathbb{R}_+ -case.

The Σ -transform is a \boxtimes -homomorphism: as shown in [7, 36],

$$\Sigma_{\mu \boxtimes \nu}(z) = \Sigma_{\mu}(z) \Sigma_{\nu}(z) \quad (2.34)$$

for any probability measures μ, ν both supported in \mathbb{U} (resp. \mathbb{R}_+), and any z in a sufficiently small neighborhood of 0 (resp. open set in $\eta_{\mu}(i\mathbb{C}_+)$).

Theorem 2.22 (Biane, 1997). [9] *The measures $\{\nu_t\}_{t \in \mathbb{R}}$ of Definition 1.1 have Σ -transforms*

$$\Sigma_{\nu_t}(z) = e^{\frac{t}{2} \frac{1+z}{1-z}} \quad (2.35)$$

defined and analytic everywhere on $\mathbb{C} \setminus \{1\}$. Hence, from (2.34), they form a \boxtimes -group: for $s, t \in \mathbb{R}$, $\nu_{s+t} = \nu_s \boxtimes \nu_t$.

Remark 2.23. In terms of the above discussion of free multiplicative convolution, $\nu_s \boxtimes \nu_t$ only makes sense if $st \geq 0$. If, instead, we take (2.34) as the definition of \boxtimes , then (2.35) shows the \boxtimes -group property holds for all $s, t \in \mathbb{R}$.

Equation 2.35 was the starting point for investigation of the measures ν_t (with $t > 0$). In [7, Lemmas 6.3 and 7.1], the authors showed that (2.35) defines a measure ν_t that is an analogue of the Gaussian on \mathbb{R} : it is the free multiplicative convolution power limit of a(n appropriately scaled) two-point measure. Later, in [9, Lemma 1], Biane showed that these measures have the moments given in (1.2). Using complex analytic techniques, a great deal of information can be gleaned about these measures. The state of the art is summarized in the following proposition, where the $t > 0$ statements were proved in [9], while the $t < 0$ case follows from results in [5, 6, 7] and the recent preprint [40].

Proposition 2.24. *For $t > 0$, ν_t has a continuous density ϱ_t with respect to the normalized Haar measure on \mathbb{U} . For $0 < t < 4$, its support is the connected arc*

$$\text{supp } \nu_t = \left\{ e^{i\theta} : -\frac{1}{2} \sqrt{t(4-t)} - \arccos \left(1 - \frac{t}{2} \right) \leq \theta \leq \frac{1}{2} \sqrt{t(4-t)} + \arccos \left(1 - \frac{t}{2} \right) \right\},$$

while $\text{supp } \nu_t = \mathbb{U}$ for $t \geq 4$. The density ϱ_t is real analytic on the interior of the arc. It is symmetric about 1, and is determined by $\varrho_t(e^{i\theta}) = \Re \kappa_t(e^{i\theta})$ where $z = \kappa_t(e^{i\theta})$ is the unique solution (with positive real part) to

$$\frac{z-1}{z+1} e^{\frac{t}{2}z} = e^{i\theta}.$$

For $t < 0$, ν_t has a continuous density ϱ_t with respect to Lebesgue measure on \mathbb{R}_+ . The support is the connected interval $\text{supp } \nu_t = [r_-(t), r_+(t)]$ where

$$r_{\pm}(t) = \frac{2-t \pm \sqrt{t(t-4)}}{2} e^{-\frac{1}{2} \sqrt{t(t-4)}}.$$

The density ϱ_t is real analytic on the interval $(r_-(t), r_+(t))$, unimodal with peak at its mean 1; it is determined by $\varrho_t(x) = \frac{1}{\pi x} \Im \zeta_t(x)$ where $z = \zeta_t(x)$ is the unique solution to

$$\frac{z}{z-1} e^{-t(z-\frac{1}{2})} = x.$$

When $t > 0$, the measure ν_t is the distribution of the **free unitary Brownian motion** introduced in [9]. The (left) free unitary Brownian motion is a stochastic process $(u_t)_{t \geq 0}$ satisfying $\varphi_{u_t} = \nu_t$ for $t \geq 0$ such that the multiplicative increments $u_{t_1}, u_{t_1}^{-1} u_{t_2}, \dots, u_{t_{n-1}}^{-1} u_{t_n}$ are freely independent and stationary for $0 < t_1 < t_2 < \dots < t_n < \infty$. The process $(u_t)_{t \geq 0}$ is constructed as the solution of a free stochastic differential equation. Let

(\mathcal{A}, τ) be a noncommutative probability space that contains a free semicircular Brownian motion x_t (see [12, 27] for an introduction to free stochastic calculus). Then u_t is defined to be the unique solution to the free SDE

$$du_t = iu_t dx_t - \frac{1}{2}u_t dt \quad (2.36)$$

with $u_0 = 1$. This precisely mirrors the matrix SDE satisfied by the Brownian motion on \mathbb{U}_N (although the proof that u_t is the noncommutative limit of this process does not follow easily from this observation).

For Section 4.2, it will also be useful to consider the **free multiplicative Brownian motion**, which is nominally the large- N limit of the Brownian motion on \mathbb{GL}_N . Let (\mathcal{A}, τ) be a noncommutative probability space that contains two freely independent semicircular Brownian motions x_t, y_t . Then $z_t = \frac{1}{\sqrt{2}}(x_t + iy_t)$ is called a **circular Brownian motion**. The free multiplicative Brownian motion g_t is defined to be the unique solution to the free SDE

$$dg_t = g_t dz_t \quad (2.37)$$

with $z_0 = 1$. Again, this precisely mirrors the matrix SDE satisfied by the Brownian motion on \mathbb{GL}_N . It was left as an open problem in [9] whether g_t is the limit in noncommutative distribution of the \mathbb{GL}_N Brownian motion. Theorem 1.7 is a partial answer to (the (s, t) -generalization of) this question. The recent paper [11] also partially solves this conjecture in the $s = t$ case. In fact, the full conjecture is solved by the author in the complementary paper [26].

3 Intertwining Operators and Concentration of Measure

In this section, we summarize the relevant results from the author's recent joint paper [16], in addition to giving some estimates of the involved constants.

3.1 The Action of $\Delta_{\mathbb{U}_N}$ and $A_{s,t}^N$ on Trace Polynomials

If $Z \in \mathbb{GL}_N$, the noncommutative distribution φ_Z (viewed as a homomorphism on \mathcal{P} , as in Section 2.4) induces a family of functions of Z : linear combinations of products of traces $\text{tr}(Z^{\varepsilon^{(1)}}) \cdots \text{tr}(Z^{\varepsilon^{(m)}})$. We call such functions **trace polynomials**; cf. Notation 3.2 below. In this section, we will describe the action of the generalized Laplacian $A_{s,t}^N$ (and its special case $\Delta_{\mathbb{U}_N} = A_{1,0}^N|_{\mathbb{U}_N}$) on trace polynomials. We will rely heavily upon Notation 2.16, as well as the following.

Notation 3.1. Given $\varepsilon^{(1)}, \dots, \varepsilon^{(m)} \in \mathcal{E}$, we say that the monomial $v_{\varepsilon^{(1)}} \cdots v_{\varepsilon^{(m)}}$ has **trace degree** equal to $|\varepsilon^{(1)}| + \cdots + |\varepsilon^{(m)}|$. More generally, given any polynomial $P \in \mathcal{P}$, the **trace degree** of P , denoted $\text{deg}(P)$, is the highest trace degree among its monomial terms; if all terms have trace degree n , we say the polynomial has **homogeneous trace degree** n .

For $n \in \mathbb{N}$, let $\mathcal{P}_n = \{P \in \mathcal{P} : \text{deg}(P) \leq n\}$. Note that \mathcal{P}_n is finite dimensional, $\mathcal{P}_n \subset \mathbb{C}[\{v_\varepsilon\}_{|\varepsilon| \leq n}]$, and $\mathcal{P} = \bigcup_{n \geq 1} \mathcal{P}_n$. The sets $\mathcal{H}\mathcal{P}_n$ are defined similarly. In particular, $\mathcal{H}\mathcal{P}_n \subset \mathbb{C}[v_{\pm 1}, \dots, v_{\pm n}]$, and, in terms of (2.29), this means

$$\text{deg}(v_1^{k_1} v_{-1}^{k_{-1}} \cdots v_n^{k_n} v_{-n}^{k_{-n}}) = \sum_{1 \leq |j| \leq n} |j| k_j.$$

Notation 3.2. Let $(\mathbb{M}_N)^\mathcal{E}$ denote the set of functions $\mathcal{E} \rightarrow \mathbb{M}_N$. Denote by \mathbf{V}_N the map $\mathbb{GL}_N \rightarrow (\mathbb{M}_N)^\mathcal{E}$ given by

$$[\mathbf{V}_N(Z)](\varepsilon) = V_\varepsilon(Z) = \text{tr}(Z^\varepsilon), \quad Z \in \mathbb{GL}_N, \varepsilon \in \mathcal{E}.$$

For $P \in \mathcal{P}$, we write $P \circ \mathbf{V}_N$ for the evaluation of P as a function on \mathbb{GL}_N . That is: if $\varepsilon^{(1)}, \dots, \varepsilon^{(n)} \in \mathcal{E}$ are such that $P = P(v_{\varepsilon^{(1)}}, \dots, v_{\varepsilon^{(n)}})$ is in $\mathbb{C}[v_{\varepsilon^{(1)}}, \dots, v_{\varepsilon^{(n)}}]$, then

$$(P \circ \mathbf{V}_N)(Z) = P(V_{\varepsilon^{(1)}}(Z), \dots, V_{\varepsilon^{(n)}}(Z)).$$

We refer to any such function as a **trace polynomial**.

Note: in [16], the trace polynomial $P \circ \mathbf{V}_N$ was often denoted simply as P_N .

Example 3.3. If $P(\mathbf{v}) = v_{(1,*)}v_{(*)} + 2v_{(*,-1,1)}$ then $\deg(P) = 3$, and

$$(P \circ \mathbf{V}_N)(Z) = \text{tr}(ZZ^*)\text{tr}(Z^*) + 2\text{tr}(Z^*Z^{-1}Z) = \text{tr}(ZZ^*)\text{tr}(Z^*) + 2\text{tr}(Z^*).$$

Thus, if we set $Q(\mathbf{v}) = v_{(1,*)}v_{(*)} + 2v_{(*)}$, then $P \circ \mathbf{V}_N = Q \circ \mathbf{V}_N$ for all N . That is, the map $P \mapsto P \circ \mathbf{V}_N$ from \mathcal{P} to the space of trace polynomials is not one-to-one for any N . If we restrict this map to $\mathcal{H}\mathcal{P}$, cancellations like this do not occur; nevertheless, the map is still not one-to-one, due to the Cayley-Hamilton theorem, as explained in [16, Section 2.4]. Nevertheless, restricted to $\mathcal{H}\mathcal{P}_n$ for some $n \in \mathbb{N}$, the map is one-to-one for all sufficiently large N (depending on n).

Remark 3.4. Note that, if $P \in \mathcal{H}\mathcal{P}$, then the function $P \circ \mathbf{V}_N$ is holomorphic on $\mathbb{G}\mathbb{L}_N$. This is the reason we use the notation $\mathcal{H}\mathcal{P}$.

We now introduce two families of polynomials $\{Q_\varepsilon^\pm : \varepsilon \in \mathcal{E}\}$ and $\{R_{\varepsilon,\delta}^\pm : \varepsilon, \delta \in \mathcal{E}\}$ in \mathcal{P} that were introduced in [16, Theorem 3.12]. Since we do not need to know all the details about these polynomials, the following is only as precise as will be needed below (in particular in Proposition 3.18).

Definition 3.5. Let $\varepsilon \in \mathcal{E}$, and let $1 \leq j < k \leq |\varepsilon|$. Define $n_\pm(\varepsilon)$ be the integer from [16, Eq. (3.36)]; in particular, $|n_\pm(\varepsilon)| \leq |\varepsilon|$, and let $\{\varepsilon_{j,k}^\ell : \ell = 0, 1, 2\}$ be the substrings of ε given in [16, Eq. (3.37)]; in particular, $\varepsilon = \varepsilon_{j,k}^0 \varepsilon_{j,k}^1 \varepsilon_{j,k}^2$ and so $|\varepsilon_{j,k}^0| + |\varepsilon_{j,k}^1| + |\varepsilon_{j,k}^2| = |\varepsilon|$. Define

$$Q_\varepsilon^\pm(\mathbf{v}) = n_\pm(\varepsilon)v_\varepsilon + 2 \sum_{1 \leq j < k \leq n} \pm v_{\varepsilon_{j,k}^0 \varepsilon_{j,k}^2} v_{\varepsilon_{j,k}^1}, \quad (3.1)$$

where the \pm signs inside the sum depend on ε, j, k . For $s, t \in \mathbb{R}$, define

$$Q_\varepsilon^{s,t} = \left(s - \frac{t}{2}\right) Q_\varepsilon^+ + \frac{t}{2} Q_\varepsilon^-. \quad (3.2)$$

Thus, except when $(s, t) = (0, 0)$, $Q_\varepsilon^{s,t}$ is a homogeneous trace degree $|\varepsilon|$ polynomial.

Additionally, let $\delta \in \mathcal{E}$. For $1 \leq j \leq |\varepsilon|$ and $1 \leq k \leq |\delta|$, let $\varepsilon^{(j)}$ and $\delta^{(k)}$ be the cyclic permutations of ε and δ in [16, Eq. (3.40)]. Define

$$R_{\varepsilon,\delta}^\pm(\mathbf{v}) = \sum_{j=1}^{|\varepsilon|} \sum_{k=1}^{|\delta|} \pm v_{\varepsilon^{(j)}\delta^{(k)}}, \quad (3.3)$$

where the \pm signs inside the sum depend on $\varepsilon, \delta, j, k$. For $s, t \in \mathbb{R}$, define

$$R_{\varepsilon,\delta}^{s,t} = \left(s - \frac{t}{2}\right) R_{\varepsilon,\delta}^+ + \frac{t}{2} R_{\varepsilon,\delta}^-. \quad (3.4)$$

Thus, except when $(s, t) = (0, 0)$, $R_{\varepsilon,\delta}^{s,t}$ is a homogeneous trace degree $|\varepsilon| + |\delta|$ polynomial.

The following **intertwining formulas** were the core computational tools in [16].

Theorem 3.6 (Intertwining Formulas). [16, Theorems 1.20 & 3.13] Let $s, t \in \mathbb{R}$. Define the following differential operators on \mathcal{P} :

$$\mathcal{D}_{s,t} = \frac{1}{2} \sum_{\varepsilon \in \mathcal{E}} Q_\varepsilon^{s,t}(\mathbf{v}) \frac{\partial}{\partial v_\varepsilon} \quad \text{and} \quad \mathcal{L}_{s,t} = \frac{1}{2} \sum_{\varepsilon, \delta \in \mathcal{E}} R_{\varepsilon,\delta}^{s,t}(\mathbf{v}) \frac{\partial^2}{\partial v_\varepsilon \partial v_\delta}, \quad (3.5)$$

where $Q_\varepsilon^{s,t}$ and $R_{\varepsilon,\delta}^{s,t}$ are as in Definition 3.5. Then for any $P \in \mathcal{P}$, we have

$$\frac{1}{2}A_{s,t}^N(P \circ \mathbf{V}_N) = - \left[\mathcal{D}_{s,t}P + \frac{1}{N^2}\mathcal{L}_{s,t}P \right] \circ \mathbf{V}_N. \quad (3.6)$$

In the special case $(s, t) = (1, 0)$,

$$\mathcal{D}_{1,0}|_{\mathcal{H}\mathcal{P}} = \frac{1}{2} \sum_{|k| \geq 1} |k| v_k \frac{\partial}{\partial v_k} + \frac{1}{2} \sum_{k=2}^{\infty} k \left[\left(\sum_{j=1}^{k-1} v_j v_{k-j} \right) \frac{\partial}{\partial v_k} + \left(\sum_{j=1}^{k-1} v_{-j} v_{-(k-j)} \right) \frac{\partial}{\partial v_{-k}} \right], \quad (3.7)$$

$$\mathcal{L}_{1,0}|_{\mathcal{H}\mathcal{P}} = \frac{1}{2} \sum_{|j|, |k| \geq 1} jk v_{j+k} \frac{\partial^2}{\partial v_j \partial v_k}. \quad (3.8)$$

Notation 3.7. For $N \geq 1$, we set

$$\mathcal{D}_{s,t}^N = \mathcal{D}_{s,t} + \frac{1}{N^2}\mathcal{L}_{s,t}. \quad (3.9)$$

Thus (3.6) asserts that $\frac{1}{2}A_{s,t}^N(P \circ \mathbf{V}_N) = -[\mathcal{D}_{s,t}^N P] \circ \mathbf{V}_N$.

Remark 3.8. (1) In the notation of [16, Definition 1.16], $\mathcal{D}_{1,0}|_{\mathcal{H}\mathcal{P}} = \mathcal{N}_0 + 2\mathcal{Z}$ (rewritten here using the trick of Remark 5.13 in that paper). Note, also, that the terms with $j = -k$ in (3.8) involve v_0 , which we interpret as 1.

(2) Since $\Delta_{\mathbb{U}_N} = A_{1,0}^N|_{\mathbb{U}_N}$, (3.6) shows that

$$\frac{1}{2}\Delta_{\mathbb{U}_N}(P \circ \mathbf{V}_N) = - \left[\mathcal{D}_{1,0}P + \frac{1}{N^2}\mathcal{L}_{1,0}P \right] \circ \mathbf{V}_N = -[\mathcal{D}_{1,0}^N P] \circ \mathbf{V}_N. \quad (3.10)$$

This is the formal sense in which (1.14) is true. For a trace polynomial $(P \circ \mathbf{V}_N)|_{\mathbb{U}_N}$ with $P \in \mathcal{H}\mathcal{P}$, the Laplacian can be calculated explicitly using (3.7) and (3.8).

Example 3.9. Consider the trace polynomials $f(U) = \text{tr}(U^n)\text{tr}(U^m)$ for $U \in \mathbb{U}_N$; for convenience we assume $n, m \geq 2$. Then $f = P \circ \mathbf{V}_N$ where $P(\mathbf{v}) = v_n v_m \in \mathcal{H}\mathcal{P}_+$. Then (3.7) and (3.8) give

$$2\mathcal{D}_{1,0}(v_n v_m) = (n+m)v_n v_m + n \sum_{j=1}^{n-1} v_j v_{n-j} v_m + m \sum_{j=1}^{m-1} v_j v_{m-j} v_n, \quad (3.11)$$

$$2\mathcal{L}_{1,0}(v_n v_m) = 2nmv_{n+m}. \quad (3.12)$$

Note that all terms have homogeneous trace degree $n+m$, the same as $v_n v_m$; this follows from Theorem 3.6. Thus, (3.10) yields

$$\begin{aligned} \Delta_{\mathbb{U}_N}(\text{tr}(U^n)\text{tr}(U^m)) &= -(n+m)\text{tr}(U^n)\text{tr}(U^m) - \frac{2nm}{N^2}\text{tr}(U^{n+m}) \\ &\quad - n \sum_{j=1}^{n-1} \text{tr}(U^j)\text{tr}(U^{n-j})\text{tr}(U^m) - m \sum_{j=1}^{m-1} \text{tr}(U^j)\text{tr}(U^{m-j})\text{tr}(U^n). \end{aligned}$$

In the special case $N = 1$, $\text{tr}(U^j) = U^j$, and so the calculation shows that

$$\Delta_{\mathbb{U}_1}(U^{n+m}) = -(n+m)U^{n+m} - 2nmU^{n+m} - n \sum_{j=1}^{n-1} U^{n+m} - m \sum_{j=1}^m U^{n+m} = -(n+m)^2 U^{n+m},$$

which is consistent with (2.10).

We record here another intertwining formula (that did not appear in [16]) regarding the complex conjugation map.

Definition 3.10. Given $\varepsilon \in \mathcal{E}$, define $\varepsilon^* \in \mathcal{E}$ by $(\varepsilon_1, \dots, \varepsilon_n)^* = (\varepsilon_n^*, \dots, \varepsilon_1^*)$, where $(\pm 1)^* = \pm 1$ and $(\pm i)^* = \mp i$. Let $\mathcal{C}: \mathcal{P} \rightarrow \mathcal{P}$ be the conjugate linear homomorphism defined by $\mathcal{C}(v_\varepsilon) = v_{\varepsilon^*}$ for all $\varepsilon \in \mathcal{E}$. Note that, for any $P \in \mathcal{P}$ and $Z \in \mathbb{Z}_N$,

$$\overline{P \circ \mathbf{V}_N(Z)} = (\mathcal{C}P \circ \mathbf{V}_N)(Z). \quad (3.13)$$

That is: \mathcal{C} intertwines complex conjugation. This follows from the fact that $\overline{\text{tr}(Z^\varepsilon)} = \text{tr}(Z^{\varepsilon^*})$. We will often write $\mathcal{C}(P) = P^*$.

Lemma 3.11. The complex conjugation intertwiner \mathcal{C} commutes with the operators $\mathcal{D}_{s,t}$, $\mathcal{L}_{s,t}$, and hence $\mathcal{D}_{s,t}^N$.

Proof. Fix $N \in \mathbb{N}$ and let $P \in \mathcal{P}$ and $Z \in \mathbb{M}_N$. From Remark 2.5(3) and (3.13), together with (3.6) and (3.9), we have

$$(e^{-\mathcal{D}_{s,t}^N} \mathcal{C}P) \circ \mathbf{V}_N(Z) = (e^{\frac{1}{2}A_{s,t}^N} (\mathcal{C}P \circ \mathbf{V}_N))(Z) = \overline{(e^{\frac{1}{2}A_{s,t}^N} (P \circ \mathbf{V}_N))(Z)} = \mathcal{C}(e^{-\mathcal{D}_{s,t}^N} P) \circ \mathbf{V}_N(Z).$$

This shows that

$$([\mathcal{C}, e^{-\mathcal{D}_{s,t}^N}]P)(Z) = 0, \quad N \in \mathbb{N}, Z \in \mathbb{GL}_N.$$

It follows from [16, Theorem 2.10] (asymptotic uniqueness of trace polynomial representations) that the polynomial $[\mathcal{C}, e^{-\mathcal{D}_{s,t}^N}]P = 0$. Scaling $(s, t) \mapsto (\alpha s, \alpha t)$ and differentiating with respect to α at $\alpha = 0$ shows that $[\mathcal{C}, \mathcal{D}_{s,t}^N]P = 0$. As this holds for each N , sending $N \rightarrow \infty$ (using continuity of all involved maps on the finite dimensional $\mathcal{D}_{s,t}^N$ -invariant subspace of polynomials with trace degree $\leq \deg(P)$) shows that $[\mathcal{C}, \mathcal{D}_{s,t}]P = 0$, and it then follows that $[\mathcal{C}, \mathcal{L}_{s,t}]P = 0$. Since these hold for all $P \in \mathcal{P}$, the lemma is proved. \square

Remark 3.12. It is possible to prove Lemma 3.11 with direct computation from the definitions (3.5) of the intertwining operators $\mathcal{D}_{s,t}$ and $\mathcal{L}_{s,t}$; the proof we've given is much shorter.

As noted in Example 3.9, the operators $\mathcal{D}_{s,t}$ and $\mathcal{L}_{s,t}$ in Theorem 3.6 preserve trace degree (so long as $(s, t) \neq (0, 0)$). Hence, so do the operators $\mathcal{D}_{s,t}^N$ which intertwine $-\frac{1}{2}A_{s,t}^N$. In particular, this means that, for each $n \in \mathbb{N}$, \mathcal{P}_n is an invariant subspace for $\mathcal{D}_{s,t}^N$; equivalently, by (3.6), the finite dimensional subspace $\mathcal{P}_n \circ \mathbf{V}^N$ of trace polynomials “of trace degree $\leq n$ ” is an invariant subspace for $A_{s,t}^N$. (Note: from the second term in P in Example 3.3, we see that trace degree is not well-defined for trace polynomial functions, only for their intertwining polynomials. However, the subspace $\mathcal{P}_n \circ \mathbf{V}_N$ is a well-defined, finite dimensional invariant subspace for $A_{s,t}^N$.)

Let $n \in \mathbb{N}$. The restriction $\mathcal{D}_{s,t}^N|_{\mathcal{P}_n}$ is a finite dimensional operator, and so can be exponentiated in the usual manner. Similar considerations applied to $A_{s,t}^N|_{\mathcal{P}_n \circ \mathbf{V}_N}$, together with (3.6), show that

$$e^{\frac{1}{2}A_{s,t}^N} (P \circ \mathbf{V}_N) = (e^{-\mathcal{D}_{s,t}^N} P) \circ \mathbf{V}_N, \quad P \in \mathcal{P}, \quad (3.14)$$

where the restrictions are done with $n = \deg(P)$. Combining this with (2.8) shows that, for $s, t > 0$ with $s > t/2$,

$$\mathbb{E}_{\mu_{s,t}^N} (P \circ \mathbf{V}_N) = (e^{-\mathcal{D}_{s,t}^N} P)(\mathbf{1}), \quad (3.15)$$

where by $P(\mathbf{1})$ we mean the complex number given by setting all $v_\varepsilon = 1$ in $P(\mathbf{v})$. Analogous considerations from (2.7) and (3.10) show that, for $t > 0$,

$$\mathbb{E}_{\rho_t^N} (P \circ \mathbf{V}_N) = (e^{-\mathcal{D}_{t,0}^N} P)(\mathbf{1}). \quad (3.16)$$

3.2 Concentration of Heat Kernel Measure

The expectation-intertwining formulas (3.16) and (3.15) show there is $O(1/N^2)$ -concentration of the \mathbb{U}_N or \mathbb{GL}_N heat kernel measure's mass. The following lemma makes this precise. It is a version of [16, Lemma 4.1]; we expand on the statement and proof here to give some quantitative estimates (cf. Proposition 3.18).

Lemma 3.13. *Let V be a finite dimensional normed \mathbb{C} -space. For parameters $s, t \in \mathbb{R}$, let $D_{s,t}$ and $L_{s,t}$ be two operators on V that depend continuously on s and t . Then there exists a constant $C(s, t) < \infty$, depending continuously on $(s, t) \in \mathbb{R}^2$, such that, for any linear functional $\psi: V \rightarrow \mathbb{C}$,*

$$|\psi(e^{D_{s,t} + \epsilon L_{s,t}} x) - \psi(e^{D_{s,t}} x)| \leq C(s, t) \|x\|_V \|\psi\|_{V^*} |\epsilon|, \quad x \in V, |\epsilon| \leq 1. \quad (3.17)$$

Note that the constant $C(s, t)$ also depends on the norm $\|\cdot\|_V$.

Proof. We follow our proof in [16, Lemma 4.1]. For the moment, write $D = D_{s,t}$ and $L = L_{s,t}$. Using the well known differential of the exponential map (see for example [17, Theorem 1.5.3, p. 23] or [25, Theorem 3.5, p. 70]),

$$\begin{aligned} \frac{d}{du} e^{D+uL} &= e^{D+uL} \int_0^1 e^{-v(D+uL)} L e^{v(D+uL)} dv \\ &= \int_0^1 e^{(1-v)(D+uL)} L e^{v(D+uL)} dv, \end{aligned}$$

we may write

$$e^{D+\epsilon L} - e^D = \int_0^\epsilon \frac{d}{du} e^{D+uL} du = \int_0^\epsilon \left[\int_0^1 e^{(1-v)(D+uL)} L e^{v(D+uL)} dv \right] du.$$

Crude bounds now show

$$\|e^{D+\epsilon L} - e^D\|_{\text{End}(V)} \leq \int_0^{|\epsilon|} \left[\int_0^1 \left\| e^{(1-v)(D+uL)} L e^{v(D+uL)} \right\|_{\text{End}(V)} dv \right] du, \quad (3.18)$$

where $\|\cdot\|_{\text{End}(V)}$ is the operator norm induced by $\|\cdot\|_V$. Now, set

$$C(s, t) = \sup_{\substack{0 \leq u \leq |\epsilon| \\ 0 \leq v \leq 1}} \left\| e^{(1-v)(D_{s,t} + uL_{s,t})} L_{s,t} e^{v(D_{s,t} + uL_{s,t})} \right\|_{\text{End}(V)}. \quad (3.19)$$

(This constant nominally depends on ϵ , but we can take $\epsilon = 1$ here to provide a uniform bound.) The function $(u, v, s, t) \mapsto e^{(1-v)(D_{s,t} + uL_{s,t})} L_{s,t} e^{v(D_{s,t} + uL_{s,t})}$ is continuous, and hence $C(s, t)$ is a continuous in (s, t) . Equations (3.18) and (3.19) show that

$$\|e^{D+\epsilon L} - e^D\|_{\text{End}(V)} \leq C(s, t) |\epsilon|; \quad (3.20)$$

and (3.17) follows immediately from (3.20). \square

Since $\psi(P) = P(\mathbf{1})$ defines a linear functional on \mathcal{P}_n for each n , (3.16), (3.15), and Lemma 3.13 immediately yield the following.

Corollary 3.14. *For $s, t \in \mathbb{R}$ and $P \in \mathcal{P}$, there is a constant $C(s, t, P) < \infty$, continuous in $(s, t) \in \mathbb{R}^2$, so that*

$$\left| (e^{-\mathcal{D}_{s,t}^N} P)(\mathbf{1}) - (e^{-\mathcal{D}_{s,t}} P)(\mathbf{1}) \right| \leq \frac{1}{N^2} \cdot C(s, t, P). \quad (3.21)$$

Proof. Let $n = \deg P$, and choose any norm $\|\cdot\|_{\mathcal{P}_n}$ on the finite dimensional space \mathcal{P}_n ; then $C(s, t, P)$ can be taken to equal $C(s, t) \|\psi\|_{\mathcal{P}_n}^* \|P\|_{\mathcal{P}_n}$ where $\psi(P) = P(\mathbf{1})$ and the constant $C(s, t)$ is from (3.19) with the operators $D_{s,t} = -\mathcal{D}_{s,t}|_{\mathcal{P}_n}$ and $L_{s,t} = -\mathcal{L}_{s,t}|_{\mathcal{P}_n}$. \square

Corollary 3.14 (in the special case $(s, t) \mapsto (t, 0)$) shows that the large- N limit of the heat kernel expectation $\mathbb{E}_{\rho_t^N}$ of any trace polynomial is given by the flow operator $e^{-\mathcal{D}_{t,0}}$; in this sense, $\mathcal{D}_{1,0}$ is the generator of the limit heat kernel (and hence of the free unitary Brownian motion). In particular, taking $P = v_n$ so that $(P \circ \mathbf{V}_N)(U) = \text{tr}(U^n)$, (3.16) and (3.21) show that

$$(e^{-\mathcal{D}_{t,0}} v_k)(\mathbf{1}) = \lim_{N \rightarrow \infty} \int_{\mathbb{U}_N} \text{tr}(U^n) \rho_t^N(dU) = \nu_k(t) \quad (3.22)$$

are the moments of ν_t ; cf. Definition 1.1. Since $\mathcal{D}_{t,0}$ is a first-order differential operator, the operator $e^{-\mathcal{D}_{t,0}}$ is an algebra homomorphism, and since the evaluation-at- $\mathbf{1}$ -map is also a homomorphism, the complete description of the semigroup acting on $\mathcal{H}\mathcal{P}$ is given by

$$\left(e^{-\mathcal{D}_{t,0}} (v_1^{k_1} v_{-1}^{k_{-1}} \cdots v_n^{k_n} v_{-n}^{k_{-n}}) \right) (\mathbf{1}) = \nu_1(t)^{k_1} \nu_{-1}(t)^{k_{-1}} \cdots \nu_n(t)^{k_n} \nu_{-n}(t)^{k_{-n}}. \quad (3.23)$$

This simplifies further, since $\nu_{-m}(t) = \nu_m(t)$ for all m .

3.3 Estimates on the Constants $C(s, t, P)$

Corollary 3.14 suffices to prove weak a.s. convergence of distributions when using (Laurent) polynomial test functions; in particular, this will suffice to prove Theorem 1.7. To extend this convergence to a larger class of test functions, as in Theorems 1.2–1.5, we will need some quantitative information about the constants $C(s, t, P)$ in (3.21). To prove such estimates, we begin by introducing a norm on \mathcal{P} that will be used throughout the remainder of this section.

Definition 3.15. Let $\|\cdot\|_1$ denote the ℓ^1 -norm on \mathcal{P} . Precisely: let $\mathbb{N}_0^\mathcal{E}$ denote the set of functions $\mathbf{k}: \mathcal{E} \rightarrow \mathbb{N}$ that are finitely-supported. For $\mathbf{k} \in \mathbb{N}_0^\mathcal{E}$, define $\mathbf{v}^{\mathbf{k}}$ to be the monomial

$$\mathbf{v}^{\mathbf{k}} = \prod_{\varepsilon \in \text{supp } \mathbf{k}} v_\varepsilon^{\mathbf{k}(\varepsilon)}.$$

Any $P \in \mathcal{P}$ has a unique representation of the form

$$P(\mathbf{v}) = \sum_{\mathbf{k} \in \mathbb{N}_0^\mathcal{E}} a_{\mathbf{k}} \mathbf{v}^{\mathbf{k}} \quad (3.24)$$

for some coefficients $a_{\mathbf{k}} \in \mathbb{C}$ that are 0 for all but finitely-many \mathbf{k} . Then we define

$$\|P\|_1 = \sum_{\mathbf{k} \in \mathbb{N}_0^\mathcal{E}} |a_{\mathbf{k}}|. \quad (3.25)$$

The uniqueness of the representation (3.24) of P shows that $\|\cdot\|_1$ is well-defined on \mathcal{P} , and it is easily verified to be a norm.

We will use the norm $\|\cdot\|_1$ of (3.25) to provide concrete bounds on $C(s, t, P)$ for $P \in \mathcal{P}_n$; this will suffice to prove Theorems 1.3 and 1.5 (as well as a weaker version of Theorem 1.2, with ultra-analytic test functions). We remind the reader of the following lemma: the operator norm on matrices induced by the ℓ^1 -norm on vectors is bounded by the maximal column sum of the matrix.

Lemma 3.16. *Let V be a finite dimensional vector space, and let $e_1, \dots, e_d \in V$ be a basis. Let $\|\cdot\|_1$ denote the norm $\|a_1e_1 + \dots + a_de_d\|_1 = |a_1| + \dots + |a_d|$ on V . Then for $A \in \text{End}(V)$, the operator norm $\|A\|_{1 \rightarrow 1} = \sup_{\|w\|_1=1} \|Aw\|_1$ is bounded by*

$$\|A\|_{1 \rightarrow 1} \leq \max_{1 \leq j \leq d} \|A(e_j)\|_1. \quad (3.26)$$

Proof. Letting $w = a_1e_1 + \dots + a_de_d$, compute

$$\|Aw\|_1 = \|a_1A(e_1) + \dots + a_dA(e_d)\|_1 \leq \sum_{k=1}^d |a_k| \|A(e_k)\|_1 \leq \max_{1 \leq j \leq d} \|A(e_j)\|_1 \sum_{k=1}^d |a_k|,$$

and since $\sum_{k=1}^d |a_k| = \|w\|_1$, this proves the result. \square

Remark 3.17. If we represent a vector in V in a non-unique way, for example $v = a_1e_1 + a_2e_2 + b_1e_1 = (a_1 + b_1)e_1 + a_2e_2$, note that $\|v\|_1 = |a_1 + b_1| + |a_2| \leq |a_1| + |b_1| + |a_2|$; thus, if we use such a redundant representation for a vector when ‘‘computing’’ the $\|\cdot\|_1$ -norm, we will always get an upper bound. This will be relevant in the proof of Proposition 3.18 below, where it will be challenging to detect repeated occurrences of basis vectors.

We now prove a quantitative bound for the constants $C(s, t, P)$ for any $P \in \mathcal{P}$.

Proposition 3.18. *Let $s, t \in \mathbb{R}$, let $n \in \mathbb{N}$, and let $P \in \mathcal{P}_n$. Define $r = |s - \frac{t}{2}| + \frac{1}{2}|t|$. Then for all $N \geq 1$,*

$$\left| (e^{-\mathcal{D}_{s,t}^N} P)(\mathbf{1}) - (e^{-\mathcal{D}_{s,t}} P)(\mathbf{1}) \right| \leq \frac{1}{N^2} \cdot \frac{r}{2} n^2 e^{\frac{r}{2} n^2 \left(1 + \frac{1}{N^2}\right)} \|P\|_1. \quad (3.27)$$

Proof. Let $V = \mathcal{P}_n$ equipped with the norm $\|\cdot\|_1$ of (3.25), let $\psi(P) = P(\mathbf{1})$, and set $D = -\mathcal{D}_{s,t}$ and $L = -\mathcal{L}_{s,t}$. Then Lemma 3.13 shows that

$$\left| (e^{-\mathcal{D}_{s,t}^N} P)(\mathbf{1}) - (e^{-\mathcal{D}_{s,t}} P)(\mathbf{1}) \right| = \left| \psi(e^{D + \frac{1}{N^2}L} P) - \psi(e^D P) \right| \leq \frac{1}{N^2} C \|\psi\|_1^* \|P\|_1, \quad (3.28)$$

where

$$C = \sup_{\substack{0 \leq u \leq 1/N^2 \\ 0 \leq v \leq 1}} \left\| e^{(1-v)(D+uL)} L e^{v(D+uL)} \right\|_{1 \rightarrow 1}. \quad (3.29)$$

Note that, for $P(\mathbf{v}) = \sum_{\mathbf{k}} a_{\mathbf{k}} \mathbf{v}^{\mathbf{k}}$ as in (3.24),

$$|\psi(P)| = |P(\mathbf{1})| = \left| \sum_{\mathbf{k}} a_{\mathbf{k}} \right| \leq \|P\|_1, \quad \text{and therefore} \quad \|\psi\|_1^* \leq 1. \quad (3.30)$$

Hence, to prove the proposition, it suffices to show that (3.29) is bounded by $\frac{r}{2} n^2 e^{\frac{r}{2} n^2 (1 + 1/N^2)}$.

Since the operator norm $\|\cdot\|_{1 \rightarrow 1}$ is submultiplicative, for $0 \leq u, v \leq 1$ we can estimate

$$\begin{aligned} \left\| e^{(1-v)(D+uL)} L e^{v(D+uL)} \right\|_{1 \rightarrow 1} &\leq \left\| e^{(1-v)(D+uL)} \right\|_{1 \rightarrow 1} \cdot \left\| e^{v(D+uL)} \right\|_{1 \rightarrow 1} \cdot \|L\|_{1 \rightarrow 1} \\ &\leq e^{(1-v)(\|D\|_{1 \rightarrow 1} + u\|L\|_{1 \rightarrow 1})} \cdot e^{v(\|D\|_{1 \rightarrow 1} + u\|L\|_{1 \rightarrow 1})} \cdot \|L\|_{1 \rightarrow 1} \\ &= e^{\|D\|_{1 \rightarrow 1}} \cdot e^{u\|L\|_{1 \rightarrow 1}} \cdot \|L\|_{1 \rightarrow 1} \end{aligned}$$

where the second line follows from expanding the power series of the exponentials and repeatedly using the triangle inequality and submultiplicativity of the norm $\|\cdot\|_{1 \rightarrow 1}$. Hence, taking the supremum over $0 \leq u \leq 1/N^2$, we have

$$C \leq e^{\|D\|_{1 \rightarrow 1}} \cdot e^{\frac{1}{N^2} \|L\|_{1 \rightarrow 1}} \cdot \|L\|_{1 \rightarrow 1}. \quad (3.31)$$

It behooves us to estimate $\|L\|_{1 \rightarrow 1}$ and $\|D\|_{1 \rightarrow 1}$; we do this using Lemma 3.16.

The basis of \mathcal{P}_n defining the norm $\|\cdot\|_1$ is the set of monomials in \mathcal{P}_n ; that is, using the notation of Definition 3.15, the basis is

$$\mathcal{B}_n = \{\mathbf{v}^{\mathbf{k}} : \deg(\mathbf{v}^{\mathbf{k}}) \leq n\} = \{\mathbf{v}^{\mathbf{k}} : \sum_{\varepsilon \in \mathcal{E}} |\mathbf{k}(\varepsilon)| |\varepsilon| \leq n\}.$$

We must therefore estimate the $\|\cdot\|_1$ -norm of the images of $D = -\mathcal{D}_{s,t}$ and $L = -\mathcal{L}_{s,t}$ on these basis vectors. So, fix a finitely-supported function $\mathbf{k} : \mathcal{E} \rightarrow \mathbb{N}$. Then for any $\varepsilon \in \mathcal{E}$, we have

$$\frac{\partial}{\partial v_\varepsilon} \mathbf{v}^{\mathbf{k}} = \mathbf{k}(\varepsilon) \frac{\mathbf{v}^{\mathbf{k}}}{v_\varepsilon}, \quad \text{where} \quad \frac{\mathbf{v}^{\mathbf{k}}}{v_\varepsilon} \in \mathcal{B}_n.$$

(I.e. we write $\frac{\partial}{\partial v} v^k = kv^{k-1} = kv^k/v$ to simplify notation.) Thus, from (3.1), we have

$$\sum_{\varepsilon \in \mathcal{E}} Q_\varepsilon^\pm(\mathbf{v}) \frac{\partial}{\partial v_\varepsilon} \mathbf{v}^{\mathbf{k}} = \sum_{\varepsilon \in \mathcal{E}} \mathbf{k}(\varepsilon) \left[n_\pm(\varepsilon) \mathbf{v}^{\mathbf{k}} + 2 \sum_{1 \leq j < k \leq |\varepsilon|} \pm \frac{v_{\varepsilon_{j,k}^0} v_{\varepsilon_{j,k}^2} v_{\varepsilon_{j,k}^1}}{v_\varepsilon} \mathbf{v}^{\mathbf{k}} \right].$$

(This is a finite sum: $\mathbf{k}(\varepsilon) = 0$ for all but finitely-many $\varepsilon \in \mathcal{E}$.) Thus, from (3.2) and (3.5), we have

$$\begin{aligned} \mathcal{D}_{s,t}(\mathbf{v}^{\mathbf{k}}) &= \frac{1}{2} \sum_{\varepsilon \in \mathcal{E}} \mathbf{k}(\varepsilon) \left[\left(s - \frac{t}{2} \right) n_+(\varepsilon) + \frac{t}{2} n_-(\varepsilon) \right] \cdot \mathbf{v}^{\mathbf{k}} \\ &\quad + \sum_{\varepsilon \in \mathcal{E}} \mathbf{k}(\varepsilon) \sum_{1 \leq j < k \leq |\varepsilon|} \left[\left(s - \frac{t}{2} \right) (\pm 1) + \frac{t}{2} (\pm 1) \right] \frac{v_{\varepsilon_{j,k}^0} v_{\varepsilon_{j,k}^2} v_{\varepsilon_{j,k}^1}}{v_\varepsilon} \mathbf{v}^{\mathbf{k}}. \end{aligned}$$

All of the vectors $\mathbf{v}^{\mathbf{k}}$ and $v_{\varepsilon_{j,k}^0} v_{\varepsilon_{j,k}^2} v_{\varepsilon_{j,k}^1} \mathbf{v}^{\mathbf{k}}/v_\varepsilon$ in the above sum are basis vectors in \mathcal{B}_n . They may not be distinct, but by Remark 3.17 we can compute an upper bound for the norm by simply summing the absolute values of the coefficients:

$$\begin{aligned} \|\mathcal{D}_{s,t}(\mathbf{v}^{\mathbf{k}})\|_1 &\leq \frac{1}{2} \sum_{\varepsilon \in \mathcal{E}} \mathbf{k}(\varepsilon) \left| \left(s - \frac{t}{2} \right) n_+(\varepsilon) + \frac{t}{2} n_-(\varepsilon) \right| \\ &\quad + \sum_{\varepsilon \in \mathcal{E}} \mathbf{k}(\varepsilon) \sum_{1 \leq j < k \leq |\varepsilon|} \left| \left(s - \frac{t}{2} \right) (\pm 1) + \frac{t}{2} (\pm 1) \right|. \end{aligned}$$

We can estimate the internal terms as follows: since $|n_\pm(\varepsilon)| \leq |\varepsilon|$ (cf. Definition 3.5),

$$\left| \left(s - \frac{t}{2} \right) n_+(\varepsilon) + \frac{t}{2} n_-(\varepsilon) \right| \leq \left| s - \frac{t}{2} \right| |n_+(\varepsilon)| + \frac{1}{2} |t| |n_-(\varepsilon)| \leq r |\varepsilon|$$

and similarly the term inside the double sum is $\leq r$. Hence, we have

$$\|\mathcal{D}_{s,t}(\mathbf{v}^{\mathbf{k}})\|_1 \leq \frac{r}{2} \sum_{\varepsilon \in \mathcal{E}} |\varepsilon| \mathbf{k}(\varepsilon) + r \sum_{\varepsilon \in \mathcal{E}} \mathbf{k}(\varepsilon) \frac{|\varepsilon|(|\varepsilon| - 1)}{2}. \quad (3.32)$$

Since $\mathbf{v}^{\mathbf{k}} \in \mathcal{B}_n$, we have $\sum_{\varepsilon \in \mathcal{E}} |\varepsilon| \mathbf{k}(\varepsilon) \leq n$, and so too $|\varepsilon| \leq n$ for any nonzero term in the sum. Thus, (3.32) yields

$$\|D\|_{1 \rightarrow 1} \leq \max_{\mathbf{v}^{\mathbf{k}} \in \mathcal{B}_n} \|\mathcal{D}_{s,t}(\mathbf{v}^{\mathbf{k}})\|_1 \leq \frac{r}{2} n + \frac{r}{2} (n-1)n = \frac{r}{2} n^2. \quad (3.33)$$

Turning now to $L = -\mathcal{L}_{s,t}$, we have

$$\frac{\partial^2}{\partial v_\varepsilon \partial v_\delta} \mathbf{v}^{\mathbf{k}} = \begin{cases} \mathbf{k}(\varepsilon)\mathbf{k}(\delta)\mathbf{v}^{\mathbf{k}}/v_\varepsilon v_\delta, & \varepsilon \neq \delta, \\ \mathbf{k}(\varepsilon)(\mathbf{k}(\varepsilon) - 1)\mathbf{v}^{\mathbf{k}}/v_\varepsilon^2, & \varepsilon = \delta. \end{cases}$$

Thus, from (3.3) we have

$$\sum_{\varepsilon, \delta \in \mathcal{E}} R_{\varepsilon, \delta}^\pm(\mathbf{v}) \frac{\partial^2}{\partial v_\varepsilon \partial v_\delta} \mathbf{v}^{\mathbf{k}} = \sum_{\varepsilon \in \mathcal{E}} \mathbf{k}(\varepsilon)(\mathbf{k}(\varepsilon) - 1) \sum_{j,k=1}^{|\varepsilon|} \pm v_{\varepsilon(j)\varepsilon(k)} \frac{\mathbf{v}^{\mathbf{k}}}{v_\varepsilon^2} + \sum_{\varepsilon \neq \delta \in \mathcal{E}} \mathbf{k}(\varepsilon)\mathbf{k}(\delta) \sum_{j=1}^{|\varepsilon|} \sum_{k=1}^{|\delta|} \pm v_{\varepsilon(j)\delta(k)} \frac{\mathbf{v}^{\mathbf{k}}}{v_\varepsilon v_\delta}$$

and so, from (3.4) and (3.5),

$$\begin{aligned} \mathcal{L}_{s,t}(\mathbf{v}^{\mathbf{k}}) &= \frac{1}{2} \sum_{\varepsilon \in \mathcal{E}} \mathbf{k}(\varepsilon)(\mathbf{k}(\varepsilon) - 1) \sum_{j,k=1}^{|\varepsilon|} \left[\left(s - \frac{t}{2} \right) (\pm) + \frac{t}{2} (\pm) \right] v_{\varepsilon(j)\varepsilon(k)} \frac{\mathbf{v}^{\mathbf{k}}}{v_\varepsilon^2} \\ &\quad + \frac{1}{2} \sum_{\varepsilon \neq \delta \in \mathcal{E}} \mathbf{k}(\varepsilon)\mathbf{k}(\delta) \sum_{j=1}^{|\varepsilon|} \sum_{k=1}^{|\delta|} \left[\left(s - \frac{t}{2} \right) (\pm) + \frac{t}{2} (\pm) \right] v_{\varepsilon(j)\delta(k)} \frac{\mathbf{v}^{\mathbf{k}}}{v_\varepsilon v_\delta}. \end{aligned}$$

As above, it follows that

$$\begin{aligned} \|\mathcal{L}_{s,t}(\mathbf{v}^{\mathbf{k}})\|_1 &\leq \frac{r}{2} \sum_{\varepsilon \in \mathcal{E}} \mathbf{k}(\varepsilon)(\mathbf{k}(\varepsilon) - 1) \cdot |\varepsilon|^2 + \frac{r}{2} \sum_{\varepsilon \neq \delta \in \mathcal{E}} \mathbf{k}(\varepsilon)\mathbf{k}(\delta) |\varepsilon| |\delta| \\ &\leq \frac{r}{2} \sum_{\varepsilon \in \mathcal{E}} \mathbf{k}(\varepsilon)^2 |\varepsilon|^2 + \frac{r}{2} \sum_{\varepsilon \neq \delta \in \mathcal{E}} \mathbf{k}(\varepsilon)\mathbf{k}(\delta) |\varepsilon| |\delta| \\ &= \frac{r}{2} \sum_{\varepsilon, \delta \in \mathcal{E}} \mathbf{k}(\varepsilon)\mathbf{k}(\delta) |\varepsilon| |\delta| = \frac{r}{2} \left(\sum_{\varepsilon \in \mathcal{E}} \mathbf{k}(\varepsilon) |\varepsilon| \right)^2 \leq \frac{r}{2} n^2. \end{aligned} \quad (3.34)$$

As this holds for all $\mathbf{v} \in \mathcal{B}_n$, we therefore have

$$\|L\|_{1 \rightarrow 1} = \max_{\mathbf{v}^{\mathbf{k}} \in \mathcal{B}_n} \|\mathcal{L}_{s,t}(\mathbf{v}^{\mathbf{k}})\|_1 \leq \frac{r}{2} n^2. \quad (3.35)$$

Combining (3.31) with (3.33) and (3.35) proves the result. \square

When $s, t > 0$ and $s > t/2$, $r = (s - \frac{t}{2}) + \frac{t}{2} = s$. Proposition 3.18 then shows that the constant $C(s, t, P)$ in Corollary 3.14 can be bounded by

$$C(s, t, P) \leq \frac{s}{2} (\deg(P))^2 e^{s(\deg(P))^2} \|P\|_1, \quad P \in \mathcal{P} \quad (3.36)$$

by using $1/N^2 \leq 1$. We can do better than this if we take N sufficiently large.

Corollary 3.19. *Let $s, t \in \mathbb{R}$, and set $r = |s - \frac{t}{2}| + \frac{1}{2}|t|$. Let $\delta > 0$, $n, N \in \mathbb{N}$, and $P \in \mathcal{P}_n$. If $N > \sqrt{2/\delta}$, then*

$$\left| (e^{-\mathcal{D}_{s,t}^N} P)(\mathbf{1}) - (e^{-\mathcal{D}_{s,t}} P)(\mathbf{1}) \right| \leq \frac{1}{N^2} \cdot \frac{1}{\delta} e^{\frac{r}{2}(1+\delta)n^2} \|P\|_1. \quad (3.37)$$

Proof. When $N > \sqrt{2/\delta}$, we have $1 + 1/N^2 < 1 + \delta/2$, and so

$$\frac{r}{2} n^2 e^{\frac{r}{2} n^2 \left(1 + \frac{1}{N^2}\right)} \leq \frac{r}{2} n^2 e^{-\frac{r}{4} \delta n^2} e^{\frac{r}{2} (1+\delta) n^2}. \quad (3.38)$$

Elementary calculus shows that the function $x \mapsto x e^{-\delta x/2}$ is maximized at $x = 2/\delta$, and takes value $2/e\delta < 1/\delta$ there. Substituting $x = \frac{r}{2} n^2$ in (3.38), the result now follows from (3.27). \square

That being said, the author does not believe the estimate (3.37) on the constant $C(s, t, P)$ in (3.21) is anywhere close to optimal: the above proofs involved fairly blunt estimates that ignored many potential cancellations. Indeed, if we work explicitly in the case $N = 1$, for any linear polynomial $\mathcal{H}\mathcal{P} \ni P = \sum_{k=-n}^n a_k \nu_k$, (2.12) shows that

$$(e^{-\mathcal{D}_{t,0}^1} P)(\mathbf{1}) = \left(e^{\frac{t}{2} \Delta_{\mathbb{U}_1}} \left(\sum_{k=-n}^n a_k U^k \right) \right) \Big|_{U=I_1} = \sum_{k=-n}^n a_k e^{-\frac{t}{2} k^2}$$

while (3.22) shows that

$$(e^{-\mathcal{D}_{t,0}} P)(\mathbf{1}) = \sum_{k=-n}^n a_k \nu_k(t).$$

Thus, we have

$$\left| (e^{-\mathcal{D}_{t,0}^1} P)(\mathbf{1}) - (e^{-\mathcal{D}_{t,0}} P)(\mathbf{1}) \right| \leq \sum_{k=-n}^n |e^{-\frac{t}{2} k^2} - \nu_k(t)| |a_k| \leq 2 \|P\|_1 \quad (3.39)$$

since $0 < e^{-\frac{t}{2} k^2} \leq 1$ and $|\nu_k(t)| \leq 1$ (as it is a moment of a probability measure on \mathbb{U}). On \mathbb{U}_1 , every trace polynomial reduces to a polynomial in U which intertwines with a linear polynomial (since $\text{tr}(U^k) = U^k$ for $U \in \mathbb{U}_1$). This reduction process can only increase the $\|\cdot\|_1$ -norm; cf. Remark 3.17. Thus, (3.39) shows that, in the special case $N = 1$, there is a *uniform bound* (uniform in n and t) for the concentration of expectations of polynomials in \mathcal{P}_n . It does not follow easily, unfortunately, that $C(t, 0, P)$ is uniformly bounded in the \mathbb{U}_N case; but the author strongly suspects this is so. We leave the investigation of the precise behavior of the constants $C(s, t, P)$ to a future publication.

4 Convergence of Empirical Distributions

This section is devoted to the proofs of Theorems 1.2-1.7. Theorem 1.2 is treated first, separately, with specialized techniques adapted from [30]. We then proceed with Theorem 1.7, and then derive Theorems 1.3 and 1.5 essentially as special cases.

4.1 Empirical Eigenvalues on \mathbb{U}_N

Let $f: \mathbb{U} \rightarrow \mathbb{C}$ be a measurable function. Since the group \mathbb{U}_N consists of normal matrices, measurable functional calculus is available to us. From (1.16), the empirical integral $\int_{\mathbb{U}} f d\tilde{\nu}_t^N$ is the random variable

$$\int_{\mathbb{U}} f d\tilde{\nu}_t^N = \text{tr} \circ f_N \quad \text{on} \quad (\mathbb{U}_N, \rho_t^N). \quad (4.1)$$

We will initially bound the empirical integral in terms of the *Lipschitz norm* on test functions. A function $F: \mathbb{U}_N \rightarrow \mathbb{C}$ is *Lipschitz* if

$$\|F\|_{\text{Lip}(\mathbb{U}_N)} \equiv \sup_{U \neq V \in \mathbb{U}_N} \frac{|F(U) - F(V)|}{d_{\mathbb{U}_N}(U, V)} < \infty,$$

where $d_{\mathbb{U}_N}$ is the Riemannian distance on \mathbb{U}_N given by the left-invariant metric induced by the inner product (2.2) on \mathfrak{u}_N . In the special case $N = 1$, this is just arclength distance:

$$\|f\|_{\text{Lip}(\mathbb{U})} = \frac{1}{2\pi} \sup_{\alpha \neq \beta \in [0, 2\pi)} \frac{|f(e^{i\alpha}) - f(e^{i\beta})|}{|\alpha - \beta|}. \quad (4.2)$$

The following general lemma was given in [30, Proposition 4.1]; it is adapted from the now well-known techniques in [21], and attributable to earlier work of Talagrand.

Lemma 4.1 (Lévy, Maïda, 2010). *Let $N \in \mathbb{N}$. If $f: \mathbb{U} \rightarrow \mathbb{C}$ is Lipschitz, then $\text{tr} \circ f_N: \mathbb{U}_N \rightarrow \mathbb{C}$ is Lipschitz, and*

$$\|\text{tr} \circ f_N\|_{\text{Lip}(\mathbb{U}_N)} = \frac{1}{N} \|f\|_{\text{Lip}(\mathbb{U})}. \quad (4.3)$$

Remark 4.2. Lemma 4.1 is proved in [30] only for *real*-valued f ; but the proof works without modification for complex valued test functions.

Lemma 4.1 is then used in conjunction with the following, proved as [30, Proposition 6.1].

Lemma 4.3 (Lévy, Maïda, 2010). *Let $F: \mathbb{U}_N \rightarrow \mathbb{R}$ be Lipschitz, and let $N \in \mathbb{N}$. Then for $t \geq 0$,*

$$\text{Var}_{\rho_t^N}(F) \leq t \|F\|_{\text{Lip}(\mathbb{U}_N)}^2. \quad (4.4)$$

Lemma 4.3 is proved using a fairly well-known martingale method. If U_t^N is a Brownian motion on \mathbb{U}_N (i.e. the Markov process with generator $\frac{1}{2}\Delta_{\mathbb{U}_N}$), and $T > 0$, then for any L^2 -function $F: \mathbb{U}_N \rightarrow \mathbb{R}$, the real-valued stochastic process

$$t \mapsto \left(e^{\frac{1}{2}(T-t)\Delta_{\mathbb{U}_N}} F \right) (U_t^N)$$

is a martingale, which is well-behaved when F is Lipschitz (in particular since $\|e^{\frac{t}{2}\Delta_{\mathbb{U}_N}} F\|_{\text{Lip}(\mathbb{U}_N)} \leq \|F\|_{\text{Lip}(\mathbb{U}_N)}$ for any $t \geq 0$). Our first task is to generalize Lemma 4.3 in two ways: from variances to covariances, and from real-valued to complex-valued random variables.

Corollary 4.4. *Let $N \in \mathbb{N}$ and $t \geq 0$. If $F, G: \mathbb{U}_N \rightarrow \mathbb{C}$ are Lipschitz functions, then*

$$\left| \text{Cov}_{\rho_t^N}(F, G) \right| \leq 2t \|F\|_{\text{Lip}(\mathbb{U}_N)} \|G\|_{\text{Lip}(\mathbb{U}_N)}. \quad (4.5)$$

Remark 4.5. To be clear: for two complex-valued L^2 random variables F and G , $\text{Cov}(F, G) = \mathbb{E}(F\overline{G}) - \mathbb{E}(F)\mathbb{E}(\overline{G}) = \mathbb{E}[(F - \mathbb{E}(F))(\overline{G} - \mathbb{E}(\overline{G}))]$.

Proof. From the Cauchy-Schwarz inequality, we have

$$|\text{Cov}(F, G)| = \left| \mathbb{E}[(F - \mathbb{E}(F))(\overline{G} - \mathbb{E}(\overline{G}))] \right| \leq \|F - \mathbb{E}(F)\|_{L^2} \|G - \mathbb{E}(G)\|_{L^2} = \sqrt{\text{Var}(F)\text{Var}(G)}. \quad (4.6)$$

Note that, for a complex-valued random variable $F = F_1 + iF_2$, $\text{Var}(F_1 + iF_2) = \text{Var}(F_1) + \text{Var}(F_2)$. A complex-valued function is Lipschitz iff its real and imaginary parts are both Lipschitz, and so Lemma 4.3 shows that

$$\text{Var}_{\rho_t^N}(F_1 + iF_2) = \text{Var}_{\rho_t^N}(F_1) + \text{Var}_{\rho_t^N}(F_2) \leq t \left(\|F_1\|_{\text{Lip}(\mathbb{U}_N)}^2 + \|F_2\|_{\text{Lip}(\mathbb{U}_N)}^2 \right). \quad (4.7)$$

We now estimate

$$\begin{aligned} \|F_1\|_{\text{Lip}(\mathbb{U}_N)}^2 + \|F_2\|_{\text{Lip}(\mathbb{U}_N)}^2 &\leq 2 \max \left\{ \|F_1\|_{\text{Lip}(\mathbb{U}_N)}^2, \|F_2\|_{\text{Lip}(\mathbb{U}_N)}^2 \right\} \\ &\leq 2 \max \left\{ \sup_{U \neq V} \frac{(F_1(U) - F_1(V))^2}{d_{\mathbb{U}_N}(U, V)^2}, \sup_{U \neq V} \frac{(F_2(U) - F_2(V))^2}{d_{\mathbb{U}_N}(U, V)^2} \right\} \\ &\leq 2 \sup_{U \neq V} \left[\frac{(F_1(U) - F_1(V))^2}{d_{\mathbb{U}_N}(U, V)^2} + \frac{(F_2(U) - F_2(V))^2}{d_{\mathbb{U}_N}(U, V)^2} \right] \\ &= 2 \|F_1 + iF_2\|_{\text{Lip}(\mathbb{U}_N)}^2, \end{aligned}$$

where the penultimate inequality is just the statement that if $f_1, f_2 \geq 0$ then $\sup(f_1 + f_2) \geq \max\{\sup f_1, \sup f_2\}$. Combining this with (4.6) and (4.7) proves the (4.5). \square

Remark 4.6. It is likely that the variance estimate (4.4) holds as stated for complex-valued F , but this is not immediately clear from the proof as given. Since we do not care too much about exact constants, we are content to have a possibly-extraneous factor of 2 in (4.5).

Combining Lemma 4.1 and Corollary 4.4 (in the special case $F = G$) with (4.1) immediately proves (1.5) in Theorem 1.2. We will now show that, at the expense of decreasing the speed of convergence below $O(1/N^2)$ (but still summably fast), convergence holds for the much less regular functions in the Sobolev spaces $H_p(\mathbb{U})$ for $p > 1$. (If $p < \frac{3}{2}$, $H_p(\mathbb{U})$ consists primarily of non-Lipschitz functions; cf. Section 2.2.) We begin by considering trigonometric polynomial test functions.

Proposition 4.7. *Let $n \in \mathbb{N}$, and let $f(u) = \sum_{k=-n}^n \hat{f}(k)u^k$ be a trigonometric polynomial on \mathbb{U} . If $1 < p < \frac{3}{2}$, then*

$$\text{Var} \left(\int_{\mathbb{U}} f d\tilde{\nu}_t^N \right) \leq \frac{n^{3-2p}}{\pi^2 N^2} \cdot \frac{4t}{3-2p} \|f\|_{H_p(\mathbb{U})}^2. \quad (4.8)$$

Proof. From (1.16), the empirical integral is the random variable

$$\int_{\mathbb{U}} f d\tilde{\nu}_t^N = \sum_{k=-n}^n \hat{f}(k) \text{tr}[(\cdot)^k],$$

and so we can expand the variance as

$$\text{Var} \left(\int_{\mathbb{U}} f d\tilde{\nu}_t^N \right) = \sum_{|j|, |k| \leq n} \hat{f}(j) \overline{\hat{f}(k)} \text{Cov}_{\rho_t^N}(\text{tr}[(\cdot)^j], \text{tr}[(\cdot)^k]). \quad (4.9)$$

Using Corollary 4.4 and then Lemma 4.1, we have

$$\left| \text{Cov}_{\rho_t^N}(\text{tr}[(\cdot)^j], \text{tr}[(\cdot)^k]) \right| \leq 2t \|\text{tr}[(\cdot)^j]\|_{\text{Lip}(\mathbb{U}_N)} \|\text{tr}[(\cdot)^k]\|_{\text{Lip}(\mathbb{U}_N)} = \frac{2t}{N^2} \|\chi_j\|_{\text{Lip}(\mathbb{U})} \|\chi_k\|_{\text{Lip}(\mathbb{U})}, \quad (4.10)$$

where $\chi_k(u) = u^k$; cf. Section 2.2. Since the functions χ_k are in $C^1(\mathbb{U})$, we can compute their Lipschitz norms as

$$\|\chi_k\|_{\text{Lip}(\mathbb{U})} = \frac{1}{2\pi} \sup_{\mathbb{U}} |\chi_k'| = \frac{1}{2\pi} |k|.$$

Combining this with (4.9) and (4.10) yields

$$\text{Var} \left(\int_{\mathbb{U}} f d\tilde{\nu}_t^N \right) \leq \frac{t}{2\pi^2 N^2} \sum_{|j|, |k| \leq n} |\hat{f}(j)| |\hat{f}(k)| |j| |k| = \frac{t}{2\pi^2 N^2} \left(\sum_{k=-n}^n |k| |\hat{f}(k)| \right)^2. \quad (4.11)$$

Note that the $k = 0$ term in the squared-sum is 0, so we omit it from here on. We estimate this squared-sum with the Cauchy-Schwarz inequality, applied with $|k| = |k|^{1-p} |k|^p$:

$$\begin{aligned} \left(\sum_{1 \leq |k| \leq n} |k| |\hat{f}(k)| \right)^2 &\leq \left(\sum_{1 \leq |k| \leq n} |k|^{2(1-p)} \right) \cdot \left(\sum_{1 \leq |k| \leq n} |k|^{2p} |\hat{f}(k)|^2 \right) \\ &\leq \left(\sum_{1 \leq |k| \leq n} |k|^{2(1-p)} \right) \cdot \|f\|_{H_p(\mathbb{U})}^2, \end{aligned} \quad (4.12)$$

where the Sobolev H_p -norm is defined in (2.9). Let $r = 2(p-1)$; then $0 < r \leq 1$. We utilize the calculus estimate

$$\sum_{k=1}^{\infty} \frac{1}{k^r} \leq 2^r \int_1^{n+1} \frac{dx}{x^r} = \frac{2^r}{1-r} [(n+1)^{1-r} - 1] \leq \frac{2}{1-r} n^{1-r},$$

which yields

$$\sum_{1 \leq |k| \leq n} |k|^{2(1-p)} = 2 \sum_{k=1}^{\infty} k^{2(1-p)} \leq \frac{4}{3-2p} n^{3-2p}. \quad (4.13)$$

Equations (4.12) and (4.13) prove the proposition. \square

Remark 4.8. In the regime $p > \frac{3}{2}$, where $2(1-p) < -1$, the sum in (4.12) is uniformly bounded in n , and the resulting estimate on the variance is

$$\text{Var} \left(\int_{\mathbb{U}} f d\tilde{\nu}_t^N \right) \leq \frac{1}{\pi^2 N^2} \cdot \frac{4^p t}{2p-3} \|f\|_{H^p(\mathbb{U})}^2, \quad p > \frac{3}{2}.$$

In the case $p = \frac{3}{2}$, $H_p(\mathbb{U})$ corresponds roughly with Lipschitz functions, and so (1.5) is the optimal result.

We will use Proposition 4.7 to prove (1.4) by doing a band-limit cut-off of the test function f at a frequency n that grows with N (in fact, the optimal result is achieved at $n = N$). To proceed, we first need the following lemma.

Lemma 4.9. *Let $N \in \mathbb{N}$ and $t \geq 0$. For $f \in L^\infty(\mathbb{U})$,*

$$\text{Var} \left(\int_{\mathbb{U}} f d\tilde{\nu}_t^N \right) \leq 4 \|f\|_{L^\infty(\mathbb{U})}^2. \quad (4.14)$$

Proof. For any L^2 random variable F , we utilize the crude estimate

$$\text{Var}(F) = \|F - \mathbb{E}(F)\|_{L^2}^2 \leq (\|F\|_{L^2} + |\mathbb{E}(F)|)^2 \leq 4\|F\|_{L^2}^2.$$

With $F = \int_{\mathbb{U}} f d\tilde{\nu}_t^N$, (1.15) shows that, for $U \in \mathbb{U}_N$,

$$|F(U)| = \frac{1}{N} \left| \sum_{\lambda \in \Lambda(U)} f(\lambda) \right| \leq \|f\|_{L^\infty(\mathbb{U})}$$

since $\Lambda(U)$ is a set of size N . Since ρ_t^N is a probability measure, it follows that $\|F\|_{L^2(\rho_t^N)} \leq \|f\|_{L^\infty(\mathbb{U})}$, and the result follow. \square

We now proceed to prove (1.4) in Theorem 1.2.

Proposition 4.10. *Let $t \geq 0$, $N \in \mathbb{N}$ and $1 < p < \frac{3}{2}$. For $f \in H_p(\mathbb{U})$,*

$$\text{Var} \left(\int_{\mathbb{U}} f d\tilde{\nu}_t^N \right) \leq \frac{1}{N^{2p-1}} \cdot \frac{4}{\pi^2} \|f\|_{H_p(\mathbb{U})} \left(\frac{\sqrt{t}}{\sqrt{3-2p}} + \frac{1}{\sqrt{2p-1}} \right)^2. \quad (4.15)$$

Proof. Fix $f \in H_p(\mathbb{U})$, with Fourier expansion $f = \sum_{k \in \mathbb{Z}} \hat{f}(k) \chi_k$. Let

$$f_N = \sum_{k=-N}^N \hat{f}(k) \chi_k$$

be the band-limited frequency cut-off at level N , and define

$$F_N = \int_{\mathbb{U}} f_N d\tilde{\nu}_t^N, \quad \text{and} \quad F^N = \int_{\mathbb{U}} (f - f_N) d\tilde{\nu}_t^N,$$

so that $F_N + F^N = \int_{\mathbb{U}} f d\tilde{\nu}_t^N$. From the triangle inequality for L^2 ,

$$\left(\text{Var} \left(\int_{\mathbb{U}} f d\tilde{\nu}_t^N \right) \right)^{1/2} = \sqrt{\text{Var}(F_N + F^N)} \leq \sqrt{\text{Var}(F_N)} + \sqrt{\text{Var}(F^N)}. \quad (4.16)$$

From Proposition 4.7, the square of the first term in (4.16) is bounded by

$$\text{Var}(F_N) \leq \frac{N^{3-2p}}{\pi^2 N^2} \cdot \frac{4t}{3-2p} \|f_N\|_{H^p(\mathbb{U})}^2 \leq N^{1-2p} \cdot \frac{4t}{\pi^2(3-2p)} \|f\|_{H^p(\mathbb{U})}^2. \quad (4.17)$$

From Lemma 4.9, the square of the second term in (4.16) is bounded by

$$\text{Var}(F^N) \leq 4 \|f - f_N\|_{L^\infty(\mathbb{U})}^2, \quad (4.18)$$

which we can bound as follows:

$$\begin{aligned} \sup_{u \in \mathbb{U}} |f(u) - f_N(u)|^2 &= \sup_{u \in \mathbb{U}} \left| \sum_{|k| > N} \hat{f}(k) u^k \right|^2 \leq \left(\sum_{|k| > N} |\hat{f}(k)| \right)^2 \leq \left(\sum_{|k| > N} |k|^{-2p} \right) \cdot \left(\sum_{|k| > N} |k|^{2p} |\hat{f}(k)|^2 \right) \\ &\leq \left(\sum_{|k| > N} |k|^{-2p} \right) \|f\|_{H^p(\mathbb{U})}^2. \end{aligned} \quad (4.19)$$

We can bound the above sum as in (4.13), using the calculus estimate

$$\sum_{k=N+1}^{\infty} \frac{1}{k^{2p}} \leq \int_N^{\infty} \frac{dx}{x^{2p}} = \frac{1}{2p-1} N^{1-2p}.$$

Combining this with (4.18) and (4.19) yields

$$\text{Var}(F^N) \leq N^{1-2p} \cdot \frac{8}{2p-1} \|f\|_{H^p(\mathbb{U})}^2. \quad (4.20)$$

Combining (4.16), (4.17), (4.20) proves (4.15). \square

This brings us to the proof of Theorem 1.2.

Proof of Theorem 1.2. Proposition 4.10 proves (1.4), while, as remarked above, Lemma 4.1 and Corollary 4.4 prove (1.5). Thus, we are left to prove only (1.3). Fix $f \in C(\mathbb{U})$, and let $\epsilon > 0$. By the Weierstrass approximation theorem, there is a trigonometric polynomial g_ϵ on \mathbb{U} such that $\|f - g_\epsilon\|_{L^\infty(\mathbb{U})} < \sqrt{\epsilon}/4$. Let

$$F = \int_{\mathbb{U}} f d\tilde{\nu}_t^N, \quad \text{and} \quad G = \int_{\mathbb{U}} g_\epsilon d\tilde{\nu}_t^N.$$

Then, as in (4.16), we estimate

$$\sqrt{\text{Var}(F)} \leq \sqrt{\text{Var}(G)} + \sqrt{\text{Var}(F - G)} \leq \sqrt{\text{Var}(G)} + 2\|f - g_\epsilon\|_{\infty} < \sqrt{\text{Var}(G)} + \sqrt{\epsilon}/2 \quad (4.21)$$

by Lemma 4.9. Now, g_ϵ is Lipschitz, and so (1.5) gives

$$\sqrt{\text{Var}(G)} \leq \frac{2t}{N^2} \|g_\epsilon\|_{\text{Lip}(\mathbb{U})}. \quad (4.22)$$

Thus, for any $N > 2\sqrt{t\|g_\epsilon\|_{\text{Lip}(\mathbb{U})}/\epsilon^{1/4}}$, $\sqrt{\text{Var}(G)} < \sqrt{\epsilon}/2$, and so (4.21) and (4.22) show that $\text{Var}(F) = \text{Var}(\int_{\mathbb{U}} f d\tilde{\nu}_t^N) < \epsilon$ for all sufficiently large N . Convergence in probability (1.3) now follows immediately from Chebyshev's inequality. \square

For a discussion of the (lack of) sharpness of (1.4), see the end of Section 4.2.

4.2 Empirical Noncommutative Distribution on \mathbb{GL}_N

Definition 4.11. Let $s, t \in \mathbb{R}$, and let $\mathcal{D}_{s,t}$ be the intertwining operator on \mathcal{P} given in Theorem 3.6. For each n , the finite dimensional subspace \mathcal{P}_n is invariant under $\mathcal{D}_{s,t}$, and so $e^{-\mathcal{D}_{s,t}}: \mathcal{P} \rightarrow \mathcal{P}$ is well-defined. Define the **noncommutative distribution** $\varphi_{s,t}: \mathbb{C}\langle A, A^* \rangle \rightarrow \mathbb{C}$ to be the following linear functional:

$$\varphi_{s,t}(f) = (e^{-\mathcal{D}_{s,t}}\Upsilon(f))(\mathbf{1}), \quad f \in \mathbb{C}\langle A, A^* \rangle \quad (4.23)$$

where $\Upsilon: \mathbb{C}\langle A, A^* \rangle \hookrightarrow \mathcal{P}^+$ is the inclusion of (2.31).

To be clear: $\mathcal{D}_{s,t}$ does not preserve the space $\Upsilon(\mathbb{C}\langle A, A^* \rangle)$ of linear polynomials, and so $e^{-\mathcal{D}_{s,t}}f$ contains terms of higher (ordinary) degree, although it preserves the *trace* degree of $\Upsilon(f)$. The functional $\varphi_{s,t}$ is defined by evaluating the resultant polynomial function $\mathbf{v} \mapsto (e^{-\mathcal{D}_{s,t}}\Upsilon(f))(\mathbf{v}) \in \mathcal{P}^+$ at $\mathbf{v} = \mathbf{1}$.

Remark 4.12. It is tempting to think that $\varphi_{s,t}$ is therefore a homomorphism on $\mathbb{C}\langle A, A^* \rangle$, since $e^{-\mathcal{D}_{s,t}}$ is a homomorphism on \mathcal{P}^+ . However, Υ is *not* a homomorphism. The product on $\mathbb{C}\langle A, A^* \rangle$ is incompatible with the product on the larger space \mathcal{P}^+ ; it is the difference between convolution product and pointwise product of functions.

To properly call the linear functional (4.23) a non-commutative distribution, we must realize it as the distribution of a random variable in a noncommutative probability space $(\mathcal{A}_{s,t}, \tau_{s,t})$. This is done in precisely the same way that we constructed the mean $\mathbb{E}(\tilde{\varphi}^N)$ of an empirical distribution (2.27) as a genuine noncommutative distribution. We take $\mathcal{A}_{s,t} = \mathbb{C}\langle A, A^* \rangle$, and define $\tau_{s,t}(f) = \varphi_{s,t}(f)$ for $f \in \mathcal{A}$; then $\varphi_{s,t} = \varphi_a$ where $a \in \mathbb{C}\langle A, A^* \rangle$ is the coordinate random variable $a(A, A^*) = A$. Note that $\varphi_{s,t}(\mathbf{1}) = 1$ since $\mathcal{D}_{s,t}$ annihilates constants. That $\tau_{s,t}$ is tracial and positive semi-definite actually follows from Theorem 1.7: (1.10) identifies $\varphi_{s,t}$ as the limit of the mean distributions $\mathbb{E}(\tilde{\varphi}_{s,t}^N)$ which are tracial and positive definite (since $\mu_{s,t}^N$ has infinite support); see the discussion on page 12. It is straightforward to verify that a limit of tracial states is tracial, and hence $\tau_{s,t}$ is a tracial state. What is not so clear from this definition is whether $\tau_{s,t}$ is *faithful*, as this property does not generally survive under limits. In the special case $s = t$, the concurrent paper [11] proves that $\varphi_{t,t}$ is the noncommutative distribution of the free multiplicative Brownian motion g_t of (2.37), and so in this case, $\tau_{t,t}$ is known to be faithful. In fact, using free stochastic calculus techniques, it is possible to construct $\varphi_{s,t}$ as the distribution of a stochastic process in a faithful W^* -probability space; we address this in a separate publication [26].

The key to proving Theorem 1.7 is the following extension of Corollary 3.14. We will use it here only in the diagonal case ($P = Q$), but the general covariance estimate will be useful in Sections 4.3 and 4.4.

Proposition 4.13. For $P, Q \in \mathcal{P}$, there is a constant $C_2(s, t, P, Q)$ depending continuously on s, t so that, for each $N \in \mathbb{N}$,

$$\left| \text{Cov}_{\mu_{s,t}^N}(P \circ \mathbf{V}_N, Q \circ \mathbf{V}_N) \right| \leq \frac{1}{N^2} \cdot C_2(s, t, P, Q). \quad (4.24)$$

Proof. From (3.13), we may write

$$P \circ \mathbf{V}_N \cdot \overline{Q \circ \mathbf{V}_N} = (PQ^*) \circ \mathbf{V}_N$$

where $Q^* = \mathbb{C}Q$. Thus, (3.15) shows that

$$\mathbb{E}_{\mu_{s,t}^N}(P \circ \mathbf{V}_N \cdot \overline{Q \circ \mathbf{V}_N}) = (e^{-\mathcal{D}_{s,t}^N}(PQ^*))(\mathbf{1}). \quad (4.25)$$

Similarly,

$$\mathbb{E}_{\mu_{s,t}^N}(P \circ \mathbf{V}_N) \cdot \mathbb{E}_{\mu_{s,t}^N}(\overline{Q \circ \mathbf{V}_N}) = (e^{-\mathcal{D}_{s,t}^N}P)(\mathbf{1}) \cdot (e^{-\mathcal{D}_{s,t}^N}Q^*)(\mathbf{1}). \quad (4.26)$$

To simplify notation, we suppress s, t and denote

$$\Psi_1^N \equiv (e^{-\mathcal{D}_{s,t}^N} P)(\mathbf{1}), \quad \Psi_*^N \equiv (e^{-\mathcal{D}_{s,t}^N} Q^*)(\mathbf{1}), \quad \Psi_{1,*}^N \equiv (e^{-\mathcal{D}_{s,t}^N} (PQ^*))(\mathbf{1}), \quad (4.27)$$

$$\Psi_1 \equiv (e^{-\mathcal{D}_{s,t}} P)(\mathbf{1}), \quad \Psi_* \equiv (e^{-\mathcal{D}_{s,t}} Q^*)(\mathbf{1}), \quad \Psi_{1,*} \equiv (e^{-\mathcal{D}_{s,t}} (PQ^*))(\mathbf{1}). \quad (4.28)$$

Thus, (4.25) and (4.26) show that

$$\text{Cov}_{\mu_{s,t}^N}(P \circ \mathbf{V}_N, Q \circ \mathbf{V}_N) = \Psi_{1,*}^N - \Psi_1^N \Psi_*^N. \quad (4.29)$$

We estimate this as follows. First

$$|\Psi_{1,*}^N - \Psi_1^N \Psi_*^N| \leq |\Psi_{1,*}^N - \Psi_{1,*}| + |\Psi_{1,*} - \Psi_1 \Psi_*| + |\Psi_1 \Psi_* - \Psi_1^N \Psi_*^N|. \quad (4.30)$$

Referring to (4.28), since $e^{-\mathcal{D}_{s,t}}$ is a homomorphism, the second term in (4.30) is 0. The first term is bounded by $\frac{1}{N^2} \cdot C(s, t, PQ^*)$ by Corollary 3.14. For the third term, we add and subtract $\Psi_1^N \Psi_*^N$ to make the additional estimate

$$\begin{aligned} |\Psi_1 \Psi_* - \Psi_1^N \Psi_*^N| &\leq |\Psi_*| |\Psi_1 - \Psi_1^N| + |\Psi_1^N| |\Psi_* - \Psi_*^N| \\ &\leq |\Psi_*| |\Psi_1 - \Psi_1^N| + (|\Psi_1| + |\Psi_1^N - \Psi_1|) |\Psi_* - \Psi_*^N| \\ &\leq \frac{1}{N^2} \cdot |\Psi_*| C(s, t, P) + \left(|\Psi_1| + \frac{1}{N^2} \cdot C(s, t, P) \right) \cdot \frac{1}{N^2} \cdot C(s, t, Q^*) \\ &= \frac{1}{N^2} \cdot (|\Psi_*| C(s, t, P) + |\Psi_1| C(s, t, Q^*)) + \frac{1}{N^4} \cdot C(s, t, P) C(s, t, Q^*). \end{aligned} \quad (4.31)$$

Combining (4.31) with (4.29) – (4.30) and the following discussion shows that the constant

$$C_2(s, t, P, Q) = C(s, t, PQ^*) + C(s, t, P) C(s, t, Q^*) + |\Psi_*| C(s, t, P) + |\Psi_1| C(s, t, Q^*) \quad (4.32)$$

verifies (4.13), proving the proposition. \square

Proposition 4.13 shows that any trace polynomial in $Z_{s,t}^N$ has variance of order $1/N^2$, as discussed following the statement of Theorem 1.7. The theorem follows as a very special case, due to the following.

Lemma 4.14. *Let $Z \in \mathbb{G}\mathbb{L}_N$, and let $f \in \mathbb{C}\langle A, A^* \rangle$. Let φ_Z denote the noncommutative distribution of Z with respect to $(\mathbb{M}_N, \text{tr})$ (Definition 2.12), let $\Upsilon: \mathbb{C}\langle A, A^* \rangle \hookrightarrow \mathcal{P}^+$ be the map of (2.31), and let \mathbf{V}_N be the map in Notation 3.2. Then*

$$\varphi_Z(f) = (\Upsilon(f) \circ \mathbf{V}_N)(Z).$$

Proof. As both sides are linear functions of f , it suffices to prove the claim on basis elements $f(A, A^*) = A^\varepsilon$ for some $\varepsilon \in \mathcal{E}^+$. Then $\Upsilon(f) = v_\varepsilon$, and $(v_\varepsilon) \circ \mathbf{V}_N(Z) = \text{tr}(Z^\varepsilon) = \varphi_Z(A^\varepsilon)$ as claimed. \square

This brings us to the proof of Theorem 1.7.

Proof of Theorem 1.7. We begin by establishing that (1.10) holds with the linear functional $\varphi_{s,t}$ of Definition 4.11. From (2.27), we have

$$\mathbb{E}(\tilde{\varphi}_{s,t}^N)(f) = \int_{\mathbb{G}\mathbb{L}_N} \varphi_Z(f) \mu_{s,t}^N(dZ)$$

where φ_Z is the noncommutative distribution of Z in $(\mathbb{M}_N, \text{tr})$. Applying Lemma 4.14 and (3.15) yields

$$\mathbb{E}(\tilde{\varphi}_{s,t}^N)(f) = \mathbb{E}_{\mu_{s,t}^N}(\Upsilon(f) \circ \mathbf{V}_N) = (e^{-\mathcal{D}_{s,t}^N} \Upsilon(f))(\mathbf{1}). \quad (4.33)$$

From the definition (4.23) of the limit distribution $\varphi_{s,t}$, (4.33) shows that

$$|\mathbb{E}(\tilde{\varphi}_{s,t}^N)(f) - \varphi_{s,t}(f)| = \left| (e^{-\mathcal{D}_{s,t}^N} \Upsilon(f))(\mathbf{1}) - (e^{-\mathcal{D}_{s,t}} \Upsilon(f))(\mathbf{1}) \right| \leq \frac{1}{N^2} \cdot C(s, t, \Upsilon(f))$$

by Corollary 3.14; this proves (1.10).

The random variable $\tilde{\varphi}_{s,t}^N$ on the probability space $(\mathbb{GL}_N, \mu_{s,t}^N)$ has value φ_Z at $Z \in \mathbb{GL}_N$. Thus, using Lemma 4.14, we have

$$\text{Var}[\tilde{\varphi}_{s,t}^N(f)] = \text{Var}_{\mu_{s,t}^N}(\Upsilon(f) \circ \mathbf{V}_N), \quad (4.34)$$

and (1.11) follows immediately from Proposition 4.13. \square

We now give some quantitative estimate for the constant $C_2(s, t, P, Q)$ of (4.32). First we need to bound the terms $|\Psi_*|$ and $|\Psi_1|$ in that equation.

Lemma 4.15. *Let $s, t \in \mathbb{R}$, with $r = |s - \frac{t}{2}| + \frac{1}{2}|t|$. Let $n, N \in \mathbb{N}$, and let $P \in \mathcal{P}_n$. Then*

$$\left| (e^{-\mathcal{D}_{s,t}^N} P)(\mathbf{1}) \right| \leq e^{\frac{r}{2}(1+1/N^2)n^2} \|P\|_1, \quad \text{and} \quad \left| (e^{-\mathcal{D}_{s,t}} P)(\mathbf{1}) \right| \leq e^{\frac{r}{2}n^2} \|P\|_1. \quad (4.35)$$

Proof. Following (3.30), (3.33), and (3.35), we estimate

$$\left| (e^{-\mathcal{D}_{s,t}^N} P)(\mathbf{1}) \right| \leq \|e^{-\mathcal{D}_{s,t}^N} P\|_1 \leq \|e^{-\mathcal{D}_{s,t}^N}|_{\mathcal{P}_n}\|_{1 \rightarrow 1} \|P\|_1 \leq e^{\|\mathcal{D}_{s,t}^N|_{\mathcal{P}_n}\|_{1 \rightarrow 1}} \|P\|_1 \leq e^{\frac{r}{2}(1+1/N^2)n^2} \|P\|_1,$$

proving the first inequality in (4.35). The second follows by taking $N \rightarrow \infty$. \square

Corollary 4.16. *Let $s, t \in \mathbb{R}$ with $r = |s - \frac{t}{2}| + \frac{1}{2}|t|$, $n, m, N \in \mathbb{N}$, and $0 < \delta < 1$. For $P \in \mathcal{P}_n$, $Q \in \mathcal{P}_m$, and $N > \sqrt{2/\delta}$,*

$$\text{Cov}_{\mu_{s,t}^N}(P \circ \mathbf{V}_N, Q \circ \mathbf{V}_N) \leq \frac{1}{N^2} \cdot \frac{4}{\delta^2} e^{r(1+\delta)(n^2+m^2)} \|P\|_1 \|Q\|_1. \quad (4.36)$$

Proof. The polynomial Q^* has trace degree m , and so PQ^* has trace degree $n + m$. It therefore follows from (4.32), together with Corollary 3.19 and Lemma 4.15, that

$$\begin{aligned} \text{Cov}_{\mu_{s,t}^N}(P \circ \mathbf{V}_N, Q \circ \mathbf{V}_N) &\leq \frac{1}{N^2} \cdot \left[\frac{1}{\delta} e^{\frac{r}{2}(1+\delta)(n+m)^2} \|PQ^*\|_1 \right. \\ &\quad \left. + \frac{1}{\delta^2} e^{\frac{r}{2}(1+\delta)(n^2+m^2)} \|P\|_1 \|Q^*\|_1 + \frac{1}{\delta} (e^{\frac{r}{2}m^2} e^{\frac{r}{2}(1+\delta)n^2} + e^{\frac{r}{2}n^2} e^{\frac{r}{2}(1+\delta)m^2}) \|P\|_1 \|Q^*\|_1 \right]. \end{aligned}$$

The reader can readily verify that $\|PQ^*\|_1 \leq \|P\|_1 \|Q\|_1$. Together with the estimate $(n + m)^2 \leq 2(n^2 + m^2)$, the argument of Corollary 3.19 then proves (4.36). \square

We conclude this section with a brief discussion of what bounds we expect are sharp, and the consequences this would have for the proof of Theorem 1.2. As mentioned in the remarks following the suggestive calculation (3.39), it is possible that the constants $C(t, 0, P)$ of Corollary 3.21 are uniformly bounded over $P \in \mathcal{H}\mathcal{P}$. To be precise, we conjecture that there is a constant $C(t)$ (depending continuously on $t > 0$) so that

$$\left| (e^{-\mathcal{D}_{t,0}^N} P)(\mathbf{1}) - (e^{-\mathcal{D}_{t,0}} P)(\mathbf{1}) \right| \leq \frac{C(t)}{N^2}, \quad P \in \mathcal{H}\mathcal{P}. \quad (4.37)$$

If this holds true, then as in the proof of Proposition 4.13, it would follow that there is a constant $C_2(t)$ such that, for $j, k \in \mathbb{Z}$,

$$\left| \text{Cov}_{\rho_{s,t}^N}(v_j \circ \mathbf{V}_N, v_k \circ \mathbf{V}_N) \right| \leq \frac{C_2(t)}{N^2}. \quad (4.38)$$

Indeed: the terms $|\Psi|$ and $|\Psi_*|$ in (4.27) and (4.32) are ≤ 1 , since $\Psi_1 = \lim_{N \rightarrow \infty} \Psi_1^N = \lim_{N \rightarrow \infty} \mathbb{E}_{\rho_t^N} \text{tr}[(\cdot)^j]$ and $|\text{tr}(U^j)| \leq 1$ for $U \in \mathbb{U}_N$ (see the proof of Lemma 4.9), and similarly $|\Psi_*| = \lim_{N \rightarrow \infty} |\text{tr}[(\cdot)^{-k}]| \leq 1$.

Consider, then, $f \in H_p(\mathbb{U})$ with $p > \frac{1}{2}$; the covariance expansion (4.9) together with the conjectured (4.38) yields

$$\text{Var} \left(\int_{\mathbb{U}} f d\tilde{\nu}_t^N \right) \leq \sum_{j,k \in \mathbb{Z}} |\hat{f}(j)| |\hat{f}(k)| \left| \text{Cov}_{\rho_t^N} (\text{tr}[(\cdot)^j], \text{tr}[(\cdot)^k]) \right| \leq \frac{C(t)^2}{N^2} \left(\sum_{k \in \mathbb{Z}} |\hat{f}(k)| \right)^2.$$

We can then estimate this squared-sum as in (4.19): writing $|\hat{f}(k)| = (1 + k^2)^{-p/2} (1 + k^2)^{p/2} |\hat{f}(k)|$,

$$\left(\sum_{k \in \mathbb{Z}} |\hat{f}(k)| \right)^2 \leq \left(\sum_{k \in \mathbb{Z}} (1 + k^2)^{-p} \right) \cdot \|f\|_{H_p(\mathbb{U})}^2,$$

and this sum is finite provided $p > \frac{1}{2}$. To summarize, if the conjectured bound (4.37) holds true, then we have

$$\text{Var} \left(\int_{\mathbb{U}} f d\tilde{\nu}_t^N \right) = O \left(\frac{1}{N^2} \right), \quad \text{if } f \in H_p(\mathbb{U}) \text{ for some } p > \frac{1}{2}. \quad (4.39)$$

In [30, Theorem 2.6 & Proposition 9.9], the authors showed that, if $f \in H_{1/2}(\mathbb{U})$ is real-valued, then the fluctuations of the empirical integral are $O(1/N^2)$ -Gaussian with variance close to $\|f\|_{H_{1/2}(\mathbb{U})}^2$ for large t :

$$N \left[\int_{\mathbb{U}} f d\tilde{\nu}_t^N - \mathbb{E} \left(\int_{\mathbb{U}} f d\tilde{\nu}_t^N \right) \right] \xrightarrow[N \rightarrow \infty]{(d)} \mathcal{N}(0, \sigma_t(f)), \quad \lim_{t \rightarrow \infty} \sigma_t(f) = \|f\|_{H_{1/2}(\mathbb{U})}^2.$$

We see from here that, at least as $t \rightarrow \infty$, we have $N^2 \text{Var} \left(\int_{\mathbb{U}} f d\tilde{\nu}_t^N \right) \sim \|f\|_{H_{1/2}(\mathbb{U})}^2$. Thus, we cannot expect the conjectural $O(1/N^2)$ -behavior of (4.39) to hold for $f \notin H_{1/2}(\mathbb{U})$, and so this is the minimal regularity needed for this rate of convergence.

4.3 Empirical Eigenvalues on \mathbb{GL}_N

We begin with the following observation: for *holomorphic* trace polynomials, $P \circ \mathbf{V}_N$ with $P \in \mathcal{HP}$, $\mathcal{D}_{s,t}$ reduces to $\mathcal{D}_{s-t,0}$.

Lemma 4.17. *For $s, t > 0$ with $s > t/2$, and for $P \in \mathcal{HP}$,*

$$\begin{aligned} (e^{-\mathcal{D}_{s,t}^N} P)(\mathbf{1}) &= (e^{-\mathcal{D}_{s-t,0}^N} P)(\mathbf{1}), \\ (e^{-\mathcal{D}_{s,t}^N} P)(\mathbf{1}) &= (e^{-\mathcal{D}_{s-t,0}^N} P)(\mathbf{1}). \end{aligned} \quad (4.40)$$

Proof. For $P \in \mathcal{HP}$, the function $Z \mapsto P \circ \mathbf{V}_N(Z)$, $Z \in \mathbb{GL}_N$, is a trace polynomial in Z and not Z^* ; hence, it is holomorphic on \mathbb{GL}_N . For any holomorphic function f and any $X \in \mathfrak{u}_N$,

$$(\partial_{iX} f)(Z) = \frac{d}{dt} \Big|_{t=0} f(Ze^{itX}) = i(\partial_X f)(Z).$$

Hence $\partial_{iX}^2 f = -\partial_X^2 f$, and so (2.6) yields

$$A_{s,t}^N f = \left(s - \frac{t}{2} \right) \sum_{X \in \beta_N} \partial_X^2 f + \frac{t}{2} \sum_{X \in \beta_N} \partial_{iX}^2 f = (s-t) \sum_{X \in \beta_N} \partial_X^2 f = (s-t) \Delta_{\mathbb{U}_N} f.$$

Applying the intertwining formulas (3.6) and (3.10) now shows that

$$(e^{-\mathcal{D}_{s,t}^N} P) \circ \mathbf{V}_N = (e^{-\mathcal{D}_{s-t,0}^N} P) \circ \mathbf{V}_N.$$

holds for all N . Evaluating both sides at I_N gives $(e^{-\mathcal{D}_{s,t}^N} P)(\mathbf{1}) = (e^{-\mathcal{D}_{s-t,0}^N} P)(\mathbf{1})$. Taking the limit as $N \rightarrow \infty$ (using Corollary 3.14) now proves (4.40). \square

This brings us to the proof of Theorem 1.3, which we break into two propositions.

Proposition 4.18. *Let $s, t > 0$ with $s > t/2$. Fix $\delta > 0$ and $f \in G_{\frac{s}{2}(1+2\delta)}(\mathbb{C}^*)$. Then*

$$\left| \mathbb{E} \left(\int_{\mathbb{C}^*} f d\tilde{\phi}_{s,t}^N \right) - \int f d\nu_{s-t} \right| \leq \frac{1}{N^2} \cdot \frac{1}{\delta} \left(1 + \frac{1}{2} \sqrt{\frac{\pi}{s\delta}} \right)^{1/2} \|f\|_{G_{\frac{s}{2}(1+2\delta)}}. \quad (4.41)$$

Proof. The random variable $\int_{\mathbb{C}^*} f d\tilde{\phi}_{s,t}^N$ is given by

$$\left(\int_{\mathbb{C}^*} f d\tilde{\phi}_{s,t}^N \right) (Z) = \sum_{k \in \mathbb{Z}} \hat{f}(k) \text{tr}(Z^k), \quad Z \in \mathbb{GL}_N, \quad (4.42)$$

which converges since, for any fixed Z , $|\text{tr}(Z^k)|$ grows only exponentially in k , while by assumption $\hat{f}(k)$ decays super-exponentially fast. Note that

$$\int f d\nu_{s-t} = \sum_{k \in \mathbb{Z}} \hat{f}(k) \nu_k(s-t) = \sum_{k \in \mathbb{Z}} \hat{f}(k) (e^{-\mathcal{D}_{s-t,0}^N} v_k)(\mathbf{1}), \quad (4.43)$$

which converges as above since the $\nu_k(s-t)$ has only exponential growth. Formally, we also have

$$\begin{aligned} \mathbb{E} \left(\int_{\mathbb{C}^*} f d\tilde{\phi}_{s,t}^N \right) &= \sum_{k \in \mathbb{Z}} \hat{f}(k) \int_{\mathbb{GL}_N} \text{tr}(Z^k) \mu_{s,t}^N(dZ) \\ &= \sum_{k \in \mathbb{Z}} \hat{f}(k) (e^{-\mathcal{D}_{s,t}^N} v_k)(\mathbf{1}) = \sum_{k \in \mathbb{Z}} \hat{f}(k) (e^{-\mathcal{D}_{s-t,0}^N} v_k)(\mathbf{1}), \end{aligned} \quad (4.44)$$

by Lemma 4.17. The convergence of this series will follow from (4.41), which we now proceed to prove. Comparing (4.43) and (4.44),

$$\left| \mathbb{E} \left(\int_{\mathbb{C}^*} f d\tilde{\phi}_{s,t}^N \right) - \int f d\nu_{s-t} \right| \leq \sum_{k \in \mathbb{Z}} |\hat{f}(k)| \left| (e^{-\mathcal{D}_{s,t}^N} v_k)(\mathbf{1}) - (e^{-\mathcal{D}_{s-t,0}^N} v_k)(\mathbf{1}) \right|. \quad (4.45)$$

We bound these terms using Corollary 3.19:

$$\left| (e^{-\mathcal{D}_{s,t}^N} v_k)(\mathbf{1}) - (e^{-\mathcal{D}_{s-t,0}^N} v_k)(\mathbf{1}) \right| \leq \frac{1}{N^2} \cdot \frac{1}{\delta} e^{\frac{s}{2}(1+\delta)k^2} \|v_k\|_1,$$

which holds true whenever $N > \sqrt{2/\delta}$; note also that $\|v_k\|_1 = 1$. Thus (4.45) implies that

$$\left| \mathbb{E} \left(\int_{\mathbb{C}^*} f d\tilde{\phi}_{s,t}^N \right) - \int f d\nu_{s-t} \right| \leq \frac{1}{N^2} \cdot \frac{1}{\delta} \sum_{k \in \mathbb{Z}} |\hat{f}(k)| e^{\frac{s}{2}(1+\delta)k^2},$$

and this sum is bounded by

$$\begin{aligned} \sum_{k \in \mathbb{Z}} e^{-\frac{s}{2}\delta k^2} |\hat{f}(k)| e^{\frac{s}{2}(1+\delta)k^2} &\leq \left(\sum_{k \in \mathbb{Z}} e^{-s\delta k^2} \right)^{\frac{1}{2}} \|f\|_{G_{\frac{s}{2}(1+2\delta)}} \\ &\leq \left(1 + \frac{1}{2} \sqrt{\frac{\pi}{s\delta}} \right)^{\frac{1}{2}} \|f\|_{G_{\frac{s}{2}(1+2\delta)}} \end{aligned} \quad (4.46)$$

where we have made the estimate

$$\sum_{j=-n}^n e^{-s\delta j^2} \leq 1 + 2 \int_0^\infty e^{-s\delta x^2} dx = 1 + \frac{1}{2} \sqrt{\frac{\pi}{s\delta}}.$$

This proves (4.41). \square

Remark 4.19. In (4.45), we have used Lemma 4.17 to convert $\mathcal{D}_{s-t,0}^N$ and $\mathcal{D}_{s-t,0}$ back to $\mathcal{D}_{s,t}^N$ and $\mathcal{D}_{s,t}$ to apply Corollary 3.19. We could instead have used that corollary with $r = |s - t|$ (or $r = \epsilon$ for some $\epsilon > 0$ in the case $s = t$) to show the same result with the milder assumption that $f \in G_{\frac{r}{2}(1+2\delta)}$. This is not possible in Proposition 4.20 below where covariances are used, thus destroying the holomorphic structure; we have kept the regularity conditions consistent between the two.

Proposition 4.20. *Let $s, t > 0$ with $s > t/2$. Fix $\delta > 0$ and $f \in G_{s(1+2\delta)}(\mathbb{C}^*)$. Then, for $N > \sqrt{2/\delta}$,*

$$\text{Var} \left(\int_{\mathbb{C}^*} f d\tilde{\phi}_{s,t}^N \right) \leq \frac{1}{N^2} \cdot \frac{4}{\delta^2} \left(1 + \frac{1}{2} \sqrt{\frac{\pi}{2s\delta}} \right) \|f\|_{G_{s(1+2\delta)}}^2. \quad (4.47)$$

Proof. Starting from (4.42), we expand

$$\text{Var} \left(\int_{\mathbb{C}^*} f d\tilde{\phi}_{s,t}^N \right) = \sum_{j,k \in \mathbb{Z}} \hat{f}(j) \overline{\hat{f}(k)} \text{Cov}_{\mu_{s,t}^N} (\text{tr}[(\cdot)^j], \text{tr}[(\cdot)^k]). \quad (4.48)$$

Note that $\text{tr}(Z^k) = v_k \circ \mathbf{V}_N(Z)$. Since $v_k \in \mathcal{P}_{|k|}$ and $\|v_k\|_1 = 1$, Corollary 4.16 shows that

$$\left| \text{Cov}_{\mu_{s,t}^N} (v_j \circ \mathbf{V}_N, v_k \circ \mathbf{V}_N) \right| \leq \frac{1}{N^2} \cdot \frac{4}{\delta^2} e^{s(1+\delta)(j^2+k^2)}. \quad (4.49)$$

Combining this with (4.48) yields

$$\begin{aligned} \text{Var} \left(\int_{\mathbb{C}^*} f d\tilde{\phi}_{s,t}^N \right) &\leq \frac{1}{N^2} \cdot \frac{4}{\delta^2} \sum_{j,k \in \mathbb{Z}} |\hat{f}(j)| |\hat{f}(k)| e^{s(1+\delta)(j^2+k^2)} \\ &= \frac{1}{N^2} \cdot \frac{4}{\delta^2} \left(\sum_{k \in \mathbb{Z}} |\hat{f}(k)| e^{s(1+\delta)k^2} \right)^2, \end{aligned} \quad (4.50)$$

and the result follows from (4.46) with r replaced by $2s$. \square

Thus, we have the ingredients to prove Theorem 1.3.

Proof of Theorem 1.3. Since $\sigma > s$, $\delta = \frac{1}{2}(\frac{\sigma}{s} - 1) > 0$ and $s(1 + 2\delta) = \sigma$. Thus Proposition 4.18 proves (1.6) with a constant that depends continuously on s (note here that $G_{\frac{s}{2}(1+2\delta)} = G_{\sigma/2} \subset G_\sigma$), and Proposition 4.20 similarly proves (1.7). \square

4.4 Empirical Singular Values on \mathbb{GL}_N

As in Section 4.3, we begin by noting a reduction in the action of the generator $\mathcal{D}_{s,t}$ of the noncommutative distribution $\varphi_{s,t}$ when restricted, in this case, to holomorphic trace polynomials in ZZ^* . For this we need some new notation.

Notation 4.21. For $k \in \mathbb{Z} \setminus \{0\}$, let $\varepsilon^{1^*}(k) = \overbrace{(1, *, \dots, 1, *)}^{2k}$ if $k > 0$ and $\varepsilon^{1^*}(k) = \overbrace{(-1, -, \dots, -1, -)}^{2|k|}$ if $k < 0$; note that $|\varepsilon^{1^*}(k)| = 2|k|$. Denote $\mathcal{E}^{1^*} = \{\varepsilon^{1^*}(k) : k \in \mathbb{Z} \setminus \{0\}\} \subset \mathcal{E}$. Let $\mathcal{P}^{1^*} \subset \mathcal{P}$ be the subalgebra of polynomials

$$\mathcal{P}^{1^*} = \mathbb{C}[\{v_\varepsilon\}_{\varepsilon \in \mathcal{E}^{1^*}}].$$

For convenience denote $v_k^{1^*} = v_{\varepsilon^{1^*}(k)}$.

The homomorphism $\Phi_{1^*} : \mathcal{H}\mathcal{P} \rightarrow \mathcal{P}^{1^*}$ determined by $\Phi_{1^*}(v_k) = v_k^{1^*}$ is an algebra isomorphism. Note that, for any $N \in \mathbb{N}$,

$$\Phi_{1^*}(P) \circ \mathbf{V}_N = (P \circ \mathbf{V}_N) \circ \Phi, \quad (4.51)$$

where $\Phi(Z) = ZZ^*$ is the map from Definition 1.4.

Lemma 4.22. Let $s, t > 0$ with $s > t/2$. For $P \in \mathcal{H}\mathcal{P}$ and $N \in \mathbb{N}$,

$$\begin{aligned} (e^{-\mathcal{D}_{s,t}^N} \Phi_{1^*}(P))(\mathbf{1}) &= (e^{\mathcal{D}_{2t,0}^N} P)(\mathbf{1}), \\ (e^{-\mathcal{D}_{s,t}} \Phi_{1^*}(P))(\mathbf{1}) &= (e^{\mathcal{D}_{2t,0}} P)(\mathbf{1}). \end{aligned} \quad (4.52)$$

Proof. For $Z \in \mathbb{GL}_N$ and $X \in \mathfrak{u}_N$, note that

$$Ze^{tX}(Ze^{tX})^* = Ze^{tX}e^{-tX}Z^* = ZZ^*, \quad Ze^{itX}(Ze^{itX})^* = Ze^{2itX}Z^*. \quad (4.53)$$

Let $f : \mathbb{GL}_N \rightarrow \mathbb{C}$ be holomorphic. The first equation in (4.53) shows that $\partial_X(f \circ \Phi) = 0$, and so in particular the first terms $\sum_{X \in \beta_N} \partial_X^2(f \circ \Phi) = 0$ in (2.6). For the second terms,

$$\partial_{iX}(f \circ \Phi)(Z) = \left. \frac{d}{dt} \right|_{t=0} f(Ze^{itX}(Ze^{itX})^*) = \left. \frac{d}{dt} \right|_{t=0} f(Ze^{2itX}Z^*),$$

and so

$$\partial_{iX}^2(f \circ \Phi)(Z) = \left. \frac{\partial^2}{\partial s \partial t} \right|_{s,t=0} f(Ze^{isX}e^{2itX}e^{isX}Z^*) = \left. \frac{\partial^2}{\partial s \partial t} \right|_{s,t=0} f(Ze^{2i(s+t)X}Z^*).$$

If we additionally assume that f is tracial, $f(ZW) = f(WZ)$ for all $Z, W \in \mathbb{GL}_N$ (for example if f is a holomorphic trace polynomial $f = P \circ \mathbf{V}_N$ for some $P \in \mathcal{H}\mathcal{P}$), then

$$\partial_{iX}^2(f \circ \Phi)(Z) = \left. \frac{\partial^2}{\partial s \partial t} \right|_{s,t=0} f(Z^*Ze^{2i(s+t)X}) = 4 \left. \frac{\partial^2}{\partial s \partial t} \right|_{s,t=0} f(Z^*Ze^{i(s+t)X}). \quad (4.54)$$

By comparison,

$$-\partial_X^2 f(Z) = \partial_{iX}^2 f(Z) = \left. \frac{\partial^2}{\partial s \partial t} \right|_{s,t=0} f(Ze^{i(s+t)X})$$

and so we have

$$\partial_{iX}^2(f \circ \Phi) = -4(\partial_X^2 f) \circ \Phi^\perp \quad (4.55)$$

where $\Phi^\perp(Z) = Z^*Z$. Hence, from (2.6), we have

$$\frac{1}{2}A_{s,t}^N(f \circ \Phi) = \frac{t}{4} \sum_{X \in \beta_N} -4(\partial_X^2 f) \circ \Phi^\perp = -t(\Delta_{\mathfrak{u}_N} f) \circ \Phi^\perp = -t(\Delta_{\mathfrak{u}_N} f) \circ \Phi, \quad (4.56)$$

where the last equality comes from the fact that $\Delta_{\mathbb{U}_N}$ preserves the class of smooth tracial functions. (This follows from its bi-invariance, so it commutes with the left- and right-actions of the group; in our case, where f will be a trace polynomial, it follows from the fact that $\Delta_{\mathbb{U}_N}$ preserves trace polynomials.) Hence, taking $f = P \circ \mathbf{V}_N$ for some $P \in \mathcal{H}\mathcal{P}$ and using (3.6), (3.10), and (4.51), we have

$$(\mathcal{D}_{s,t}^N \Phi_{1*}(P)) \circ \mathbf{V}_N = -2t \Phi_{1*}(\mathcal{D}_{1,0}^N P) \circ \mathbf{V}_N = \Phi_{1*}(-\mathcal{D}_{2t,0}^N P) \circ \mathbf{V}_N.$$

Since Φ_{1*} is an algebra isomorphism, it follows that

$$(e^{-\mathcal{D}_{s,t}^N} \Phi_{1*}(P)) \circ \mathbf{V}_N = \Phi_{1*}(e^{\mathcal{D}_{2t,0}^N} P) \circ \mathbf{V}_N. \quad (4.57)$$

Evaluating both sides at I_N gives

$$(e^{-\mathcal{D}_{s,t}^N} \Phi_{1*}(P))(\mathbf{1}) = \Phi_{1*}(e^{\mathcal{D}_{2t,0}^N} P)(\mathbf{1}) = (e^{\mathcal{D}_{2t,0}^N} P)(\mathbf{1}),$$

the last equality following from the general fact that $(\Phi_{1*}(Q))(\mathbf{1}) = Q(\mathbf{1})$. Now letting $N \rightarrow \infty$ proves the lemma. \square

We now approach the proof of Theorem 1.5 as we did for Theorem 1.3. We begin by verifying (1.8).

Proposition 4.23. *Let $s, t > 0$ with $s > t/2$. Fix $\delta > 0$, and let $f \in G_{\frac{s}{2}(1+2\delta)}(\mathbb{C}^*)$. Then*

$$\left| \mathbb{E} \left(\int_{\mathbb{C}^*} f d\tilde{\eta}_{s,t}^N \right) - \int f d\nu_{-2t} \right| \leq \frac{1}{N^2} \cdot \frac{1}{\delta} \left(1 + \frac{1}{2} \sqrt{\frac{\pi}{s\delta}} \right)^{\frac{1}{2}} \|f\|_{G_{\frac{s}{2}(1+2\delta)}}. \quad (4.58)$$

Proof. The random variable $\int_{\mathbb{C}^*} f d\tilde{\eta}_{s,t}^N$ is given by

$$\left(\int_{\mathbb{C}^*} f d\tilde{\eta}_{s,t}^N \right) = \sum_{k \in \mathbb{Z}} \hat{f}(k) \text{tr}(Y^k), \quad Y \in \mathbb{M}_N^{>0}, \quad (4.59)$$

which converges since, for any fixed Y , $|\text{tr}(Y^k)|$ grows only exponentially in k , while by assumption $\hat{f}(k)$ decays super-exponentially fast. We also have

$$\int f d\nu_{-2t} = \sum_{k \in \mathbb{Z}} \hat{f}(k) \nu_k(-2t) = \sum_{k \in \mathbb{Z}} \hat{f}(k) (e^{\mathcal{D}_{2t,0}^N} \nu_k)(\mathbf{1}), \quad (4.60)$$

which converges as above since $\nu_k(-2t)$ have only exponential growth (being the moments of a compactly-supported probability measure). By definition, subject to convergence,

$$\begin{aligned} \mathbb{E} \left(\int_{\mathbb{C}^*} f d\tilde{\eta}_{s,t}^N \right) &= \sum_{k \in \mathbb{Z}} \hat{f}(k) \int_{\text{GL}_N} \text{tr}(\Phi(Z)^k) \mu_{s,t}^N(dZ) \\ &= \sum_{k \in \mathbb{Z}} \hat{f}(k) (e^{-\mathcal{D}_{s,t}^N} \Phi_{1*}(\nu_k))(\mathbf{1}) = \sum_{k \in \mathbb{Z}} \hat{f}(k) (e^{\mathcal{D}_{2t,0}^N} \nu_k)(\mathbf{1}), \end{aligned} \quad (4.61)$$

by (4.51) and Lemma 4.22. The convergence of this series will follow from (4.58), which we now proceed to prove. Comparing (4.60) and (4.61),

$$\left| \mathbb{E} \left(\int_{\mathbb{C}^*} f d\tilde{\eta}_{s,t}^N \right) - \int f d\nu_{-2t} \right| \leq \sum_{k \in \mathbb{Z}} |\hat{f}(k)| \left| (e^{\mathcal{D}_{2t,0}^N} \nu_k)(\mathbf{1}) - (e^{\mathcal{D}_{2t,0}^N} \nu_k)(\mathbf{1}) \right|. \quad (4.62)$$

The remainder of the proof proceeds exactly as in the proof of Proposition 4.41, following (4.45). \square

Proposition 4.24. *Let $s, t > 0$ with $s > t/2$. Fix $\delta > 0$ and $f \in G_{4s(1+2\delta)}(\mathbb{C}^*)$. Then, for $N > \sqrt{2/\delta}$,*

$$\text{Var} \left(\int_{\mathbb{C}^*} f d\tilde{\eta}_{s,t}^N \right) \leq \frac{1}{N^2} \cdot \frac{4}{\delta^2} \left(1 + \frac{1}{2} \sqrt{\frac{\pi}{8s\delta}} \right) \|f\|_{G_{4s(1+2\delta)}}^2. \quad (4.63)$$

Proof. As in (4.48), we begin by expanding the variance from (4.59) as follows:

$$\text{Var} \left(\int_{\mathbb{C}^*} f d\tilde{\eta}_{s,t}^N \right) = \sum_{j,k \in \mathbb{Z}} \hat{f}(k) \overline{\hat{f}(k)} \text{Cov}_{\Phi_*(\mu_{s,t}^N)}(\text{tr}[(\cdot)^j], \text{tr}[(\cdot)^k]). \quad (4.64)$$

By definition, for any random variables F, G on $\mathbb{M}_N^{\geq 0}$,

$$\text{Cov}_{\Phi_*(\mu_{s,t}^N)}(F, G) = \text{Cov}_{\mu_{s,t}^N}(F \circ \Phi, G \circ \Phi).$$

With $F(Y) = \text{tr}(Y^k)$, we have $F \circ \Phi = (v_k \circ \mathbf{V}_N) \circ \Phi = \Phi_{1*}(v_k) \circ \mathbf{V}_N$ by (4.51), and so the covariances in (4.64) are

$$\left| \text{Cov}_{\mu_{s,t}^N}(v_j^{1*} \circ \mathbf{V}_N, v_k^{1*} \circ \mathbf{V}_N) \right| \leq \frac{1}{N^2} \cdot \frac{4}{\delta^2} e^{s(1+\delta)((2j)^2+(2k)^2)}$$

by Corollary 4.16, since $\deg(v_k^{1*}) = 2|k|$. The remainder of the proof follows exactly as in the proof of Proposition 4.20, following (4.49). \square

This brings us to the proof of Theorem 1.5

Proof of Theorem 1.5. Since $\sigma > 4s$, $\delta = \frac{1}{2}(\frac{\sigma}{4s} - 1) > 0$ and $4s(1+2\delta) = \sigma$. Thus Proposition 4.23 proves (1.8) with a constant that depends continuously on s (note here that $G_{\frac{\sigma}{2}(1+2\delta)} = G_{\sigma/8} \subset G_\sigma$). Similarly, Proposition 4.24 proves (1.9). \square

5 L^p Convergence

In this final section, we observe that the techniques developed in Section 3.2 in fact yield, with little extra effort, convergence in a sense significantly stronger than those given in Theorems 1.2–1.7. We begin with a brief discussion of strong convergence.

5.1 Strong Convergence and Noncommutative L^p -norms

Let ρ^N be a probability measure on \mathbb{M}_N . Suppose that the noncommutative empirical distribution $\tilde{\varphi}^N$ of ρ^N has an almost-sure limit distribution φ , in the sense of Definition 2.15. In other words, if A_N is a random matrix with distribution ρ^N , we have $\varphi_{A_N}(f) \rightarrow \varphi(f)$ *a.s.* for all noncommutative polynomials $f \in \mathbb{C}\langle A, A^* \rangle$. The following stronger form of convergence has significant applications in operator algebras.

Definition 5.1 (Strong Convergence). *For each N , let ρ^N be a probability measure on \mathbb{M}_N , and let A_N be a random matrix with distribution ρ^N . Say that A_N **converges strongly** if it converges in distribution and in operator norm almost surely. That is: there exists a C^* -probability space (\mathcal{A}, τ) , and an element $a \in \mathcal{A}$, such that, for any noncommutative polynomial $f \in \mathbb{C}\langle A, A^* \rangle$,*

$$\text{tr}[f(A_N, A_N^*)] \rightarrow \tau[f(a, a^*)] \text{ a.s.} \quad \text{and} \quad \|f(A_N, A_N^*)\|_{\mathbb{M}_N} \rightarrow \|f(a, a^*)\|_{\mathcal{A}} \text{ a.s.} \quad (5.1)$$

Definition 5.1 naturally generalizes to multivariate noncommutative distributions. In their seminal paper [22], Haagerup and Thorbjørnsen showed that if ρ^N is (a finite product of) the GUE_N measure (1.18), then the independent GUE_N random matrices with this distribution converge strongly. More recently, in [13], the authors showed that strong convergence also holds for (finite products of) the Haar measure on \mathbb{U}_N . Given our mantra that the heat kernel measure ρ_t^N on \mathbb{U}_N interpolates between these two ensembles, it is natural to ask whether the matrices U_t^N also exhibit strong convergence. By extension, we may also ask whether random matrices $Z_{s,t}^N$ also exhibit strong convergence (now that we have proved, in Theorem 1.7, that they have an almost-sure limit distribution).

Note that, for any matrix $A \in \mathbb{M}_N$, $\|A\| = \lim_{q \rightarrow \infty} (\text{tr}[(AA^*)^q])^{1/2q}$; since $AA^* \in \mathbb{M}_N^{\geq 0}$ this makes sense for all real $q > 0$, but for convenience we may restrict q to be an integer. In fact, the same holds true in any *faithful* noncommutative C^* -probability space (\mathcal{A}, τ) :

$$\|a\|_{\mathcal{A}} = \lim_{q \rightarrow \infty} (\tau[(aa^*)^q])^{1/2q}.$$

These are (limits of) the **noncommutative L^p -norms** over (\mathcal{A}, τ) :

$$\|a\|_{L^p(\mathcal{A}, \tau)} \equiv (\tau[(aa^*)^{p/2}])^{1/p}. \quad (5.2)$$

$\|\cdot\|_{L^p(\mathcal{A}, \tau)}$ is a norm on \mathcal{A} for $p \geq 1$. In the case that \mathcal{A} is a W^* -algebra, its completion $L^p(\mathcal{A}, \tau)$ can be realized as a space of unbounded operators affiliated to \mathcal{A} when $p < \infty$, while $L^\infty(\mathcal{A}, \tau) = \mathcal{A}$.

The second statement in (5.1) can thus be rephrased as an almost sure interchange of limits: since $(\mathbb{M}_N, \text{tr})$ is a faithful C^* -probability space, then $A_N \in \mathbb{M}_N$ converges to $a \in \mathcal{A}$ strongly if and only if $\varphi_{A_N} \rightarrow \varphi_a$ a.s. and

$$\mathbb{P} \left(\lim_{N \rightarrow \infty} \lim_{p \rightarrow \infty} \|f(A_N, A_N^*)\|_{L^p(\mathbb{M}_N, \text{tr})} = \lim_{p \rightarrow \infty} \|f(a, a^*)\|_{L^p(\mathcal{A}, \tau)} \right) = 1, \quad (5.3)$$

provided that (\mathcal{A}, τ) is a faithful C^* -probability space.

5.2 Almost Sure L^p Convergence

Theorem 1.7 establishes that the random matrices U_t^N and $Z_{s,t}^N$ converge weakly almost surely to limit noncommutative distributions. Indeed, the U_t^N case (of convergence in expectation) is the main theorem in [9], where it is shown that, if U_t^N is chosen to be a Brownian motion on \mathbb{U}_N , then the weak limit exists as a noncommutative stochastic process, the free unitary Brownian motion discussed at the end of Section 2.5. In this case, the limit noncommutative probability space can be taken as a free group factor, and so is indeed a faithful C^* -probability space. As for $Z_{s,t}^N$, Definition 4.11 and the subsequent discussion show how to realize the almost sure limit noncommutative distribution $\varphi_{s,t}$ as the distribution of an operator $\varphi_{s,t} = \varphi_{z_{s,t}}$ on a noncommutative probability space $(\mathcal{A}_{s,t}, \tau_{s,t})$; the main theorem of the author's complementary paper [26] demonstrates that $\mathcal{A}_{s,t}$ may be taken to be a free group factor, with $\tau_{s,t}$ the usual faithful tracial state. As such, we can construct a larger C^* -probability space that contains both of the limit operators u_t and $z_{s,t}$. (By taking the reduced free product C^* -algebra of the two spaces, we can even make u_t and $z_{s,t}$ freely independent if we wish.) Thus, in the statement of Theorem 1.9, there is no loss of generality in realizing the limits in a single C^* -probability space (\mathcal{A}, τ) .

While we are, as yet, unable to prove strong convergence of U_t^N and $Z_{s,t}^N$ to u_t and $z_{s,t}$, we can prove almost sure L^p -convergence for all even integers p , i.e. Theorem 1.9. From (5.3), this result could be viewed as only infinitesimally weaker; but, in fairness, it does not suffice for the same powerful operator algebra techniques that strong convergence supports.

Once again, they key is a variance estimate, which follows easily from Proposition 4.13.

Lemma 5.2. *Let $s, t > 0$ with $s > t/2$, and let $f \in \mathbb{C}\langle A, A^* \rangle$ be a noncommutative polynomial. Let $p \geq 2$ be an even integer. Then, for $N \in \mathbb{N}$,*

$$\text{Var}\left(\|f(U_t^N, (U_t^N)^*)\|_{L^p(\mathbb{M}_N, \text{tr})}^p\right) = O\left(\frac{1}{N^2}\right) \quad \text{and} \quad \text{Var}\left(\|f(Z_{s,t}^N, (Z_{s,t}^N)^*)\|_{L^p(\mathbb{M}_N, \text{tr})}^p\right) = O\left(\frac{1}{N^2}\right).$$

Proof. We begin with the case of $Z_{s,t}^N$. The variance in question is

$$\text{Var}\left(\|f(Z_{s,t}^N, (Z_{s,t}^N)^*)\|_{L^p(\mathbb{M}_N, \text{tr})}^p\right) = \text{Var}_{\mu_{s,t}^N}(F^p), \quad (5.4)$$

where $F^p: \mathbb{G}\mathbb{L}_N \rightarrow \mathbb{C}$ is the random variable

$$F^p(Z) = \|f(Z, Z^*)\|_{L^p(\mathbb{M}_N, \text{tr})}^p = \text{tr}\left(\left(f(Z, Z^*)f(Z, Z^*)^*\right)^{p/2}\right).$$

Note that $g_p(A, A^*) = (f(A, A^*)f(A, A^*)^*)^{p/2}$ is an element of $\mathbb{C}\langle A, A^* \rangle$. Thus using the inclusion Υ of $\mathbb{C}\langle A, A^* \rangle \hookrightarrow \mathcal{P}^+$ (2.31), we have

$$F^p(Z) = \Upsilon(g_p) \circ \mathbf{V}_N(Z). \quad (5.5)$$

By Proposition 4.13,

$$\text{Var}_{\mu_{s,t}^N}(\Upsilon(g_p) \circ \mathbf{V}_N) \leq \frac{1}{N^2} \cdot C_2(s, t, \Upsilon(g_p), \Upsilon(g_p)), \quad (5.6)$$

and this, together with (5.4) and (5.5), proves the lemma for $Z_{s,t}^N$. The statement for U_t^N actually follows as a special case. Indeed, for any $P \in \mathcal{P}$, (3.13) and (3.16) show that

$$\text{Var}_{\rho_t^N}(P \circ \mathbf{V}_N) = (e^{-\mathcal{D}_{t,0}^N}(PP^*))(\mathbf{1}) - (e^{-\mathcal{D}_{t,0}^N}P)(\mathbf{1})(e^{-\mathcal{D}_{t,0}^N}P^*)(\mathbf{1}). \quad (5.7)$$

Proposition 4.13 is proved by showing that this quantity, with $\mathcal{D}_{s,t}^N$ in place of $\mathcal{D}_{t,0}^N$, is $\leq C_2(s, t, P, P)/N^2$. Although we must have $s, t > 0$ and $s > t/2$ for $\mu_{s,t}^N$ to be a well-defined measure, the operators $e^{-\mathcal{D}_{s,t}^N}$, and ergo the quantities in (5.7) and the constant $C_2(s, t, P, P)$, are all well-defined for $s, t \in \mathbb{R}$. Thus, we may restrict (5.6) to find

$$\text{Var}_{\rho_t^N}(F^p) = \text{Var}_{\rho_t^N}(\Upsilon(g_p) \circ \mathbf{V}_N) \leq \frac{1}{N^2} \cdot C_2(t, 0, \Upsilon(g_p), \Upsilon(g_p)), \quad (5.8)$$

and this proves the U_t^N -case of the lemma. \square

Remark 5.3. The size of the constant $C_2(t, 0, P, P)$ has only been shown (Corollary 4.16) to be bounded (almost) by $e^{2t \cdot \deg(P)^2} \|P\|_1^2$. We conjecture (as in (4.38)) that the growth with $\deg(P)$ is erroneous; but the dependence on $\|P\|_1$ is surely not. It is relatively straightforward to calculate that, with g_p defined from f as in the proof of Lemma 5.2,

$$\|\Upsilon(g_p)\|_1 = \|\Upsilon(f)\|_1^p.$$

This is not unexpected, since the L^p -norm itself is the p th root of the quantities considered here.

This brings us, finally, to the proof of Theorem 1.9.

Proof of Theorem 1.9. The almost sure weak convergence of $Z_{s,t}^N$ to $z_{s,t}$ was established in Theorem 1.7; U_t^N follows as the special case $Z_{t,0}^N$ (and was established already in [32]). It follows that, for any $f \in \mathbb{C}\langle A, A^* \rangle$,

$$\begin{aligned} \mathbb{E}\left(\|f(U_t^N, (U_t^N)^*)\|_{L^p(\mathbb{M}_N, \text{tr})}^p\right) &\rightarrow \|f(u_t, u_t^*)\|_{L^p(\mathcal{A}, \tau)}^p, \quad \text{and} \\ \mathbb{E}\left(\|f(Z_{s,t}^N, (Z_{s,t}^N)^*)\|_{L^p(\mathbb{M}_N, \text{tr})}^p\right) &\rightarrow \|f(z_{s,t}, z_{s,t}^*)\|_{L^p(\mathcal{A}, \tau)}^p, \end{aligned}$$

since these quantities (raised to the p th power as they are) are trace polynomials in U_t^N (resp. $Z_{s,t}^N$) and u_t (resp. $z_{s,t}$). Lemma 5.2, together with Chebyshev's inequality and the Borel-Cantelli Lemma, now shows that

$$\begin{aligned} \|f(U_t^N, (U_t^N)^*)\|_{L^p(\mathbb{M}_N, \text{tr})}^p &\rightarrow \|f(u_t, u_t^*)\|_{L^p(\mathcal{A}, \tau)}^p \text{ a.s.} \quad \text{and} \\ \|f(Z_{s,t}^N, (Z_{s,t}^N)^*)\|_{L^p(\mathbb{M}_N, \text{tr})}^p &\rightarrow \|f(z_{s,t}, z_{s,t}^*)\|_{L^p(\mathcal{A}, \tau)}^p \text{ a.s.} \end{aligned}$$

The theorem now follows by taking p th roots. □

Remark 5.4. The above proof, coupled with Remark 5.3, shows that it is plausible that the rate of a.s. convergence in Theorem 1.9 is uniformly bounded in p (contingent on the conjectured trace degree-independence of the constants $C_2(t, 0, P, P)$) in the U_t^N -case. If this is true, then strong convergence $U_t^N \rightarrow u_t$ follows readily from (5.3). This is left as a promising avenue for future study.

Acknowledgments

The author wishes to thank Bruce Driver for many helpful and insightful conversations, particularly with regards to Section 4.1, and Brian Hall, for careful proofreading and clarifying remarks regarding Theorem 1.3.

References

- [1] ANDERSON, G. W., GUIONNET, A., AND ZEITOUNI, O. *An introduction to random matrices*, vol. 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010.
- [2] ARNOLD, L. On the asymptotic distribution of the eigenvalues of random matrices. *J. Math. Anal. Appl.* 20 (1967), 262–268.
- [3] ARNOLD, L. On Wigner's semicircle law for the eigenvalues of random matrices. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* 19 (1971), 191–198.
- [4] BAI, Z. D. Methodologies in spectral analysis of large-dimensional random matrices, a review. *Statist. Sinica* 9, 3 (1999), 611–677. With comments by G. J. Rodgers and Jack W. Silverstein; and a rejoinder by the author.
- [5] BELINSCHI, S. T., AND BERCOVICI, H. Atoms and regularity for measures in a partially defined free convolution semigroup. *Math. Z.* 248, 4 (2004), 665–674.
- [6] BELINSCHI, S. T., AND BERCOVICI, H. Partially defined semigroups relative to multiplicative free convolution. *Int. Math. Res. Not.*, 2 (2005), 65–101.
- [7] BERCOVICI, H., AND VOICULESCU, D. Lévy-Hinčin type theorems for multiplicative and additive free convolution. *Pacific J. Math.* 153, 2 (1992), 217–248.
- [8] BERCOVICI, H., AND VOICULESCU, D. Free convolution of measures with unbounded support. *Indiana Univ. Math. J.* 42, 3 (1993), 733–773.
- [9] BIANE, P. Free Brownian motion, free stochastic calculus and random matrices. In *Free probability theory (Waterloo, ON, 1995)*, vol. 12 of *Fields Inst. Commun.* Amer. Math. Soc., Providence, RI, 1997, pp. 1–19.
- [10] BIANE, P. Segal-Bargmann transform, functional calculus on matrix spaces and the theory of semi-circular and circular systems. *J. Funct. Anal.* 144, 1 (1997), 232–286.

- [11] CÉBRON, G. Free convolution operators and free Hall transform. *J. Funct. Anal.* 265, 11 (2013), 2645–2708.
- [12] COLLINS, B., AND KEMP, T. Liberation of projections. *J. Funct. Anal.* (2013). To appear.
- [13] COLLINS, B., AND MALE, C. The strong asymptotic freeness of haar and deterministic matrices. *Preprint* (January 2013). arXiv:1105.4345v2.
- [14] DIACONIS, P., AND EVANS, S. N. Linear functionals of eigenvalues of random matrices. *Trans. Amer. Math. Soc.* 353, 7 (2001), 2615–2633.
- [15] DRIVER, B. K., AND HALL, B. C. Yang-Mills theory and the Segal-Bargmann transform. *Comm. Math. Phys.* 201, 2 (1999), 249–290.
- [16] DRIVER, B. K., HALL, B. C., AND KEMP, T. The large- N limit of the Segal-Bargmann transform on \mathbb{U}_N . *J. Funct. Anal.* 265, 11 (2013), 2585–2644.
- [17] DUISTERMAAT, J. J., AND KOLK, J. A. C. *Lie groups*. Universitext. Springer-Verlag, Berlin, 2000.
- [18] EVANS, L. C. *Partial differential equations*, second ed., vol. 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2010.
- [19] FERRARI, A. B., AND TITI, E. S. Gevrey regularity for nonlinear analytic parabolic equations. *Comm. Partial Differential Equations* 23, 1-2 (1998), 1–16.
- [20] FOIAS, C., AND TEMAM, R. Gevrey class regularity for the solutions of the Navier-Stokes equations. *J. Funct. Anal.* 87, 2 (1989), 359–369.
- [21] GUIONNET, A., AND ZEITOUNI, O. Concentration of the spectral measure for large matrices. *Electron. Comm. Probab.* 5 (2000), 119–136 (electronic).
- [22] HAAGERUP, U., AND THORBJØRNSSEN, S. A new application of random matrices: $\text{Ext}(C_{\text{red}}^*(F_2))$ is not a group. *Ann. of Math. (2)* 162, 2 (2005), 711–775.
- [23] HALL, B. C. A new form of the Segal-Bargmann transform for Lie groups of compact type. *Canad. J. Math.* 51, 4 (1999), 816–834.
- [24] HALL, B. C. Harmonic analysis with respect to heat kernel measure. *Bull. Amer. Math. Soc. (N.S.)* 38, 1 (2001), 43–78 (electronic).
- [25] HALL, B. C. *Lie groups, Lie algebras, and representations*, vol. 222 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2003. An elementary introduction.
- [26] KEMP, T. The large- N limits of Brownian motions on $\mathbb{G}L_N$. *Preprint* (June 2013). arXiv:1306.6033.
- [27] KEMP, T., NOURDIN, I., PECCATI, G., AND SPEICHER, R. Wigner chaos and the fourth moment. *Ann. Probab.* 40, 4 (2012), 1577–1635.
- [28] LEVERMORE, C. D., AND OLIVER, M. Analyticity of solutions for a generalized Euler equation. *J. Differential Equations* 133, 2 (1997), 321–339.
- [29] LÉVY, T. Schur-Weyl duality and the heat kernel measure on the unitary group. *Adv. Math.* 218, 2 (2008), 537–575.

- [30] LÉVY, T., AND MAÏDA, M. Central limit theorem for the heat kernel measure on the unitary group. *J. Funct. Anal.* 259, 12 (2010), 3163–3204.
- [31] NICA, A., AND SPEICHER, R. *Lectures on the combinatorics of free probability*, vol. 335 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 2006.
- [32] RAINS, E. M. Combinatorial properties of Brownian motion on the compact classical groups. *J. Theoret. Probab.* 10, 3 (1997), 659–679.
- [33] ROBINSON, D. W. *Elliptic operators and Lie groups*. Oxford Mathematical Monographs. The Clarendon Press Oxford University Press, New York, 1991. Oxford Science Publications.
- [34] RUDIN, W. *Real and complex analysis*, third ed. McGraw-Hill Book Co., New York, 1987.
- [35] SALOFF-COSTE, L. *Aspects of Sobolev-type inequalities*, vol. 289 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 2002.
- [36] VOICULESCU, D. Multiplication of certain noncommuting random variables. *J. Operator Theory* 18, 2 (1987), 223–235.
- [37] WIGNER, E. P. Characteristic vectors of bordered matrices with infinite dimensions. *Ann. of Math. (2)* 62 (1955), 548–564.
- [38] WIGNER, E. P. Characteristic vectors of bordered matrices with infinite dimensions. II. *Ann. of Math. (2)* 65 (1957), 203–207.
- [39] WIGNER, E. P. On the distribution of the roots of certain symmetric matrices. *Ann. of Math. (2)* 67 (1958), 325–327.
- [40] ZHONG, P. On the free convolution with a free multiplicative analogue of the normal distribution. *Preprint* (April 2013). arXiv:1211.3.160v2.