# The Large-$N$ Limit of the Segal–Bargmann Transform on $U(N)$

Bruce K. Driver*
Department of Mathematics
University of California, San Diego
La Jolla, CA 92093-0112
bdriver@math.ucsd.edu

Brian C. Hall†
Department of Mathematics
University of Notre Dame
Notre Dame, IN 46556
bhall@nd.edu

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

May 11, 2013

**Abstract**

We study the (two-parameter) Segal–Bargmann transform $\mathbf{B}_{s,t}^N$ on the unitary group $U(N)$, for large $N$. Acting on matrix valued functions that are equivariant under the adjoint action of the group, the transform has a meaningful limit $\mathbf{B}_{s,t}$ as $N \to \infty$, which can be identified as an operator on the space of complex Laurent polynomials. We introduce the space of *trace Laurent polynomials*, and use it to give effective computational methods to determine the action of the heat operator, and thus the Segal–Bargmann transform. We prove several concentration of measure and limit theorems, giving a direct connection from the finite-dimensional transform $\mathbf{B}_{s,t}^N$ to its limit $\mathbf{B}_{s,t}$. We characterize the operator $\mathbf{B}_{s,t}$ through its inverse action on the standard polynomial basis. Finally, we show that, in the case $s = t$, the limit transform $\mathbf{B}_{t,t}$ is the free unitary Segal–Bargmann transform $\mathscr{G}^t$ introduced by Biane.

## Contents



*Supported in part by NSF Grant DMS-0804472
†Supported in part by NSF Grant DMS-1001328
‡Supported in part by NSF Grant DMS-1001894

# 1 Introduction

The *Segal–Bargmann transform* (also known in the physics literature as the *Bargmann transform* or *Coherent State transform*) is a unitary isomorphism from $L^2$ to holomorphic $L^2$. It was originally introduced by Segal [28, 29] and Bargmann [1, 2], as a map

$$S_t\colon L^2(\mathbb{R}^N, \gamma_t^N) \to \mathcal{H}L^2(\mathbb{C}^N, \gamma_{t/2}^{2N})$$

where $\gamma_t^N$ is the standard Gaussian heat kernel measure $(\frac{1}{4\pi t})^{N/2}\exp(-\frac{1}{4t}|\mathbf{x}|^2)\,d\mathbf{x}$ on $\mathbb{R}^N$, and $\mathcal{H}L^2$ denotes the subspace of square-integrable *holomorphic* functions. The transform $S_t$ is given by convolution with the heat kernel, followed by analytic continuation.

In [17], the second author introduced an analog of the Segal–Bargmann transform for any compact Lie group $K$. The transform maps functions on $K$ to holomorphic functions on the complexification $K_\mathbb{C}$ of $K$. We will be particularly interested in the case $K = U(N)$, where the complexification is $K_\mathbb{C} = GL(N, \mathbb{C})$. The "$B_t$" version of the transform, which is the one most relevant to the present paper, is a unitary map

$$B_t : L^2(K, \rho_t) \to \mathcal{H}L^2(K_\mathbb{C}, \mu_t)$$

where $\rho_t$ and $\mu_t$ are suitable heat kernel measures on $K$ and $K_\mathbb{C}$, based at the identity. Mirroring the Euclidean case, the transform itself is defined as follows:

$$B_t f = (e^{\frac{t}{2}\Delta_K} f)_\mathbb{C} \tag{1.1}$$

where $e^{\frac{t}{2}\Delta_K}$ is the time-$t$ (forward) heat operator on $K$ and $(\,\cdot\,)_\mathbb{C}$ denotes analytic continuation from $K$ to $K_\mathbb{C}$. More generally, given two positive numbers $s$ and $t$ with $s > t/2$, there is a two-parameter Segal–Bargmann transform, introduced by the first two authors in [10, 18]:

$$B_{s,t}\colon L^2(K, \rho_s) \to \mathcal{H}L^2(K_\mathbb{C}, \mu_{s,t}),$$

where $\mu_{s,t}$ is another heat kernel measure on $K_{\mathbb{C}}$. The transform $B_{s,t}$ is given by the same formula (1.1) as $B_t$; only the definition of the inner products on the domain and range spaces depends on the additional parameter $s$. The original transform $B_t$ coincides with $B_{t,t}$.

Since the classical Segal–Bargmann transform for Euclidean spaces admits an infinite dimensional version [30], it is natural to attempt to construct an infinite dimensional limit of the transform for compact Lie groups. One successful approach to such a limit is found in the paper [20] of the second author and A. N. Sengupta, in which they develop a version of the Segal–Bargmann transform for the path group with values in a compact Lie group $K$. The paper [20] is an extension of the work of L. Gross and P. Malliavin [15] and reflects the origins of the generalized Segal–Bargmann transform for compact Lie groups in the work of Gross [14].

A different approach to an infinite dimensional limit is to consider the transform on a nested family of compact Lie groups, such as $U(N)$ for $N = 1, 2, 3, \ldots$ To define the transform on $U(N)$, one must choose an $\mathrm{Ad}$-invariant inner-product on the Lie algebra $\mathfrak{u}(N)$ of $U(N)$: letting $M_N(\mathbb{C})$ denote the space of all $N \times N$ complex matrices, $\mathfrak{u}(N) = \{X \in M_N(\mathbb{C})\colon X^* = -X\}$. The most obvious approach to the $N \to \infty$ limit would be to use on each $\mathfrak{u}(N)$ a fixed (i.e. $N$-independent) multiple of the Hilbert–Schmidt norm. Work of M. Gordina [12, 13], however, showed that this approach does not work, because the target Hilbert space becomes undefined in the limit. Indeed, Gordina showed that, with the metrics normalized the this way, in the large-$N$ limit all nonconstant holomorphic functions on $GL(N, \mathbb{C})$ have infinite norm with respect to the heat kernel measure $\mu_t$.

An alternative approach to the $N \to \infty$ limit of the Segal–Bargmann transform on $U(N)$, suggested by Philippe Biane [4], is to scale the Hilbert–Schmidt norm on $\mathfrak{u}(N)$ by an $N$-dependent constant. Specifically, Biane proposed to use on $\mathfrak{u}(N)$ the norm $\|\cdot\|_{\mathfrak{u}(N)}$ given by

$$\|X\|_{\mathfrak{u}(N)}^2 = N\mathrm{Tr}\,(X^*X) = N\sum_{j,k=1}^{N} |X_{jk}|^2\,. \tag{1.2}$$

In the first part of [4], this scaling was used to successfully carry out a large-$N$ limit of the *Lie algebra version* of the transform. Taking the underlying space to be the Lie algebra $\mathfrak{u}(N)$ rather than the group $U(N)$, he considered a vector-valued version of the classical Euclidean Segal–Bargmann transform $S_t^N$ acting on functions $\mathfrak{u}(N) \to M_N(\mathbb{C})$. Specifically, if $P(x) = \sum_{k=0}^n a_n x^n$ is a single-variable polynomial, it gives rise to a function $P_N\colon \mathfrak{u}(N) \to M_N(\mathbb{C})$ in the usual way:

$$P_N(X) = a_0 I_N + \sum_{k=1}^{n} a_n X^n \in M_N(\mathbb{C}). \tag{1.3}$$

The transformed functions $S_t^N P_N$ have a limit (in an appropriate sense; cf. (1.4)) which can be thought of as a one-variable polynomial function $P^t\colon \mathbb{R} \to \mathbb{R}$. This defines a unitary transformation $\mathscr{F}^t\colon P \mapsto P^t$ [4, Theorem 3] on the limiting $L^2$ closure of polynomials with respect to the limit heat kernel measure—in this context Wigner's semicircle law.

*Remark* 1.1. The results of [4, Section 1] are formulated in terms of the large-$N$ limit of $S_t^N$ on the space $\mathscr{X}_N = i\mathfrak{u}(N)$ of Hermitian $N \times N$ matrices, which is of course equivalent to the formulation above. It also deals more generally with the class of *functional calculus* functions on $\mathscr{X}_N$; cf. Section 2.1. We will restrict our attention almost exclusively to the space of Laurent polynomial functions, that are dense in functional calculus. Section 2 also discusses *equivariant functions*: an extension of functional calculus which forms a natural domain for the Segal-Bargmann transform, and subsumes all other function spaces discussed in this paper.

Biane proceeded in [4] to construct a "large-$N$ limit $U(N)$ Segal–Bargmann transform" $\mathscr{G}^t$, not by taking this limit directly, but instead developing a free probabilistic version of the Malliavin calculus techniques used by Gross and Malliavin [15] to derive the properties of $B_t$ from an infinite dimensional version of $S_t$. This laid the foundation for the modern theory of free Malliavin calculus and free stochastic differential equations, subsequently studied in [5, 6, 21] and many other papers, and was groundbreaking in many respects. Biane conjectured that his transform $\mathscr{G}^t$ *is* the direct $N \to \infty$ limit of the Segal–Bargmann transforms $B_t^N$ on $U(N)$, and suggested that this could be proved using the methods of stochastic analysis, but left the details of such an argument out of [4] (see the Remark on page 263). One of the main motivations for the present paper is to prove that this connection indeed holds: we will show directly that $B_t^N$ (defined with the metric given in (1.2)), boosted to act on Laurent polynomial functions $U(N) \to M_N(\mathbb{C})$, converges to $\mathscr{G}^t$ as $N \to \infty$. Our methods and ideas are very different from those Biane suggested, however; they are analytic and geometric, rather than probabilistic. Moreover, we find the large-$N$ limit of the *two-parameter* Segal–Bargmann transform $B_{s,t}^N$, and this generalization is essential to our proof that $\lim_{N\to\infty} B_{t,t}^N = \mathscr{G}^t$.

*Remark* 1.2. It should be noted that, shortly before the present paper was completed, the preprint [16] by Cébron Guillaume appeared, addressing (in the special case $s = t$) some of the same questions we answer presently. On the one hand, there are some similar ideas in our paper and Guillaume's. In our Lemma 1.19 and Theorem 1.20, for example, we identify the action of the Laplacian in terms of a sum of two operators, one of which is of order $1/N^2$; this result is conceptually very similar to [16, Lemma 4.1], although expressed in a different framework. On the other hand, our method for connecting the large-$N$ limit of the Segal–Bargmann transform to the work of Biane is completely different from that of [16]. Guillaume works within an extension of the circular system in which [4] constructs the transform, and shows that the leading term in his Lemma 4.1 is the generator of Biane's transform $\mathscr{G}^t$. We, by contrast, do not use free probability at all. Instead, we derive a polynomial generating function for the limiting transform and show that this generating function coincides with the generating function for $\mathscr{G}^t$; cf. Theorem 1.31.

In [4, Section 1], Biane considered the Segal–Bargmann transform $S_t^N$ acting on matrix-valued function on the Lie algebra $\mathfrak{u}(N)$. The result of applying $S_t^N$ to a single-variable polynomial function—that is, the type of function in (1.3)—is typically *not* a single-variable polynomial function on $M_N(\mathbb{C})$. Nevertheless, [4, Theorem 2] asserts that, for each single-variable polynomial $P$, there is a unique single-variable polynomial $P^t$ such that

$$\lim_{N\to\infty} \|S_t^N P_N - [P^t]_N\|_{L^2(M_N(\mathbb{C}),\gamma_{t/2};M_N(\mathbb{C}))} = 0, \tag{1.4}$$

Biane's transform $\mathscr{F}^t$ is then the map sending $P$ to $P^t$.

In the present paper, we establish a similar result for the transform $B_t^N$ on the unitary group $U(N)$, extended to act on functions with values in $M_N(\mathbb{C})$. In the group case, it is natural to consider single-variable *Laurent* polynomial functions; that is, linear combinations of matrix-valued functions of the form $U \mapsto U^k$, where $k$ is a (possibly negative) integer. As in the Lie algebra case, $B_t^N$ does not map single-variable Laurent polynomials on $U(N)$ to single-variable Laurent polynomials on $GL(N,\mathbb{C})$. Rather, applying $B_t^N$ to a single-variable Laurent polynomial gives a *trace Laurent polynomial*.

**Definition 1.3.** *Let $G \subset M_N(\mathbb{C})$ be a matrix group. Let $n \in \mathbb{N}$ and let $P$ be a polynomial in $2(n+1)$ commuting variables. Denote by $P_N\colon G \to M_N(\mathbb{C})$ the function*

$$P_N(A) = P(A, A^{-1}, \mathrm{tr}(A), \mathrm{tr}(A^{-1}), \ldots, \mathrm{tr}(A^n), \mathrm{tr}(A^{-n}))I_N, \qquad A \in G$$

*where $\mathrm{tr}(A) = \frac{1}{N}\mathrm{Tr}\,(A)$ is the normalized trace, and $I_N$ is the identity matrix. Any such function on $G$ is called a* **trace Laurent polynomial function***. In the special case $n = 0$, $P \in \mathbb{C}[u, u^{-1}]$ is a single-variable Laurent polynomial, and $P_N$ is referred to as a* **single-variable Laurent polynomial function** $G \to M_N(\mathbb{C})$. *(In particular, single-variable Laurent polynomial functions on $G$ form a subspace of trace Laurent polynomial functions on $G$.)*

The $I_N$ in the definition of $P_N$ is added to emphasize that, even if $P$ is "scalar" (i.e. does not depend on the $A$ and $A^{-1}$ variables), $P_N$ is still $M_N(\mathbb{C})$-valued (in this case a scalar function times $I_N$).

The Segal–Bargmann transform of any single-variable Laurent polynomial function on $U(N)$ is a trace Laurent polynomial function. What's more, this class is invariant under the transform:

**Proposition 1.4.** *Let $s, t > 0$ with $s > t/2$. The Segal–Bargmann transform $\mathbf{B}_{s,t}^N$ (cf. Definition 1.8) maps trace Laurent polynomial functions on $U(N)$ to trace Laurent polynomial functions on $GL(N, \mathbb{C})$.*

The proof of Proposition 1.4 can be found on page 25. For example, let $P(u) = u^2$; we calculate in Example 3.5 below that, for all $Z \in GL(N, \mathbb{C})$,

$$(\mathbf{B}_{s,t}^N P_N)(Z) = e^{-t}\left[\cosh(t/N)Z^2 - t\frac{\sinh(t/N)}{t/N}Z\mathrm{tr}(Z)\right] \tag{1.5}$$
$$= e^{-t}[Z^2 - tZ\mathrm{tr}Z] + O\left(\frac{1}{N^2}\right),$$

where the $O(1/N^2)$ means the left- and right-hand-sides differ by $O(1/N^2)$-scalar-multiples of $Z^2$ and $Z\mathrm{tr}Z$. This suggests we identify the large-$N$ limit of the Segal–Bargmann transform acting on $U \mapsto U^2$ as the *formal* trace polynomial $e^{-t}[Z^2 - tZ\mathrm{tr}Z]$; ignoring the unknown domain of the variable $Z$, it appears the result is still not a genuine polynomial. This highlights the second aspect of the limit that comes into play: in fact, we will see that the trace term $\mathrm{tr}(Z)$ concentrates on its mean $e^{-(s-t)/2}$ in the spaces $L^2(GL(N, \mathbb{C}), \mu_{s,t}^N; M_N(\mathbb{C}))$; cf. Theorem 1.26. Setting $(\mathbf{B}_{s,t}P)(z) = e^{-t}(z^2 - te^{-(s-t)/2}z)$, we therefore have

$$\|\mathbf{B}_{s,t}^N P_N - [\mathbf{B}_{s,t}P]_N\|^2_{L^2(GL(N,\mathbb{C}),\mu_{s,t}^N;M_N(\mathbb{C}))} = O\left(\frac{1}{N^2}\right). \tag{1.6}$$

This procedure works in general: given any Laurent polynomial $P$, there is a unique holomorphic Laurent polynomial $\mathbf{B}_{s,t}P$ so that $\mathbf{B}_{s,t}^N P_N \to \mathbf{B}_{s,t}P$ in the sense of (1.6). This is our Main Theorem.

**Main Theorem.** *Let $s, t > 0$ with $s > t/2$. There is an invertible operator $\mathbf{B}_{s,t}$ on the space of Laurent polynomials in a single variable such that, for any Laurent polynomial $P$,*

$$\|\mathbf{B}_{s,t}^N P_N - [\mathbf{B}_{s,t}P]_N\|^2_{L^2(GL(N,\mathbb{C}),\mu_{s,t}^N;M_N(\mathbb{C}))} = O\left(\frac{1}{N^2}\right), \quad \textit{and}$$
$$\|(\mathbf{B}_{s,t}^N)^{-1} P_N - [\mathbf{B}_{s,t}^{-1}P]_N\|^2_{L^2(U(N),\rho_s^N;M_N(\mathbb{C}))} = O\left(\frac{1}{N^2}\right).$$

*A formula for $\mathbf{B}_{s,t}$ may be found in Definition 1.28. In the special case $s = t$, $\mathbf{B}_{t,t}$ coincides with the free unitary Segal–Bargmann transform $\mathscr{G}^t$ from [4].*

We prove this as Theorems 1.30, 1.29, and 1.31, stated on page 11.

Thus, the proper way to take the large-$N$ limit of $\mathbf{B}_{s,t}^N$ is to first take the limit of the coefficients of the resultant trace Laurent polynomial function, and then let the trace terms concentrate on their means to produce a genuine single-variable Laurent polynomial. This is how we define the large-$N$ limit of the (two-parameter) Segal–Bargmann transform on $U(N)$.

A crucial result underlying many of our theorems is a "limiting partial product rule" for the action of the Laplacian $\Delta_{U(N)}$ for $U(N)$ (with metric scaled as in (1.2)) on trace Laurent polynomial functions. Let $P, Q$ be as in Definition 1.3, at least one of which (say $Q$) is scalar-valued. Then, when applying $\Delta_{U(N)}$ to the product $Q_N P_N$, the Leibniz product rule holds modulo an error term of order $1/N^2$: there is a fixed polynomial $R$, determined by $P$ and $Q$, such that

$$\Delta_{U(N)}(P_N Q_N) = (\Delta_{U(N)}P_N)Q_N + P_N(\Delta_{U(N)}Q_N) + \frac{1}{N^2}R_N. \tag{1.7}$$

Thus, in the large-$N$ limit, $\Delta_{U(N)}$ behaves essentially like a *first-order* differential operator; we make this precise in Theorem 1.20. This property allows us to recursively determine the limit action of the heat operator $e^{\frac{t}{2}\Delta_{U(N)}}$ and thus the Segal–Bargmann transform; cf. Section 5.1. It also gives an elegant framework to prove (and, more importantly, *explain*) the concentration of measure phenomenon that is essential to our main results.

The remainder of this introduction is devoted to precise statements of the definitions and theorems of this paper.

## 1.1 Laplacian and Segal–Bargmann Transform on $U(N)$

Throughout this paper, $N$ is a positive integer. In Definitions 1.5, 1.6, and 1.8, $V$ stands for any normed compex vector space; we will shortly deal exclusively with the case $V = M_N(\mathbb{C})$, but equipped with a *different* norm from $\|\cdot\|_{\mathfrak{u}(N)}$ in (1.2).

**Definition 1.5.** *Let $G \subset M_N(\mathbb{C})$ be a matrix Lie group with Lie algebra $\mathfrak{g} \subset M_N(\mathbb{C})$. For $X \in \mathfrak{g}$, the associated* **left-invariant derivative** *in the direction $X$ is the operator $\partial_X : C^\infty(G;V) \to C^\infty(G;V)$ given by*

$$(\partial_X F)(A) = \frac{d}{dt}\bigg|_{t=0} F(Ae^{tX}), \qquad A \in G. \tag{1.8}$$

**Definition 1.6.** *Let $\beta_N$ be an orthonormal basis for $\mathfrak{u}(N)$ (with norm $\|\cdot\|_{\mathfrak{u}(N)}$ given in (1.2)). The* **Laplacian** *$\Delta_{U(N)}$ on $C^\infty(U(N);V)$ is the operator*

$$\Delta_{U(N)} = \sum_{X\in\beta_N} \partial_X^2 \tag{1.9}$$

*which is independent of the choice of orthonormal basis $\beta_N$. For $t > 0$, the* **heat operator** *is $e^{\frac{t}{2}\Delta_{U(N)}}$. The* **heat kernel measure** *$\rho_t^N$ is determined by*

$$\int_{U(N)} f(U)\rho_t^N(dU) = \left(e^{\frac{t}{2}\Delta_{U(N)}} f\right)(I_N), \qquad f \in C(U(N)), \tag{1.10}$$

*where $I_N$ is the identity matrix in $M_N(\mathbb{C})$. We will sometimes write $\mathbb{E}_{\rho_t^N}(f)$ to mean $\int_{U(N)} f(U)\,\rho_t^N(dU)$.*

*Let $s,t > 0$ with $s > t/2$. Define the operator $A_{s,t}^N$ on $C^\infty(GL(N,\mathbb{C});V)$ by*

$$A_{s,t}^N = \left(s - \frac{t}{2}\right)\sum_{X\in\beta_N} \partial_X^2 + \frac{t}{2}\sum_{X\in\beta_N}\partial_{iX}^2. \tag{1.11}$$

*The measure $\mu_{s,t}^N$ on $GL(N,\mathbb{C})$ is determined by*

$$\int_{GL(N,\mathbb{C})} f(A)\,\mu_{s,t}^N(dA) = \left(e^{\frac{1}{2}A_{s,t}} f\right)(I_N), \qquad f \in C_c(GL(N,\mathbb{C})). \tag{1.12}$$

*We will sometimes write $\mathbb{E}_{\mu_{s,t}^N}(f)$ to mean $\int_{GL(N,\mathbb{C})} f(U)\,\mu_{s,t}(dU)$.*

*Remark* 1.7. (1) When $s = t$, $A_{s,t}^N = \frac{t}{2}\Delta_{GL(N,\mathbb{C})}$, and the measure $\mu_{t,t}^N$ is therefore the heat kernel measure on $GL(N,\mathbb{C})$ (at $\frac{1}{2}$ the usual time). On the other hand, $A_{1,0}^N\big|_{C^\infty(U(N))} = \Delta_{U(N)}$. Thus, $A_{s,t}^N$ interpolates between the two heat kernels.

(2) Since $\Delta_{U(N)}$ and $A_{s,t}^N$ are elliptic operators, the semigroups $e^{\frac{t}{2}\Delta_{U(N)}}$ and $e^{\frac{1}{2}A_{s,t}^N}$ can be defined using heat equations. For our purposes, where the functions $f$ will be trace polynomial functions, it suffices to expand $e^{\frac{t}{2}\Delta_{U(N)}}$ and $e^{\frac{1}{2}A_{s,t}^N}$ as power series. This is discussed in Appendix A.

(3) The formula (1.12) extends beyond compactly-supported functions; in particular, it also holds for any scalar-valued trace Laurent polynomial function $f$ in variables $A, A^* \in GL(N,\mathbb{C})$. This follows from Langland's Theorem cf. [24, Theorem 2.1 (p. 152)]. Appendix A gives a concise sketch of the heat kernel results we need in this paper.

**Definition 1.8.** *Fix $s, t > 0$ with $s > t/2$. The* scalar *unitary* **Segal–Bargmann transform**

$$B_{s,t}^N \colon L^2(U(N), \rho_s^N) \to \mathcal{H}L^2(GL(N,\mathbb{C}), \mu_{s,t}^N)$$

*is defined by*

$$B_{s,t}^N f = \left(e^{\frac{t}{2}\Delta_{U(N)}} f\right)_{\mathbb{C}}$$

*where $(\,\cdot\,)_{\mathbb{C}}$ denotes analytic continuation from $U(N)$ to its complexification $GL(N,\mathbb{C})$. It is an isometric isomorphism of these spaces; cf. [7, 10, 17, 18, 20].*

*We let $B_{s,t}^N$ act on $V$-valued functions componentwise; that is, the* boosted *transform*

$$B_{s,t}^N \otimes \mathbb{1}_V \colon L^2(U(N), \rho_s^N) \otimes V \to \mathcal{H}L^2(GL(N,\mathbb{C}), \mu_{s,t}^N) \otimes V$$

*is also an isometric isomorphism. (All tensor products are over $\mathbb{C}$.) As usual, we identify*

$$L^2(U(N), \rho_s^N; V) \cong L^2(U(N), \rho_s^N) \otimes V \qquad \text{and} \qquad \mathcal{H}L^2(GL(N,\mathbb{C}), \mu_{s,t}^N; V) \cong \mathcal{H}L^2(GL(N,\mathbb{C}), \mu_{s,t}^N) \otimes V,$$

*where the induced norms of the $V$-valued functions $F$ are*

$$\|F\|^2_{L^2(U(N),\rho_s^N;V)} \equiv \int_{U(N)} \|F(U)\|_V^2 \, \rho_s^N(dU) \tag{1.13}$$

$$\|H\|^2_{L^2(GL(N,\mathbb{C}),\mu_{s,t}^N;V)} \equiv \int_{GL(N,\mathbb{C})} \|H(A)\|_V^2 \, \mu_{s,t}^N(dA). \tag{1.14}$$

*Henceforth, we will let $V = M_N(\mathbb{C})$, equipped with the inner-product*

$$\|A\|^2_{M_N(\mathbb{C})} = \frac{1}{N}\mathrm{Tr}\,(AA^*) = \frac{1}{N}\sum_{j,k=1}^N |A_{jk}|^2. \tag{1.15}$$

**Notation***: $\mathbf{B}_{s,t}^N$ refers to the boosted unitary Segal–Bargmann transform $B_{s,t}^N \otimes \mathbb{1}_{M_N(\mathbb{C})}$ with this choice of norm on $M_N(\mathbb{C})$.*

*Remark* 1.9. We are free to use any normed space for the boosted Segal–Bargmann transform, as long as it is the same space for values of the functions in the domain and in the range of $\mathbf{B}_{s,t}^N$. We will not use the norm $\|\cdot\|_{\mathfrak{u}(N)}$ from (1.2); rather, we use the scaling of (1.15). As we will see, this joint scaling of the two norms is the unique choice that give a meaningful large-$N$ limit for the Segal–Bargmann transform (cf. Remark 3.4).

*Remark* 1.10. Equations (1.10) and (1.12) in conjunction with Remark 1.7(3) give an easy way to compute $L^2$-norms with respect to the heat kernel measures. In particular, (1.12) and (1.14) yield

$$\|F\|^2_{L^2(GL(N,\mathbb{C}),\mu_{s,t}^N;V)} = (e^{\frac{1}{2}A_{s,t}^N}\|F\|^2_{M_N(\mathbb{C})})(I_N) \tag{1.16}$$

for any trace Laurent polynomial function $F$.

## 1.2 Trace Laurent Polynomials and Intertwining Operators

To properly treat trace Laurent polynomial functions over $U(N)$ and $GL(N,\mathbb{C})$ (cf. Definition 1.3), the following polynomial spaces are useful.

**Definition 1.11.** *Throughout this paper, we let* $u$ *and* $\mathbf{v} = \{v_n\}_{n\in\mathbb{Z}\setminus\{0\}} = \{v_{\pm1}, v_{\pm2}, \ldots\}$ *denote commuting indeterminates. Define*

$$\mathscr{P}^1 = \mathbb{C}[u, u^{-1}], \qquad \mathscr{P}^0 = \mathbb{C}[\mathbf{v}], \qquad \mathscr{P} = \mathscr{P}^1 \otimes \mathscr{P}^0 = \mathbb{C}[u, u^{-1}; \mathbf{v}]. \tag{1.17}$$

*Thus,* $\mathscr{P}$ *consists of all finite linear combinations of monomials*

$$u^{k_0} v_1^{k_1} v_{-1}^{k_{-1}} \cdots v_n^{k_n} v_{-n}^{k_{-n}}, \qquad n \geq 0,\;\; k_0 \in \mathbb{Z},\;\; k_j \in \mathbb{N} \;\; \textit{for} \;\; j \in \mathbb{Z}\setminus\{0\}.$$

*Denote by* $\mathscr{P}^{1,+} = \mathbb{C}[u]$ *the polynomials in* $u$*, and* $\mathscr{P}^{1,-} = \mathbb{C}[u^{-1}] \ominus \mathbb{C}$ *the polynomials in* $u^{-1}$ *with constant term* $0$*, so that* $\mathscr{P}^1 = \mathscr{P}^{1,+} \oplus \mathscr{P}^{1,-}$*.*

*The* **trace degree** *of a monomial in* $\mathscr{P}$ *is*

$$\deg\left(u^{k_0} v_1^{k_1} v_{-1}^{k_{-1}} \cdots v_n^{k_n} v_{-n}^{k_{-n}}\right) = |k_0| + \sum_{1\leq|j|\leq n} |j| k_j. \tag{1.18}$$

*More generally, the trace degree of any element of* $\mathscr{P}$ *is the maximum of the trace degrees of its monomial terms. For* $n \geq 0$*, denote by* $\mathscr{P}_n \subset \mathscr{P}$ *the subspace of polynomials of trace degree* $\leq n$*:*

$$\mathscr{P}_n = \{P \in \mathscr{P}\colon\; \deg P \leq n\}. \tag{1.19}$$

*Note that* $\mathscr{P}_n$ *is finite dimensional,* $\mathscr{P}_n \subset \mathbb{C}[u, u^{-1}; v_{\pm1}, \ldots, v_{\pm n}]$*, and* $\mathscr{P} = \bigcup_{n\geq0} \mathscr{P}_n$*. Define* $\mathscr{P}_n^0$*,* $\mathscr{P}_n^1$*, and* $\mathscr{P}_n^{1,\pm}$ *similarly.*

*Let* $G \subset M_N(\mathbb{C})$ *be a matrix group. Restating Definition 1.3 in the present language, a* **trace Laurent polynomial function** *on* $G$ *is any function of the form*

$$P_N(A) = P(A; \mathrm{tr}(A), \mathrm{tr}(A^{-1}), \mathrm{tr}(A^2), \mathrm{tr}(A^{-2}), \ldots) I_N \qquad A \in G \tag{1.20}$$

*for some* $P \in \mathscr{P}$*.*

*Remark* 1.12. (1) The first variable in $P$ on the right-hand-side of (1.20) is a *Laurent polynomial* variable, meaning that $A$ may be have positive or negative powers. The standard notation for this might be to write $P(A, A^{-1}, \mathrm{tr}(A), \mathrm{tr}(A^{-1}), \ldots)I_N$, but this might suggest that $u$ and $u^{-1}$ are independent variables, so we avoid this.

(2) For Definitions 1.3 and 1.11, it doesn't matter whether we use the trace $\mathrm{Tr}$ or the normalized trace $\mathrm{tr}$. It will be convenient to use $\mathrm{tr}$ later, so we fix this convention now.

(3) It is important to note that, for any finite $N$, there will be many *distinct* elements $P \in \mathscr{P}$ that induce the same trace Laurent polynomial function on $G$, i.e. there will be $P \neq Q$ with $P_N = Q_N$. Nevertheless, it is true that if $P_N = Q_N$ for *sufficiently large* $N$, then $P = Q$; cf. Proposition 2.10.

(4) The trace degree reflects the nature of the variables $v_{\pm1}, v_{\pm2}, \ldots$ in $\mathscr{P}$ as stand-ins for traces of powers of a matrix variable. Informally, the trace degree of $P \in \mathscr{P}$ is the total degree of $P_N(A)$, counting all instances of $A$ inside and outside traces.

**Definition 1.13.** *The* **tracing map** $\mathcal{T}\colon \mathscr{P} \to \mathscr{P}^0$ *is the linear operator given as follows: if* $P \in \mathscr{P}^0$ *and* $k \in \mathbb{Z}\setminus\{0\}$*, then*

$$\mathcal{T}(u^k P(\mathbf{v})) = v_k P(\mathbf{v}). \tag{1.21}$$

*An element* $P \in \mathscr{P}$ *is in* $\mathscr{P}^0$ *if and only if* $\mathcal{T}(P) = P$*.*

*Remark* 1.14. The operator $\mathcal{T}$ intertwines the normalized trace on matrix-valued functions on $M_N(\mathbb{C})$:

$$[\mathcal{T}(P)]_N = \mathrm{tr} \circ P_N \tag{1.22}$$

as can be easily verified.

**Definition 1.15.** *Let $\mathcal{A}_\pm$ denote the* positive part *and* negative part *operators*

$$\mathcal{A}_\pm \colon \mathscr{P} \to \mathscr{P}^1_\pm \otimes \mathscr{P}^0$$

*given by*

$$\mathcal{A}_+ \left( \sum_{k=-\infty}^{\infty} u^k q_k(\mathbf{v}) \right) = \sum_{k=0}^{\infty} u^k q_k(\mathbf{v}), \quad \mathcal{A}_- \left( \sum_{k=-\infty}^{\infty} u^k q_k(\mathbf{v}) \right) = \sum_{k=-\infty}^{-1} u^k q_k(\mathbf{v}). \tag{1.23}$$

*Note that $\mathcal{A}_+ + \mathcal{A}_- = \mathrm{id}_{\mathscr{P}}$, while $\mathcal{A}_+ - \mathcal{A}_- = \mathrm{sgn}$ is the signum operator, where $\mathrm{sgn}(u^n) = \mathrm{sgn}(n)u^n$, and $\mathrm{sgn}(n) = n/|n|$ when $n \neq 0$ and $\mathrm{sgn}(0) = 1$.*

The next results show that the Laplacian $\Delta_{U(N)}$ of (1.9), when acting on trace Laurent polynomial functions over $U(N)$, lifts to a linear operator $\mathcal{D}_N$ on $\mathscr{P}$. Moreover, the structure of this operator illucidates the limit prodecure that will allow us to identify the limit Segal–Bargmann transform.

**Definition 1.16.** *Define the following operators on $\mathscr{P}$. For $k \in \mathbb{Z}$, let $\mathcal{M}_{(\cdot)}$ denote the multiplication operator, and define:*

$$\mathcal{N}_1 = u\frac{\partial}{\partial u}(\mathcal{A}_+ - \mathcal{A}_-), \qquad \mathcal{N}_0 = \sum_{|k|\geq 1} |k| v_k \frac{\partial}{\partial v_k}, \qquad \mathcal{N} = \mathcal{N}_0 + \mathcal{N}_1, \tag{1.24}$$

$$\mathcal{Y} = \mathcal{Y}_+ - \mathcal{Y}_- = \sum_{k=1}^{\infty} v_k u \mathcal{A}_+ \frac{\partial}{\partial u} \mathcal{M}_{u^{-k}} \mathcal{A}_+ - \sum_{k=-\infty}^{-1} v_k u \mathcal{A}_- \frac{\partial}{\partial u} \mathcal{M}_{u^{-k}} \mathcal{A}_-, \tag{1.25}$$

$$\mathcal{Z} = \mathcal{Z}_+ - \mathcal{Z}_- = \sum_{k=2}^{\infty} \left( \sum_{j=1}^{k-1} j v_j v_{k-j} \right) \frac{\partial}{\partial v_k} - \sum_{k=-\infty}^{-2} \left( \sum_{j=k+1}^{-1} j v_k v_{k-j} \right) \frac{\partial}{\partial v_k}, \tag{1.26}$$

$$\mathcal{L} = \sum_{|j|,|k|\geq 1} jk v_{k+j} \frac{\partial^2}{\partial v_j \partial v_k} + 2 \sum_{|k|\geq 1} k u^{k+1} \frac{\partial^2}{\partial v_k \partial u}. \tag{1.27}$$

*Example* 1.17. The operator $\mathcal{N}$ is the trace degree operator: for $P \in \mathscr{P}$

$$\mathcal{N}(P) = (\deg P)P.$$

The operator $\mathcal{N}_1$ is a first-order *pseudo*-differential operator; but it is not a genuine differential operator because we do not treat $u$ and $u^{-1}$ as independent variables. By contrast, we do treat $v_k$ and $v_{-k}$ as independent, and so $\mathcal{N}_0$ is a first-order differential operator.

*Example* 1.18. The first order pseudo-differential operator $\mathcal{Y}$ appears somewhat mysterious; we illustrate its action here.

- $\mathcal{Y}$ annihilates $\mathscr{P}^0$; more generally, for $P \in \mathscr{P}$ and $Q \in \mathscr{P}^0$, $\mathcal{Y}(PQ) = \mathcal{Y}(P) \cdot Q$. It therefore suffices to understand the action of $\mathcal{Y}$ on $\mathscr{P}^1$.

- $\mathcal{Y}_\pm$ annihilates $\mathscr{P}^{1,\mp}$. The reader can calculate that

$$\mathcal{Y}(u^n) = \mathcal{Y}_+(u^n) = \sum_{k=1}^{n-1}(n-k)v_k u^{n-k}, \qquad n \geq 0$$
$$-\mathcal{Y}(u^n) = \mathcal{Y}_-(u^n) = \sum_{k=n+1}^{-1}(n-k)v_k u^{n-k}, \qquad n < 0.$$

**Lemma 1.19.** *Let $\mathcal{N}, \mathcal{Y}, \mathcal{Z}, \mathcal{L} \colon \mathscr{P} \to \mathscr{P}$ be given as in Definition 1.16. Define*

$$\mathcal{D} = -\mathcal{N} - 2\mathcal{Z} - 2\mathcal{Y} \tag{1.28}$$
$$\mathcal{D}_N = -\mathcal{N} - 2\mathcal{Z} - 2\mathcal{Y} - \frac{1}{N^2}\mathcal{L} = \mathcal{D} - \frac{1}{N^2}\mathcal{L}. \tag{1.29}$$

*The operators $\mathcal{D}_N$, $\mathcal{D}$, and $\mathcal{L}$ preserve trace degree (1.18), and commute with the tracing map $\mathcal{T}$ (1.21).*

The proof of Lemma 1.19 can be found on page 23.

The next theorem shows that $\Delta_{U(N)}$ lifts to the operator $\mathcal{D}_N$ on $\mathscr{P}$. Hence, it is a $O(1/N^2)$-perturbation of $\mathcal{D}$, which behaves in many respects like a first-order differential operator.

**Theorem 1.20** (Intertwining Formula)**.** *For any $P \in \mathscr{P}$,*

$$\Delta_{U(N)} P_N = [\mathcal{D}_N P]_N. \tag{1.30}$$

The proof of Theorem 1.20 can be found on page 23.

*Remark* 1.21. A similar intertwining formula holds for the operator $A_{s,t}$; cf. Theorem 3.13 on page 29.

The following easy corollary to Theorem 1.20 is of both computational and conceptual importance.

**Corollary 1.22.** *Let $P \in \mathscr{P}$ and $Q \in \mathscr{P}^0$. Then*

$$\mathcal{D}(PQ) = (\mathcal{D}P)Q + P(\mathcal{D}Q). \tag{1.31}$$

*Thus, for any $t \in \mathbb{R}$,*

$$e^{\frac{t}{2}\mathcal{D}}(PQ) = e^{\frac{t}{2}\mathcal{D}}P \cdot e^{\frac{t}{2}\mathcal{D}}Q. \tag{1.32}$$

The proof of Corollary 1.22 can be found on page 25.

*Remark* 1.23. Eq. (1.31) and Theorem 1.20 prove the partial product rule of (1.7). Indeed, it is easy to work out that the polynomial $R$ in (1.7) is

$$R = \mathcal{L}(P)Q + P\mathcal{L}(Q) - \mathcal{L}(PQ).$$

### 1.3 Limit Theorems

We now turn to the limit as $N \to \infty$. Since the Segal–Bargmann transform is (analytic continuation of) the heat operator $\exp(\frac{t}{2}\Delta_{U(N)})$, the intertwining Theorem 1.20 suggests the limit Segal–Bargmann transform ought to be given in terms of the semigroup $\exp(\frac{t}{2}\mathcal{D})$. This is only half the story, however. As $N \to \infty$, each trace Laurent polynomial function concentrates on a single-variable Laurent polynomial function, described by the following map.

**Definition 1.24.** *For $s \in \mathbb{R}$, the* **evaluation map** *$\pi_s \colon \mathscr{P} \to \mathscr{P}^1$ is defined to be*

$$(\pi_s P)(u) = \left(e^{-\frac{s}{2}(\mathcal{N}_0 + 2\mathcal{Z})}\right) P(u, \mathbf{1})$$

*where $\mathbf{1} = (1, 1, \ldots)$. Since $\mathcal{N}_0$ and and $\mathcal{Z}$ are first-order differential operators, $\pi_s$ is an algebra homomorphism.*

*Remark* 1.25. Note that $\mathcal{N}_1$ and $\mathcal{Y}$ annihilate $\mathscr{P}^0$; thus, for $Q \in \mathscr{P}^0$, we have $-(\mathcal{N}_0+2\mathcal{Z})Q = \mathcal{D}Q$, and therefore the constant $\pi_s Q$ is given by $\pi_s Q = \left(e^{\frac{s}{2}\mathcal{D}}Q\right)(\mathbf{1})$. In particular, define

$$\nu_k(s) = \pi_s(v_k). \qquad k \in \mathbb{Z} \tag{1.33}$$

Note that, since the coefficients of the differential operators $\mathcal{N}_0$ and $\mathcal{Z}$ are real, the constants $\nu_k(s)$ are real. What's more, it follows from the structure of $\mathcal{N}_0$ and $\mathcal{Z}$ that $\nu_{-k}(s) = \nu_k(s)$. In Lemma 5.4, we show that $|\nu_k(s)|$ is exponentially bounded in $k$ for all $s \in \mathbb{R}$.

The following concentration of measure results underly our main limit theorems.

**Theorem 1.26.** *Let $s, t > 0$, with $s > t/2$. If $P \in \mathscr{P}$, then the following hold:*

$$\|P_N - [\pi_s P]_N\|^2_{L^2(U(N),\rho_s^N;M_N(\mathbb{C}))} = O\left(\frac{1}{N^2}\right) \tag{1.34}$$

$$\|P_N - [\pi_{s-t} P]_N\|^2_{L^2(GL(N,\mathbb{C}),\mu_{s,t}^N;M_N(\mathbb{C}))} = O\left(\frac{1}{N^2}\right). \tag{1.35}$$

The proof of Theorem 1.26 can be found on page 32.

*Remark* 1.27. Taking $P(u, \mathbf{v}) = v_k$, so that $P_N(U) = \mathrm{tr}(U^k)$, (1.33) and (1.34) show, in particular, that $\nu_k(s) = \lim_{N\to\infty} \mathbb{E}_{\rho_s^N} \mathrm{tr}((\,\cdot\,)^k)$ when $s > 0$. These limiting expected values, in turn, were explicitly calculated in [3, Lemma 3]:

$$\nu_k(s) = e^{-\frac{k}{2}s} \sum_{j=0}^{k-1} (-1)^j \frac{s^j}{k!} k^{j-1} \binom{k}{j+1}, \qquad k \geq 0. \tag{1.36}$$

They are the (trace) moments of the *free unitary Brownian motion*. There is a probability measure $\nu_s$ on the unit circle $\mathbb{T}$ such that $\nu_k(s) = \int_{\mathbb{T}} \xi^k \nu_s(d\xi)$ for $k \in \mathbb{Z}$; we will refer to this measure as the **free unitary Brownian motion distribution**. The measure $\nu_s$ is characterized in [4, Prop. 10].

Because of Theorem 1.26's concentration of traces, although in general $e^{\frac{t}{2}\mathcal{D}}P \notin \mathscr{P}^1$ is (even if $P \in \mathscr{P}^1$), the corresponding trace Laurent polynomial function $[e^{\frac{t}{2}\mathcal{D}}P]_N$ is close, in $L^2(\mu_{s,t}^N)$-sense, to a single-variable Laurent polynomial function on $GL(N, \mathbb{C})$. In the limit, this will produce an operator on $\mathscr{P}^1$. This finally brings us to the limit Segal–Bargmann transform.

**Definition 1.28.** *For $s, t > 0$ with $s > t/2$, the* **free unitary Segal–Bargmann transform** $\mathbf{B}_{s,t}\colon \mathscr{P}^1 \to \mathscr{P}^1$ *is given by*

$$\mathbf{B}_{s,t} = \pi_{s-t} \circ \exp\left(\frac{t}{2}\mathcal{D}\right).$$

*The* **inverse free unitary Segal–Bargmann transform** $\mathbf{H}_{s,t}\colon \mathscr{P}^1 \to \mathscr{P}^1$ *is defined as*

$$\mathbf{H}_{s,t} = \pi_s \circ \exp\left(-\frac{t}{2}\mathcal{D}\right).$$

Without the evaluation maps, the two operators $e^{\frac{t}{2}\mathcal{D}}$ and $e^{-\frac{t}{2}\mathcal{D}}$ are, of course, inverse to each other. In fact, this is true with the evaluations as well, justifying the terminology in Definition 1.28.

**Theorem 1.29.** *For $s, t > 0$ with $s > \frac{t}{2}$, $\mathbf{B}_{s,t}$ and $\mathbf{H}_{s,t}$ are invertible operators on $\mathscr{P}^1$, inverse to each other.*

The proof of Theorem 1.29 can be found on page 35.

Following is the motivating theorem of this paper: the free unitary Segal–Bargmann transform of Definition 1.28 is the limit of the unitary Segal–Bargmann transform on $U(N)$ of Definition 1.8, as $N \to \infty$.

**Theorem 1.30.** *Let $s,t>0$ with $s>\frac{t}{2}$. Let $f\in\mathscr{P}^1$ be a single-variable Laurent polynomial. Then $\mathbf{B}_{s,t}f$ is the limit of $\mathbf{B}_{s,t}^N f_N$ in $\mathcal{H}L^2(GL(N,\mathbb{C}),\mu_{s,t}^N;M_N(\mathbb{C}))$ in the following sense:*

$$\|\mathbf{B}_{s,t}^N f_N-[\mathbf{B}_{s,t}f]_N\|^2_{L^2(GL(N,\mathbb{C}),\mu_{s,t}^N;M_N(\mathbb{C}))}=O\left(\frac{1}{N^2}\right), \tag{1.37}$$

*and $\mathbf{H}_{s,t}f$ is the limit of $(\mathbf{B}_{s,t}^N)^{-1}f_N$ in $L^2(U(N),\rho_s^N;M_N(\mathbb{C}))$ in the following sense:*

$$\|(\mathbf{B}_{s,t}^N)^{-1}f_N-[\mathbf{H}_{s,t}f]_N\|^2_{L^2(U(N),\rho_s^N;M_N(\mathbb{C}))}=O\left(\frac{1}{N^2}\right). \tag{1.38}$$

The proof of Theorem 1.30 can be found on page 32.

Finally, we explicitly describe the action of the transform, through its inverse $\mathbf{H}_{s,t}$, via the generating function of its inverse action on monomials. In the special case $s=t$, we show this aligns with the free Gross-Malliavin approach pioneered by Biane in [4].

**Theorem 1.31.** *Let $s,t>0$ with $s>t/2$, let $k\geq 1$, and let $p_k^{s,t}=\mathbf{H}_{s,t}((\,\cdot\,)^k)$ (so that $\mathbf{B}_{s,t}(p_k^{s,t})(z)=z^k$). Then the power series*

$$\Pi(s,t,u,z)=\sum_{k\geq 1}p_k^{s,t}(u)z^k$$

*converges for all sufficiently small $|u|$ and $|z|$. This generating function is determined by the implicit formula*

$$\Pi(s,t,u,ze^{\frac{1}{2}(s-t)\frac{1+z}{1-z}})=\left(1-uze^{\frac{s}{2}\frac{1+z}{1-z}}\right)^{-1}-1. \tag{1.39}$$

*In the special case $s=t$, this yields the generating function corresponding to the transform $\mathscr{G}^t$ of [4, Proposition 13]. Thus, $\mathbf{B}_{t,t}=\mathscr{G}^t$.*

The proof of Theorem 1.31 can be found on page 45.

## 2 Equivariant Functions and Trace Laurent Polynomials

In this section, we consider function spaces over $U(N)$ and $GL(N,\mathbb{C})$ that are very natural domains for the Segal-Bargmann transform and its inverse.

**Definition 2.1.** *Let $G\subset M_N(\mathbb{C})$ be a matrix group. A function $F\colon G\to M_N(\mathbb{C})$ is called* **equivariant** *if $F(BAB^{-1})=BF(A)B^{-1}$ for all $A,B\in G$ (i.e. it is equivariant under the adjoint action of $G$).*

The set of equivariant functions is a $\mathbb{C}$-algebra. It contains all trace Laurent polynomial functions (1.20), as can be easily verified. This shows that the **equivariant subspaces**

$$L^2(U(N),\rho_s^N;M_N(\mathbb{C}))_{\mathrm{eq}}\quad\text{and}\quad\mathcal{H}L^2(GL(N,\mathbb{C}),\mu_{s,t}^N;M_N(\mathbb{C}))_{\mathrm{eq}},$$

are non-trivial. The main results of this section, Theorem 2.3 and 2.7, show that $\mathbf{B}_{s,t}^N$ maps $L^2(\rho_s^N)_{\mathrm{eq}}$ onto $\mathcal{H}L^2(\mu_{s,t}^N)_{\mathrm{eq}}$ (extending Proposition 1.4), and that trace Laurent polynomials are dense in these equivariant $L^2$-spaces. We conclude this section with Theorem 2.10, showing that the map $\mathscr{P}\to L^2(\rho_s^N)_{\mathrm{eq}}$ given by $P\mapsto P_N$ is one-to-one on each subspace $\mathscr{P}_n$ for all sufficiently large $N$.

We begin with a brief discussion of **functional calculus**, another subspace of the $L^2$-equivariant space, which featured prominently in [4].

## 2.1 Functional Calculus

**Definition 2.2.** *Let $\mathbb{T}$ denote the unit circle in $\mathbb{C}$. For every measurable function $f : \mathbb{T} \to \mathbb{C}$, let $f_N$ be the unique function mapping $U(N)$ into $M_N(\mathbb{C})$ with the property that*

$$f_N\left(V\begin{pmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_N\end{pmatrix}V^{-1}\right) = V\begin{pmatrix}f(\lambda_1) & & \\ & \ddots & \\ & & f(\lambda_N)\end{pmatrix}V^{-1}$$

*for all $V \in U(N)$ and all $\lambda_1, \ldots, \lambda_N \in \mathbb{T}$. The function $f_N$ is called the* **functional calculus function** *associated to the function $f$. The space of those functional calculus functions that are in $L^2(U(N), \rho_s^N; M_N(\mathbb{C}))$ is called the* **functional calculus subspace***.*

It is easy to check that $f_N(U)$ is well defined, independent of the choice of diagonalization. If, for example, $f$ is the function given by $f(\lambda) = e^\lambda$, then $f_N(U) = e^U$, computed by the usual power series. If $f$ happens to be a single-variable Laurent polynomial, then $f_N$ is a single-variable Laurent polynomial function, in the sense of Definition 1.3; thus our notation $f_N$ for both is consistent. (By comparison: in [4], the functional calculus function $f_N$ is denoted $\theta_f^N$.) Trace polynomials are *not*, in general, functional calculus functions. For example, the function $F(U) = U\mathrm{tr}(U)$ is not a functional calculus function on $U(N)$, except when $N = 1$. Indeed, if $N \geq 2$ and $U(N) \ni U = \mathrm{diag}(\lambda_1, \lambda_2)$, the $(1,1)$-entry of the diagonal matrix $U\mathrm{tr}U$ is $\frac{1}{2}(\lambda_1 + \lambda_2)\lambda_1$, which is not a function of $\lambda_1$ alone. This violates Definition 2.2. Functional calculus functions are equivariant.

Since $\Lambda(f) \equiv \int_{U(N)} \mathrm{tr}(f_N(U))\, \rho_s^N(dU)$ defines a positive linear functional on $C(\mathbb{T})$ with $\Lambda(1) = 1$, by the Riesz Representation Theorem [25, Theorem 2.14] there is a probability measure $\nu_s^N$ on $\mathbb{T}$ such that

$$\int_{U(N)} \mathrm{tr}(f_N(U))\, \rho_s^N(dU) = \Lambda(f) = \int_{\mathbb{T}} f(\xi)\, \nu_s^N(d\xi), \qquad f \in C(\mathbb{T}). \tag{2.1}$$

(Theorem 1.26 shows, in particular, that $\nu_s^N$ converges weakly to $\nu_s$; cf. Remark 1.27.) For any function $f$ on $\mathbb{T}$, one can easily verify from Definition 2.2 that $[|f|^2]_N(U) = f_N(U)f_N(U)^*$; hence, by the density of $C(\mathbb{T})$ in $L^2(\mathbb{T}, \nu_s^N)$, (2.1) shows that

$$\|f_N\|_{L^2(U(N),\rho_s^N;M_N(\mathbb{C}))} = \|f\|_{L^2(\mathbb{T},\nu_s^N)}\,. \tag{2.2}$$

It follows that the functional calculus subspace is a closed subspace of $L^2(\rho_s^N)_{\mathrm{eq}}$, and contains the single-variable Laurent polynomial functions as a dense subspace. That this density result extends to trace Laurent polynomials in the full space $L^2(\rho_s^N)_{\mathrm{eq}}$ is Theorem 2.7 below.

If $F$ is a holomorphic function on $\mathbb{C}^*$, there is a unique holomorphic function $F_N : GL(N;\mathbb{C}) \to M_N(\mathbb{C})$ which satisfies

$$F_N\left(A\begin{pmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_N\end{pmatrix}A^{-1}\right) = A\begin{pmatrix}F(\lambda_1) & & \\ & \ddots & \\ & & F(\lambda_N)\end{pmatrix}A^{-1}$$

for every $A \in GL(N,\mathbb{C})$ and all $\lambda_1, \ldots, \lambda_N \in \mathbb{C}^*$; indeed, $F_N$ is given by the same Laurent series expansion as $F$, applied to the matrix variable. We call such a function a **holomorphic functional calculus function** on $GL(N,\mathbb{C})$. As (1.5) shows, the boosted Segal–Bargmann transform $\mathbf{B}_{s,t}^N$ *does not*, in general, map functional calculus functions on $U(N)$ to holomorphic functional calculus functions on $GL(N,\mathbb{C})$. Nevertheless, [4] suggests that *in the large-$N$ limit*, $\mathbf{B}_{s,t}^N$ ought to map functional calculus functions to holomorphic functional calculus functions (at least in the $s = t$ case), mirroring the limit theorem that holds for the Lie-algebra version $S_t^N$ of the transform; cf. [4, Theorem 2]. Since single-variable Laurent polynomial functions are dense in the functional calculus subspace, Theorem 1.30 can be interpreted as a rigorous version of this idea.

## 2.2 Results on Equivariant Functions

**Theorem 2.3.** *Let $s, t > 0$ with $s > t/2$. The Segal–Bargmann transform $\mathbf{B}_{s,t}^N$ maps $L^2(U(N), \rho_s^N; M_N(\mathbb{C}))_{\mathrm{eq}}$ isometrically onto $\mathcal{H}L^2(GL(N,\mathbb{C}), \mu_{s,t}^N; M_N(\mathbb{C}))_{\mathrm{eq}}$.*

We begin with the following lemma.

**Lemma 2.4.** *Let $G \subset M_N(\mathbb{C})$ be a group. For any function $F\colon G \to M_N(\mathbb{C})$, define*

$$C_V(F)(A) = V^{-1}F(VAV^{-1})V, \qquad V, A \in G. \tag{2.3}$$

*Let $s, t > 0$ with $s > t/2$. Then for all $F \in L^2(U(N), \rho_s; M_N(\mathbb{C}))$ and $V \in U(N)$,*

$$\mathbf{B}_{s,t}^N(C_V F) = C_V(\mathbf{B}_{s,t}^N F). \tag{2.4}$$

*Proof.* Since $\Delta_{U(N)}$ is bi-invariant, it commutes with the left- and right- actions of the group; hence it, and therefore the semigroup $e^{\frac{t}{2}\Delta_{U(N)}}$, commute with the adjoint action $\mathrm{Ad}_V(U) = VUV^{-1}$ on functions: for any $V \in U(N)$

$$e^{\frac{t}{2}\Delta_{U(N)}}(F \circ (\mathrm{Ad}_V)) = \left(e^{\frac{t}{2}\Delta_{U(N)}}F\right) \circ \mathrm{Ad}_V. \tag{2.5}$$

Conjugating both sides of (2.5) by $V^{-1}$ in the range of $F$ (which commutes with the heat operator), it follows that

$$C_V(e^{\frac{t}{2}\Delta_{U(N)}}F) = e^{\frac{t}{2}\Delta_{U(N)}}(C_V F), \qquad V \in U(N). \tag{2.6}$$

Uniqueness of analytic continuation now proves (2.4) from (2.6). □

Theorem 2.3 now follows by analytically continuing (2.4) in the $V$ variable.

*Proof of Theorem 2.3.* Let $F \in L^2(U(N), \rho_s^N; M_N(\mathbb{C}))$ be equivariant; thus $C_V F = F$ for all $V \in U(N)$. Then (2.4) shows that $C_V(\mathbf{B}_{s,t}^N F) - \mathbf{B}_{s,t}^N F \equiv 0$ for each $V \in U(N)$. Since $\mathbf{B}_{s,t}^N F$ is holomorphic, it follows by uniqueness of analytic continuation that the function $A \mapsto C_A(\mathbf{B}_{s,t}^N F) - \mathbf{B}_{s,t}^N F \equiv 0$ for $A \in GL(N,\mathbb{C})$; thus, $\mathbf{B}_{s,t}^N F$ is equivariant under $GL(N,\mathbb{C})$, as required. An entirely analogous argument applies to the inverse transform, establishing the proposition. □

Let us remark here on an intuitive approach to the concentration of measure results in Section 4. If $U_t$ is a random matrix sampled from the distribution $\rho_t^N$ on $U(N)$, its (random) eigenvalues converge to their (deterministic) mean as $N \to \infty$. To be precise: if $\lambda_1^N, \ldots, \lambda_N^N$ are the eigenvalues of $U_t$, the *empirical eigenvalue measure*

$$\widetilde{\nu}_t^N = \frac{1}{N}\sum_{j=1}^N \delta_{\lambda_j^N}$$

converges weakly in probability to the measure $\nu_t$. (The mean of the random measure $\widetilde{\nu}_t^N$ is the measure $\nu_t^N$ of (2.1) which converges weakly to $\nu_t$; cf. Remark 1.27. The stronger statement that the convergence is in probability, not just in expectation, follows from the variance estimates in [23, Proposition 6.2], for example.)

The conjugacy classes in the group $U(N)$ are in one-to-one correspondence with the (symmetrized) list of eigenvalues. Each such list is, in turn, determined by its emprical measure $\widetilde{\nu}_t^N$. The convergence of the random eigenvalues of $U_t$ to a deterministic limit therefore suggests that the heat kernel measure $\rho_t^N$ concentrates its mass on a *single conjugacy class* as $N \to \infty$. The following proposition therefore offers some insight into Theorem 1.26 (that trace Laurent polynomials concentrate on single-variable Laurent polynomials). Indeed, on a fixed conjugacy class, *any equivariant function* is given by a polynomial.

**Proposition 2.5.** *Let $G \subseteq M_N(\mathbb{C})$ be a group, and let $C$ be a conjugacy classe in $G$. If $F\colon G \to M_N(\mathbb{C})$ is equivariant, then there exists a single-variable polynomial $P_C$ such that $F(A) = P_C(A)$ for all $A \in C$.*

*Proof.* Fix a point $A_0$ in $C$, and let $A_1$ commute with $A_0$. Then since $F$ is equivariant,

$$A_1^{-1}F(A_0)A_1 = F(A_1^{-1}A_0A_1) = F(A_0)$$

which shows that $F(A_0)$ commutes with any such $A_1$: that is, $F(A_0) \in \{A_0\}''$ is in the double commutant of $A_0$. A classical theorem in linear algebra (see, for example, [22] for a short proof) then asserts that there is a single-variable polynomial $P_{A_0}$ such that $F(A_0) = P(A_0)$. Every other point in the conjugacy class $C$ is of the form $A = BA_0B^{-1}$ for some $B \in G$. Since applying a polynomial function to a matrix commutes with conjugation, we have

$$F(A) = F(BA_0B^{-1}) = BF(A_0)B^{-1} = BP_{A_0}(A_0)B^{-1} = P_{A_0}(BA_0B^{-1}) = P_{A_0}(A)$$

which shows that the map $A_0 \mapsto P_{A_0}$ is constant for $A_0 \in C$, so relabel $P_{A_0} = P_C$. Thus, the identity $F(A) = P_C(A)$ holds for all $A \in C$. □

*Remark* 2.6. Proposition 2.5 has the at-first-surprising consequence that the equivariant function $F(A) = A^{-1}$ is equal to a polynomial (not a Laurent polynomial) on any given conjugacy class. This can be seen as a consequence of the Cayley-Hamilton Theorem; cf. Section 2.4. Indeed, let $p_A(\lambda) = \det(\lambda I_N - A)$ be the characteristic polynomail of $A$; then $p_A(A) = 0$. This shows there are coefficients $c_k$ (determined by $A$) so that $\sum_{k=0}^N c_k A^k = 0$. Since $c_0 = (-1)^N \det(A)$, if $A$ is invertible we can therefore factor out $A$ from the $k \geq 1$ terms and solve for $A^{-1}$ as a polynomial in $A$. The above proof shows that this $A$-dependent polynomial is, in fact, uniform over the whole conjugacy class.

## 2.3 Density of Trace Laurent Polynomials

Conceptually, equivariant functions are a natural arena for the Segal–Bargmann transform in the large-$N$ limit. Computationally, it will be convenient to work on the subclass of *trace Laurent polynomials*; cf. (1.20). In fact, trace Laurent polynomials are dense in $L^2(U(N), \rho_s^N; M_N(\mathbb{C}))_{\mathrm{eq}}$. Thus, understanding the action of $\mathbf{B}_{s,t}^N$ on this class tells the full story.

**Theorem 2.7.** *For $s > 0$, the space of trace Laurent polynomials is dense in $L^2(U(N), \rho_s^N; M_N(\mathbb{C}))_{\mathrm{eq}}$.*

We begin by proving that equivariant functions whose entries are polynomials in $U$ and $U^*$ are dense.

**Lemma 2.8.** *Every equivariant function $F \in L^2(U(N), \rho_s^N; M_N(\mathbb{C}))_{\mathrm{eq}}$ can be approximated by a sequence of equivariant matrix-valued functions $F_n$, where each entry of $F_n(U)$ is a polynomial in the entries of $U$ and their conjugates.*

*Proof.* By the Stone–Weierstrass Theorem and the density of continuous functions in $L^2$, any $f \in L^2(\rho_s^N)$ can be approximated by scalar-valued polynomial functions of the entries of the $U(N)$ variable and their conjugates. Applying this result to the components of the matrix-valued function $F$, we see that there is a sequence $P_n$ of polynomials in the entries of $U$ and their conjugates such that

$$\|P_n - F\|_{L^2(U(N),\rho_s^N;M_N(\mathbb{C}))} \to 0. \tag{2.7}$$

Now, consider again the conjugation action $C_V$ of (2.3). It is easy to verify that this action preserves the space of homogeneous polynomials of degree $m$ in the entries $U_{jk}$ and their conjugates. Thus, the averaged function

$$F_n(U) = \int_{U(N)} C_V(P_n)(U)\, dV$$

is still a polynomial in the entries of $U$ and their conjugates; and $F_n$ is evidently equivariant. Therefore $C_V(F) = F$ for each $V \in U(N)$, and so

$$F_n(U) - F(U) = \int_{U(N)} C_V(P_n)(U)\, dV - F(U) = \int_{U(N)} [C_V(P_n) - C_V(F)](U)\, dV.$$

It follows from (2.7) (with an application of Minkowski's inequality and the dominated convergence theorem) that $F_n$ approximates $F$ in $L^2(U(N), \rho_s^N; M_N(\mathbb{C}))$ as claimed. □

*Proof of Theorem 2.7.* We will show that each of the functions $F_n$ in Lemma 2.8 is actually a trace Laurent polynomial. Suppose, then, that $F$ is equivariant and that each entry of $F(U)$ is a polynomial in the entries of $U$ and their conjugates. Let $T(N) \subset U(N)$ denote the diagonal subgroup. By the spectral theorem, any $U \in U(N)$ has a unitary diagonalization $U = V\Lambda V^{-1}$ for some $\Lambda \in T(N)$. The equivariance of $F$ then gives that $F(U) = F(V\Lambda V^{-1}) = VF(\Lambda)V^{-1}$. In particular, any equivariant function $F$ is completely determined by its restriction $F|_{T(N)}$ to the diagonal subgroup.

Because $F$ is equivariant, by the same argument used in the proof of Proposition 2.5, $F(U) \in \{U\}''$ for each $U$. Let $U \in T(N)$ be in the dense subset of matrices with all eigenvalues distinct; then $\{U\}'$ is the set of all diagonal matrices, and so $F(U)$ commutes with all diagonal matrices, meaning that $F(U)$ is diagonal. By the initial assumption on $F$, all entries of $F(U)$ are polynomials in the entries and their conjugates; hence, since the off-diagonal entries are 0 on a dense set, $F(U)$ is diagonal for all $U \in T(N)$, and its diagonal entries are polynomials in the diagonal entries $\lambda_1, \ldots, \lambda_N$ of $U$ and their conjugates. Since this holds true on a dense set of $U \in T(N)$, it holds true on all of $T(N)$. Of course, for $U \in T(N)$ the diagonal entries of $U$ satisfy $\bar{\lambda}_j = 1/\lambda_j$. Thus, each of the diagonal entries of $F|_{T(N)}(U)$ is a Laurent polynomial $q(\lambda_1, \ldots, \lambda_N)$ in the $\lambda_j$'s. The symmetric group $\Sigma_N$ is a subgroup of $U(N)$, so since $F|_{T(N)}$ is equivariant under $U(N)$, it is also equivariant under $\Sigma_N$. Hence each of the (matrix-valued) polynomials $q$ is equivariant under the action of $\Sigma_N$ on the diagonal entries.

Taking $k$ be be larger than the largest negative degree of any variable in $q$, and setting $r(\lambda_1, \ldots, \lambda_N) = (\lambda_1 \cdots \lambda_N)^k q(\lambda_1, \ldots, \lambda_N)$, $r$ is also equivariant under the action of $\Sigma_N$. We can then express

$$F|_{T(N)}(U) = (\lambda_1 \cdots \lambda_N)^{-k} r(\lambda_1, \ldots, \lambda_N) = \det(U^*)^k r(\lambda_1, \ldots, \lambda_N).$$

Since the diagonal entries of $r(\lambda_1, \ldots \lambda_N)$ are equivariant under permutations, the first entry of $r$ must be invariant under permutations of the remaining $N-1$ variables. This means that the first entry of $r$ is a linear combination of terms of the form $\lambda_1^\ell s_\ell(\lambda_2, \ldots, \lambda_N)$, where $\ell$ ranges from 0 up to the degree $d$ of $r$ and $s_\ell$ is a symmetric polynomial in $N-1$ variables. By equivariance under $\Sigma_N$, it now follows that, for $1 \le j \le N$, the $j$th diagonal component of $r$ itself must be a linear combination of terms of the form

$$\left\{ \lambda_j^\ell s_\ell(\lambda_1, \ldots, \widehat{\lambda_j}, \ldots \lambda_N) \colon 0 \le \ell \le d, 1 \le j \le N \right\}.$$

It is well-known that every symmetric polynomial in $N-1$ variables $\lambda_1, \ldots, \lambda_{N-1}$ is a polynomial in power-sums $p_\ell(\lambda_1, \ldots, \lambda_{N-1})$ with $0 \le \ell \le N-1$, where, for any integer $\ell$,

$$p_\ell(\lambda_1, \ldots, \lambda_{N-1}) = \lambda_1^\ell + \lambda_2^\ell + \cdots + \lambda_{N-1}^\ell. \tag{2.8}$$

(This result was known at least to Newton. For a proof, see [26, Theorem 4.3.7].) Furthermore, any power sum in $N-1$ variables can be written as a linear combination of power sums of $N$ variables along with the monomials $\lambda_j^\ell$; for example

$$\sum_{j=2}^{N} \lambda_j^\ell = \left( \sum_{j=1}^{N} \lambda_j^\ell \right) - \lambda_1^\ell.$$

Thus, the first entry of $r$ is actually a polynomial in power-sums of all $N$ variables and in $\lambda_1$ with the remaining entries of $r$ then being determined by equivariance with respect to permutations.

Suppose now that $r$ is the permutation-equivariant polynomial whose $j$th entry is

$$\lambda_j^{\ell_0}\left(\lambda_1^{k_1}+\cdots+\lambda_N^{k_1}\right)^{\ell_1}\cdots\left(\lambda_1^{k_M}+\cdots+\lambda_N^{k_M}\right)^{\ell_M}.$$

Then $r$ is nothing but the restriction to $T(N)$ of the trace polynomial

$$R(U)=U^{\ell_0}\mathrm{Tr}(U^{k_1})^{\ell_1}\cdots\mathrm{Tr}(U^{k_M})^{\ell_M}.$$

Meanwhile, by the above-quoted result, the symmetric polynomial $(\lambda_1\lambda_2\cdots\lambda_N)^k$ can be expressed as a polynomial in the power-sums of the $\lambda_j$'s. Taking the complex-conjugate of this result, we see that $\det(U^*)^k$ can be expressed as a scalar trace polynomial in $U^*$; thus $U\mapsto(\det U^*)^kR(U)$ is a trace Laurent polynomial. Hence $F|_{T(N)}$ is the restriction of the trace Laurent polynomial function $U\mapsto(\det U^*)^kR(U)$, and the result follows since $F$ is determined by $F|_{T(N)}$. □

### 2.4 Asymptotic Uniquness of Trace Laurent Polynomial Representations

The Cayley–Hamilton theorem asserts that, for any matrix $A\in M_N(\mathbb{C})$, it holds that $p_A(A)=0$ where $p_A(\lambda)=\det(\lambda I_N-A)$ is the characteristic polynomial of $A$. In fact, the coefficients of the characteristic polynomial $p_A$ are all scalar trace polynomial functions of $A$: this follows from the Newton identities. In fact, using the operators $\mathcal{M}_{(\cdot)}$ and $\mathcal{A}_+$ of Definition 1.16, there is an explicit formula for $p_A$. Let

$$h_A(\lambda)=\exp\left(-\sum_{m=1}^{\infty}\frac{1}{m\lambda^m}\mathrm{Tr}\left(A^m\right)\right).$$

Then for $A\in M_N(\mathbb{C})$, $p_A(\lambda)=(\mathcal{A}_+\mathcal{M}_{\lambda^N}h_A)(\lambda)$. (See the Wikipedia entry for the Cayley-Hamilton theorem.) Thus, the expression $p_A(A)$ is a(n $N$-dependent) trace polynomial in $A$, and the Cayley–Hamilton theorem asserts that this trace polynomial function vanishes identically on $M_N(\mathbb{C})$. We illustrate this result in the case $N=2$.

*Example* 2.9. For all $A\in M_2(\mathbb{C})$, the Cayley–Hamilton Theorem asserts that

$$A^2-\mathrm{Tr}(A)A+\det(A)I_2=0. \tag{2.9}$$

In the $2\times 2$ case, however, it is easily seen that

$$\det(A)=\frac{1}{2}(\mathrm{Tr}(A)^2-\mathrm{Tr}(A^2)). \tag{2.10}$$

Substituting (2.10) into (2.9) and expressing things in terms of the normalized trace gives

$$A^2-2A\mathrm{tr}(A)+2\mathrm{tr}(A)^2I_2-\mathrm{tr}(A^2)I_2=0$$

for all $A\in M_2(\mathbb{C})$. In particular, if $P\in\mathscr{P}$ denotes the nonzero polynomial $P(u;\mathbf{v})=u^2-2uv_1+2v_1^2-v_2$, then $P_2\colon U(2)\to M_2(\mathbb{C})$ is the zero function. Note, however, that $P_N$ is not the zero function on $U(N)$ for $N>2$, since the minimal polynomial of a generic element of $U(N)$ has degree $N$. This demonstrates the following theorem.

**Theorem 2.10.** *Let $P\in\mathscr{P}\setminus\{0\}$. Then for all sufficiently large $N$, the trace Laurent polynomial function $P_N$ is not identically zero on $U(N)$. In particular, if $P,Q\in\mathscr{P}$ are such that $P_N=Q_N$ for all sufficiently large $N$, then $P=Q$.*

In order to prove Theorem 2.10, the following lemma (from the theory of symmetric functions) is useful. The corresponding statement for symmetric polynomials (rather than Laurent polynomials) is a standard result. The Laurent polynomial case must be known, but is well hidden in the literature.

**Lemma 2.11.** *If $N \geq 2n$, then the power sums $p_k(\lambda_1, \ldots, \lambda_N)$ (cf. (2.8)) with $0 < |k| \leq n$ are algebraically independent elements of the ring of rational function in $N$ variables.*

*Proof.* Let $e_j$ denote the $j$th elementary symmetric polynomial in $N$ variables, that is, the sum of all products of exactly $j$ of the $N$ variables. Then the power sums $p_1, \ldots, p_n$ can be expressed as linear combinations of the functions $e_1, \ldots, e_n$. Thus, it suffices to prove the independence of the functions $e_j(\lambda_1, \ldots, \lambda_N)$ and $e_j(\lambda_1^{-1}, \ldots, \lambda_N^{-1})$ for $1 \leq j \leq n$. We may easily see, however, that

$$e_j(\lambda_1^{-1}, \ldots, \lambda_N^{-1}) = \frac{e_{N-j}(\lambda_1, \ldots, \lambda_N)}{e_N(\lambda_1, \ldots, \lambda_N)}.$$

In the case $N = 2n$, we need to establish the independence of the functions $e_1, \ldots, e_{N/2}$ and $e_{N/2}/e_N, \ldots, e_{N-1}/e_N$, which follows easily from the known indepence of $e_1, \ldots, e_n$ cf. [26, Theorem 4.3.7]. In the case $N > 2n$, if we had an algebraic relation among the functions $e_j(\lambda_1, \ldots, \lambda_N)$ and $e_j(\lambda_1^{-1}, \ldots, \lambda_N^{-1})$ for $1 \leq j \leq n$, we could clear $e_N$ from the denominator to obtain an algebraic relation among the functions $e_1, \ldots, e_n$, $e_{N-1}, \ldots, e_{N-n}$ and $e_N$, which is impossible. □

We now proceed with the scalar version of Theorem 2.10.

**Lemma 2.12.** *Let $Q \in \mathscr{P}_n^0 \setminus \{0\}$, and let $N \geq 2n$. Then $Q_N$ is not identically zero on $U(N)$.*

*Proof.* Let $Q \in \mathscr{P}^0 \setminus \{0\}$; then $Q_N$ is a trace Laurent polynomial, which also defines a holomorphic function on $GL(N, \mathbb{C})$. By uniqueness of analytic continuation, if $Q_N \equiv 0$ on $U(N)$, then $Q_N \equiv 0$ on $GL(N, \mathbb{C})$. To prove the lemma, it therefore suffices to find $A \in GL(N, \mathbb{C})$ with $Q_N(A) \neq 0$. Actually, we will find a diagonal matrix $A \in GL(N, \mathbb{C})$ with $Q_N(A) \neq 0$.

For clarity, we write out the polynomial $Q$ in terms of its coefficients:

$$Q(v_1, v_{-1}, \ldots, v_n, v_{-n}) = \sum_{\substack{i_1, \ldots, i_n \\ j_1, \ldots, j_n}} a_{i_1, \ldots, i_n}^{j_1, \ldots, j_n} \cdot v_1^{i_1} v_{-1}^{j_1} \cdots v_n^{i_n} v_{-n}^{j_n}.$$

Consider any diagonal matrix $\mathrm{diag}(\lambda_1, \ldots, \lambda_N)$ in $GL(N, \mathbb{C})$; for convenience, denote $\lambda = (\lambda_1, \ldots, \lambda_N)$. Then $\mathrm{tr}(A^k) = p_k(\lambda)$ (the power sum of (2.8)), and so

$$Q_N(\mathrm{diag}(\lambda)) = \sum_{\substack{i_1, \ldots, i_n \\ j_1, \ldots, j_n}} a_{i_1, \ldots, i_n}^{j_1, \ldots, j_n} \cdot p_1(\lambda)^{i_1} p_{-1}(\lambda)^{j_1} \cdots p_n(\lambda)^{i_n} p_{-n}(\lambda)^{j_n}. \tag{2.11}$$

By Lemma 2.11, the power sums $p_1(\lambda), p_{-1}(\lambda), \ldots, p_n(\lambda), p_{-n}(\lambda)$ are algebraically independent since $\lambda = (\lambda_1, \ldots, \lambda_N)$ and $N \geq 2n$. Since $Q \neq 0$, some of the coefficients $a_{i_1, \ldots, i_n}^{j_1, \ldots, j_n}$ in (2.11) are $\neq 0$. It follows that $Q_N(\mathrm{diag}(\lambda))$ is not identically 0, as desired. □

This finally brings us to Theorem 2.10.

*Proof of Theorem 2.10.* We write $P(u; \mathbf{v})$ as a sum of positive and negative powers of $u$, multiplied by polynomials in $\mathbf{v}$, where at least one of these coefficients polynomials is nonzero. Let us multiply $P_N(U)$ by $U^k$ for some large $k$, so that all the untraced powers of $U$ in $U^k P_N(U)$ are non-negative. Let $\ell$ be the highest untraced power of $U$ occurring in the expression for $U^k P_N(U)$. Choose $N$ large enough so that $N > \ell$ and so that (Lemma 2.12) the coefficient $q$ of $U^\ell$ in $P_N(U)$ is not identically zero. Then $q$ is nonzero on a nonempty open subset of $U(N)$. This set contains a matrix $U_0$ whose minimal polynomial has degree $N > \ell$. When we evaluate $P_N(U_0)$, the result will be a linear combination of powers of $U_0$ with the coefficient of $U_0^\ell$ being nonzero. Since the minimal polynomial of $U_0$ has degree $N > \ell$, the value of $P_N(U_0)$ is not zero. □

## 3 The Laplacian and Heat Operator on Trace Laurent Polynomials

This section is devoted to a complete description of the action of the Laplacian $\Delta_{U(N)}$ on trace Laurent polynomial functions, and its corresponding lift to $\mathcal{D}_N$ on the space $\mathscr{P}$; cf. Theorem 1.20. We begin by proving "magic formulas" expressing certain quadratic matrix sums in simple forms. We use these to give derivative formulas that allow for the routine computation of $\Delta_{U(N)} P_N$ for any trace Laurent polynomial function $P_N$, and we then use these to prove the intertwining formula of Theorem 1.20. We conclude by proving a more general intertwining formula (Theorem 3.13) for the action of $A^N_{s,t}$ on trace polynomial functions over $GL(N,\mathbb{C})$; in this latter case, we deal more generally with trace Laurent polynomials in $A$ and $A^*$ as this will be of use in Section 4.

### 3.1 Magic Formulas

We define an inner-product on $M_N(\mathbb{C})$ by

$$\langle X, Y\rangle = N\mathrm{Tr}\,(Y^*X) = N^2\mathrm{tr}(Y^*X). \tag{3.1}$$

Restricted to the Lie algebra $\mathfrak{u}(N)$ (consisting of all skew-Hermitian matrices in $M_N(\mathbb{C})$), $\langle\cdot,\cdot\rangle$ is real-valued; it is the polarized inner product corresponding to the norm $\|\cdot\|_{\mathfrak{u}_N}$ of (1.2). (This is not to be confused with the polarized inner-product corresponding to the norm $\|\cdot\|_{M_N}$ of (1.15).)

The main result of this section, which underlies all computations throughout this paper, is the following list of "magic formulas".

**Proposition 3.1.** *Let $\beta_N$ be any orthonormal basis for $\mathfrak{u}(N)$ with respect to the inner-product in (3.1). Then we have the following "magic" formulas: for any $A, B \in M_N(\mathbb{C})$,*

$$\sum_{X\in\beta_N} X^2 = -I_N, \tag{3.2}$$

$$\sum_{X\in\beta_N} XAX = -\mathrm{tr}(A)I_N, \tag{3.3}$$

$$\sum_{X\in\beta_N} \mathrm{tr}(XA)X = -\frac{1}{N^2}A, \tag{3.4}$$

$$\sum_{X\in\beta_N} \mathrm{tr}(XA)\mathrm{tr}(XB) = -\frac{1}{N^2}\mathrm{tr}(AB). \tag{3.5}$$

*Remark* 3.2. Eq. (3.2) is the $A = I_N$ special-case of (3.3); similarly, (3.5) follows from (3.4) by multiplying by $B$ and taking $\mathrm{tr}$. We separate them out as distinct formulas for convenience in repeated use below.

*Proof.* If $\beta_N$ is a basis for the real vector space $\mathfrak{u}(N)$, it is also a basis for the complex vector space $M_N(\mathbb{C}) = \mathfrak{u}(N)\oplus i\mathfrak{u}(N)$. Furthermore, if $\beta_N$ is (real) orthonormal in $\mathfrak{u}(N)$ with respect to the (restricted real) inner product in (3.1), then $\beta_N$ is (complex) orthonormal in $M_N(\mathbb{C})$ with respect to the (complex) inner-product in (3.1).

Thus, let $\tilde{\beta}_N$ be any orthonormal basis for $M_N(\mathbb{C})$ with respect to (3.1), and consider the linear map $\Phi : M_N(\mathbb{C}) \to M_N(\mathbb{C})$ given by

$$\Phi(A) = \sum_{X\in\tilde{\beta}_N} X^*AX.$$

A routine calculation shows that $\Phi$ is independent of the choice of orthonormal basis. We compute $\Phi$ by using the basis

$$\tilde{\beta}_N \equiv \left\{\frac{1}{\sqrt{N}}E_{jk}\right\}_{j,k=1}^N \tag{3.6}$$

where $E_{jk}$ is the $N \times N$ matrix with a 1 in the $(j,k)$-entry and zeros elsewhere. Writing things out in terms of indices shows that, for any $A \in M_N(\mathbb{C})$, we have

$$N \cdot [\Phi(A)]_{\ell m} = \left[ \sum_{j,k=1}^{N} E_{kj} A E_{jk} \right]_{\ell m} = \sum_{j,k,n,o=1}^{N} \delta_{k\ell} \delta_{jn} A_{no} \delta_{jo} \delta_{km} = \sum_{o} A_{oo} \delta_{\ell m},$$

which says that

$$\Phi(A) = \frac{1}{N} \mathrm{Tr}(A) I_N = \mathrm{tr}(A) I_N.$$

The basis-independence of $\Phi$ allows us to replace (3.6) by any real orthonormal basis $\beta_N$ of $\mathfrak{u}(N)$ (which, as noted above, is also a complex orthonormal basis for $M_N(\mathbb{C})$). The elements $X \in \beta_N$ are skew-Hermitian, and thus we obtain

$$\sum_{X \in \beta_N} XAX = -\Phi(A) = -\mathrm{tr}(A) I,$$

which is (3.3).

Meanwhile, if we multiply both sides of (3.4) by $-N^2$ and recall that each $X$ is skew, we see that (3.4) is equivalent to the assertion that

$$A = \sum_{X \in \beta_N} N^2 \mathrm{tr}(X^* A) X = \sum_{X \in \beta_N} \langle A, X \rangle X.$$

But this identity is just the expansion of $A$ in the orthornormal basis $\beta_N$ for $M_N(\mathbb{C})$. Finally, as we have already remarked, (3.2) and (3.5) follow from (3.3) and (3.4), respectively. ☐

### 3.2 Derivative Formulas

**Theorem 3.3.** *Let $m, n \in \mathbb{N}$. Let $\beta_N$ denote an orthonormal basis for $\mathfrak{u}(N)$, and let $X \in \beta_N$. The following hold true:*

$$\partial_X U^n = \sum_{j=1}^{n} U^j X U^{n-j}, \qquad n \geq 0 \tag{3.7}$$

$$\partial_X U^n = -\sum_{j=n+1}^{0} U^j X U^{n-j}, \qquad n < 0 \tag{3.8}$$

$$\partial_X \mathrm{tr}(U^n) = n \cdot \mathrm{tr}\left( X U^n \right), \qquad n \in \mathbb{Z} \tag{3.9}$$

$$\Delta_{U(N)} U^n = -n U^n - 2 \mathbb{1}_{n \geq 2} \sum_{j=1}^{n-1} j U^j \mathrm{tr}(U^{n-j}), \qquad n \geq 0 \tag{3.10}$$

$$\Delta_{U(N)} U^n = n U^n + 2 \mathbb{1}_{n \leq -2} \sum_{j=n+1}^{-1} j U^j \mathrm{tr}(U^{n-j}), \qquad n < 0 \tag{3.11}$$

$$\Delta_{U(N)}\mathrm{tr}(U^n) = -n\mathrm{tr}(U^n) - 2\mathbb{1}_{n\geq 2}\sum_{j=1}^{n-1} j\mathrm{tr}(U^j)\mathrm{tr}(U^{n-j}), \qquad n \geq 0 \tag{3.12}$$

$$\Delta_{U(N)}\mathrm{tr}(U^n) = n\mathrm{tr}(U^n) + 2\mathbb{1}_{n\leq -2}\sum_{j=n+1}^{-1} j\mathrm{tr}(U^j)\mathrm{tr}(U^{n-j}), \qquad n < 0 \tag{3.13}$$

$$\sum_{X\in\beta_N} \partial_X U^m \cdot \partial_X \mathrm{tr}(U^n) = -\frac{mn}{N^2} U^{n+m}, \qquad n, m \in \mathbb{Z} \tag{3.14}$$

$$\sum_{X\in\beta_N} \partial_X \mathrm{tr}(U^m) \cdot \partial_X \mathrm{tr}(U^n) = -\frac{mn}{N^2} \mathrm{tr}(U^{n+m}), \qquad n, m \in \mathbb{Z}. \tag{3.15}$$

*These formulas are valid for all matrices $U \in M_N(\mathbb{C})$; we will normally use them for $U \in U(N)$.*

*Proof.* By the product rule, for $n \geq 0$

$$\partial_X U^n = \frac{d}{dt}\bigg|_{t=0} \left(Ue^{tX}\right)^n = \sum_{j=1}^{n} U^j X U^{n-j}$$

which proves (3.7). Similalry, for $m > 0$

$$\partial_X U^{-m} = \frac{d}{dt}\bigg|_{t=0} \left(e^{-tX}U^{-1}\right)^m = -\sum_{j=0}^{m-1} U^{-j} X U^{-(m-j)}$$

and letting $n = -m$ proves (3.8). Taking traces of (3.7) and (3.8) then gives (3.9) after using $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ repeatedly. Making use of magic formulas (3.2) and (3.3), we then have, for $n \geq 0$

$$\begin{aligned}\Delta_{U(N)}U^n &= 2\mathbb{1}_{n\geq 2}\sum_{1\leq j<k\leq n}\sum_{X\in\beta_N} U\ldots\overbrace{UX}^{j}\ldots\overbrace{UX}^{k}\ldots U + \sum_{j=1}^{n}\sum_{X\in\beta_N} U\ldots\overbrace{UX^2}^{j}\ldots\ldots U \\ &= -2\mathbb{1}_{n\geq 2}\sum_{1\leq j<k\leq n} U^{n-(k-j)}\mathrm{tr}(U^{k-j}) - nU^n.\end{aligned}$$

A little index gymnastics then reduces this last expression to the result in (3.10). An entirely analogous computation proves (3.11). Equations (3.12) and (3.13) result from taking traces of (3.10) and (3.11), since the linear functional $\mathrm{tr}$ commutes with $\Delta_{U(N)}$. Finally, from (3.7) and (3.9), when $m \geq 0$

$$\begin{aligned}\sum_{X\in\beta_N} (\partial_X U^m)\mathrm{tr}(\partial_X U^n) &= n\sum_{X\in\beta_N}\sum_{j=1}^{m} U^j X U^{m-j}\mathrm{tr}(XU^n) \\ &= n\sum_{X\in\beta_N}\sum_{j=1}^{m} U^j \mathrm{tr}(XU^n) X U^{m-j} \\ &= -\frac{n}{N^2}\sum_{j=1}^{m} U^j U^n U^{m-j} = -\frac{mn}{N^2}U^{m+n}.\end{aligned}$$

An analogous computation for $m < 0$ yields the same result, proving (3.14); and taking the trace of this formula gives (3.15). □

*Remark* 3.4. Eq. (3.10) shows that the identity function $\mathrm{id}(U) = U$ on $U(N)$ satisfies $\Delta_{U(N)}\mathrm{id} = -\mathrm{id}$. It follows, for example, that all of the coordinate functions $U \mapsto U_{jk}$ are eigenfunctions of $\Delta_{U(N)}$ with eigenvalue $-1$, *independent of* $n$. This independence suggests that we are, in fact, using the "correct" scaling of the metric on $U(N)$, which in turn determines the scaling of $\Delta_{U(N)}$. If we used the unscaled Hilbert-Schmidt norm on $\mathfrak{u}(N)$, the function id would be an eigenvector for the Laplacian with eigenvalue $-N$; that scaling would not bode well for an infinite dimensional limit of any quantities involving the Laplacian.

To illustrate how Theorem 3.3 may be used, we proceed to determine the action of the heat operator $e^{\frac{t}{2}\Delta_{U(N)}}$ on the polynomial $P_N(U) = U^2$.

*Example* 3.5. Equation (3.10) shows that $\Delta_{U(N)}U^2 = -2U^2 - 2U\mathrm{tr}U$. In order to calculate $\Delta_{U(N)}(U\mathrm{tr}U)$, we use the definition (1.9) of $\Delta_{U(N)}$ and the product rule twice. For each $X \in \mathfrak{u}(N)$,

$$\partial_X^2(U\mathrm{tr}U) = \partial_X\left[(\partial_X U)\cdot \mathrm{tr}U + U\cdot(\partial_X\mathrm{tr}U)\right] = (\partial_X^2 U)\cdot\mathrm{tr}U + 2(\partial_X U)(\partial_X\mathrm{tr}U) + U\cdot\partial_X^2\mathrm{tr}U.$$

Summing over $X \in \mathfrak{u}(N)$ and using (3.10), (3.12), and (3.14) then shows that

$$\Delta_{U(N)}(U\mathrm{tr}U) = (-U)\cdot\mathrm{tr}U - \frac{2}{N^2}U^2 + 2U\cdot(-\mathrm{tr}U) = -\frac{2}{N^2}U^2 - 2U\mathrm{tr}U.$$

Thus, setting $P_N(U) = U^2$ and $Q_N(U) = U\mathrm{tr}(U)$, we have

$$\Delta_{U(N)}P_N = -2P_N - 2Q_N, \tag{3.16}$$

$$\Delta_{U(N)}Q_N = -\frac{2}{N^2}P_N - 2Q_N. \tag{3.17}$$

When $N > 1$, the span of the two functions $P_N, Q_N$ is a 2-dimensional subspace of $C^\infty(U(N))$ (when $N = 1$, $P_N = Q_N$). Equations (3.16)–(3.17) show that this subspace is invariant under the action of $\Delta_{U(N)}$, which is represented there by the matrix

$$D_N = \begin{bmatrix} -2 & -2/N^2 \\ -2 & -2 \end{bmatrix}.$$

The exponentiated matrix $e^{\frac{t}{2}D_N}$ is easily computed (cf. [19, Chapter 2, Exercises 6,7]) as

$$e^{\frac{t}{2}D_N} = e^{-t}\begin{bmatrix} \cosh(t/N) & -1/N\sinh(t/N) \\ -N\sinh(t/N) & \cosh(t/N) \end{bmatrix}.$$

It follows immediately (reading off from the first column of this matrix) that

$$e^{\frac{t}{2}\Delta_{U(N)}}P_N = e^{-t}\cosh(t/N)\,P_N - e^{-t}N\sinh(t/N)Q_N$$

as claimed in (1.5).

Any trace Laurent polynomial function $P_N$ on $U(N)$ is contained in a finite-dimensional subspace of matrix-valued functions that is invariant under $\Delta_{U(N)}$; this is the content of Lemma 3.7 below. Thus, the computation of $e^{\frac{t}{2}\Delta_{U(N)}}P_N$ for any trace Laurent polynomial $P_N$ reduces to exponentiating a matrix of finite size.

### 3.3 Intertwining Formulas I

Recall the operators $\mathcal{T}$, $\mathcal{N}$, $\mathcal{Z}$, $\mathcal{Y}$, and $\mathcal{L}$ from Definitions 1.13 and 1.16. Before we prove that their composite $\mathcal{D}_N$ (1.29) intertwines $\Delta_{U(N)}$, we first prove Lemma 1.19.

*Proof of Lemma 1.19.* The reader may readily verify that $\mathcal{N}$, $\mathcal{Z}_\pm$, $\mathcal{Y}_\pm$, and $\mathcal{L}$ all preserve trace degree. What's more, it is elementaty to calculate that $[\mathcal{T},\mathcal{N}]=0$, while

$$\mathcal{Z}_\pm\mathcal{T}=\mathcal{T}[\mathcal{Z}_\pm+\mathcal{Y}_\pm], \quad \mathcal{Y}_\pm\mathcal{T}=0.$$

Hence, it follows that $\mathcal{D}=-\mathcal{N}-2(\mathcal{Z}+\mathcal{Y})=-\mathcal{N}-2(\mathcal{Z}_++\mathcal{Y}_+)+2(\mathcal{Z}_-+\mathcal{Y}_-)$ commutes with $\mathcal{T}$. Since $\mathcal{D}_N=\mathcal{D}-\frac{1}{N^2}\mathcal{L}$ (cf. (1.28)), we are left only to prove that $[\mathcal{T},\mathcal{L}]=0$. This is also straightforward to compute; instead, we offer an alternative proof. From (1.22), we see that, for any $P\in\mathscr{P}$,

$$[\mathcal{T}\mathcal{D}_N(P)]_N=\mathrm{tr}(\Delta_{U(N)}P_N)=\Delta_{U(N)}\mathrm{tr}(P_N)=[\mathcal{D}_N\mathcal{T}(P)]_N.$$

That is: $([\mathcal{T},\mathcal{D}_N]P)_N\equiv 0$. It follows, using the fact that $[\mathcal{T},\mathcal{D}]=0$, that

$$([\mathcal{T},\mathcal{L}]P)_N=\left([\mathcal{T},N^2(\mathcal{D}_N-\mathcal{D})]P\right)_N=N^2\left([\mathcal{T},\mathcal{D}_N]P\right)_N\equiv 0, \quad \text{for all} \quad N. \tag{3.18}$$

If the polynomial $[\mathcal{T},\mathcal{L}]P$ is not identically 0, (3.18) is in direct contradiction to Proposition 2.10. Thus, for each $P\in\mathscr{P}$, $[\mathcal{T},\mathcal{L}]P=0$, which proves the result. □

We now proceed to prove Theorem 1.20 (the intertwining formula). The following notation will be useful.

**Notation 3.6.** *For $n\in\mathbb{Z}$ and $A\in M_N(\mathbb{C})$ let $W_n(A)=A^n$, $V_n(A)=\mathrm{tr}(A^n)$, and $\mathbf{V}(A)=\{V_n(A)\}_{|n|\geq 1}$. (Technically we should write $V_n^N$ for $V_n$ and $W_n^N$ for $W_n$, but we omit this extra index since the meaning should be clear from the context.) With this notation we have $P_N(U)=P(U,\mathbf{V}(U))I_N$ for $P\in\mathscr{P}$.*

*Proof of Theorem 1.20.* For convenience, we explicitly restate the desired formula as follows:

$$\Delta_{U(N)}P_N=\left[\left(-\mathcal{N}-2\mathcal{Z}-2\mathcal{Y}-\frac{1}{N^2}\mathcal{L}\right)P\right]_N. \tag{3.19}$$

Fix $n\in\mathbb{Z}\setminus\{0\}$, and let $P(u,\mathbf{v})=u^nq(\mathbf{v})$ where $q\in\mathscr{P}^0$; thus $P_N=W_n\cdot q(\mathbf{V})$. For $X\in\mathfrak{u}(N)$, by the product rule we have

$$\partial_XP_N=\partial_X\left[W_n\cdot q(\mathbf{V})\right]=\partial_XW_n\cdot q(\mathbf{V})+W_n\cdot\partial_Xq(\mathbf{V})$$

and therefore

$$\begin{aligned}\Delta_{U(N)}P_N=\sum_{X\in\beta_N}\partial_X^2P_N&=\sum_{X\in\beta_N}\left[\partial_X^2W_n\cdot q(\mathbf{V})+2\partial_XW_n\cdot\partial_Xq(\mathbf{V})+W_n\cdot\partial_X^2q(\mathbf{V})\right]\\ &=\left(\Delta_{U(N)}W_n\right)\cdot q(\mathbf{V})+2\sum_{X\in\beta_N}\partial_XW_n\cdot\partial_Xq(\mathbf{V})+W_n\cdot\left(\Delta_{U(N)}q(\mathbf{V})\right).\end{aligned} \tag{3.20}$$

Using (3.14) and the chain rule, the middle term in (3.20) can be written as

$$\begin{aligned}\sum_{X\in\beta_N}\partial_XW_n\cdot\partial_Xq(\mathbf{V})&=\sum_{X\in\beta_N}\partial_XW_n\cdot\sum_{|k|\geq 1}\left(\frac{\partial}{\partial v_k}q\right)(\mathbf{V})\cdot\partial_XV_k\\ &=\sum_{|k|\geq 1}\left(\sum_{X\in\beta_N}\partial_XW_n\cdot\partial_XV_k\right)\left(\frac{\partial}{\partial v_k}q\right)(\mathbf{V})\\ &=\sum_{|k|\geq 1}\left(-\frac{nk}{N^2}W_{n+k}\right)\left(\frac{\partial}{\partial v_k}q\right)(\mathbf{V})=-\frac{1}{N^2}\sum_{|k|\geq 1}nkW_{n+k}\left(\frac{\partial}{\partial v_k}q\right)(\mathbf{V}).\end{aligned} \tag{3.21}$$

Notice that $nW_{n+k} = W_{k+1} \cdot nW_{n-1} = W_{k+1}\left[\frac{\partial}{\partial u}u^n\right]_N$, and so (3.21) may be written in the form

$$\sum_{X\in\beta_N} \partial_X W_n \cdot \partial_X q(\mathbf{V}) = -\frac{1}{N^2}\left[\sum_{|k|\geq 1} k u^{k+1}\frac{\partial^2}{\partial u \partial v_k}P\right]_N. \tag{3.22}$$

For the last term in (3.20), we again use the chain and product rules repeatedly to find

$$\begin{aligned}\partial_X^2 q(\mathbf{V}) &= \partial_X\left(\sum_{|k|\geq 1}\left(\frac{\partial}{\partial v_k}q\right)(\mathbf{V})\partial_X V_k\right)\\ &= \sum_{|k|\geq 1}\left(\frac{\partial}{\partial v_k}q\right)(\mathbf{V})\cdot\partial_X^2 V_k + \sum_{|j|,|k|\geq 1}\left(\frac{\partial^2}{\partial v_j \partial v_k}q\right)(\mathbf{V})\cdot(\partial_X V_j)(\partial_X V_k).\end{aligned} \tag{3.23}$$

Summing this equation on $X \in \beta_N$, (3.15) shows that the the second sum in (3.23) simplifies to

$$\sum_{X\in\beta_N}\sum_{|j|,|k|\geq 1}\left(\frac{\partial^2}{\partial v_j \partial v_k}q\right)(\mathbf{V})\cdot(\partial_X V_j)(\partial_X V_k) = -\frac{1}{N^2}\sum_{|j|,|k|\geq 1} jkV_{j+k}\cdot\left(\frac{\partial^2}{\partial v_j \partial v_k}q\right)(\mathbf{V}). \tag{3.24}$$

For the first sum in (3.23), we break up the sum over positive and negative terms, and use (3.12) and (3.13) to see that

$$\begin{aligned}\sum_{X\in\beta_N}\sum_{|k|\geq 1}\left(\frac{\partial}{\partial v_k}q\right)(\mathbf{V})\cdot\partial_X^2 V_k &= \sum_{k=1}^{\infty}\left(\frac{\partial}{\partial v_k}q\right)(\mathbf{V})\left(-kV_k - 2\mathbb{1}_{k\geq 2}\sum_{j=1}^{k-1} jV_jV_{k-j}\right)\\ &\quad + \sum_{k=-\infty}^{-1}\left(\frac{\partial}{\partial v_k}q\right)(\mathbf{V})\left(kV_k + 2\mathbb{1}_{k\leq -2}\sum_{j=k+1}^{-1} jV_jV_{k-j}\right)\end{aligned}$$

which is equal to

$$\begin{aligned}&-\sum_{|k|\geq 1}|k|V_k\left(\frac{\partial}{\partial v_k}q\right)(\mathbf{V})\\ &\quad - 2\sum_{k=2}^{\infty}\left(\sum_{j=1}^{k-1} jV_jV_{k-j}\right)\left(\frac{\partial}{\partial v_k}q\right)(\mathbf{V}) + 2\sum_{k=-\infty}^{-1}\left(\sum_{j=k+1}^{-1} jV_jV_{k-j}\right)\left(\frac{\partial}{\partial v_k}q\right)(\mathbf{V}).\end{aligned} \tag{3.25}$$

Combining (3.23)-(3.25) we see that the final term in (3.20) is

$$W_n \cdot \Delta_{U(N)} q(\mathbf{V}) = -[\mathcal{N}_0 P]_N - 2[\mathcal{Z}P]_N - \frac{1}{N^2}\left[\sum_{|j|,|k|\geq 1} jkv_{j+k}\frac{\partial^2}{\partial v_j \partial v_k}P\right]_N$$

and combining this with (3.20) and (3.22) gives

$$\Delta_{U(N)}P_N = \left(\Delta_{U(N)}W_n\right)\cdot q(\mathbf{V}) - \left[\left(\mathcal{N}_0 + 2\mathcal{Z} + \frac{1}{N^2}\mathcal{L}\right)P\right]_N, \tag{3.26}$$

where (3.22) and (3.24) are the terms responsible for $\mathcal{L}$. To address the first term in (3.26), we treat the cases $n \geq 0$ and $n < 0$ separately. When $n \geq 0$, (3.10) gives

$$\left(\Delta_{U(N)}W_n\right)\cdot q(\mathbf{V}) = -nW_n\cdot q(\mathbf{V}) - 2\mathbb{1}_{n\geq 2}\sum_{j=1}^{n-1} jW_jV_{n-j}q(\mathbf{V}).$$

The first term is $-\left[u\frac{\partial}{\partial u}u^n q(\mathbf{v})\right]_N$, and the second is (reindexing $k = n - j$)

$$-2\left[\sum_{k=1}^{n-1} v_k u^{n-k} q(\mathbf{v})\right]_N = -2[\mathcal{Y}_+ P]_N$$

from Example 1.18. An analogous computation in the case $n < 0$, using (3.11), shows that in this case

$$\Delta_{U(N)} W_n \cdot q(\mathbf{V}) = \left[u\frac{\partial}{\partial u}P\right]_N + 2[\mathcal{Y}_- P]_N.$$

Combining these with (3.26) concludes the proof. □

We now prove Corollary 1.22.

*Proof of Corollary 1.22.* For convenience, we restate (1.31): the desired property is

$$\mathcal{D}(PQ) = (\mathcal{D}P)Q + P(\mathcal{D}Q), \qquad P \in \mathscr{P},\ Q \in \mathscr{P}^0.$$

Recall from Definition 1.16 and Lemma 1.19 that $\mathcal{D} = -\mathcal{N} - 2\mathcal{Y} - 2\mathcal{Z} = -(\mathcal{N}_0 + 2\mathcal{Y}) - (\mathcal{N}_1 + 2\mathcal{Z})$, where $\mathcal{N}_1$ and $\mathcal{Z}$ are first order differential operators on $\mathscr{P}$, while $\mathcal{N}_0$ and $\mathcal{Y}$ annihilate $\mathscr{P}^0$ and satisfy

$$\mathcal{N}_0(PQ) = (\mathcal{N}_0 P)Q, \qquad \mathcal{Y}(PQ) = (\mathcal{Y}P)Q, \qquad P \in \mathscr{P},\ Q \in \mathscr{P}^0.$$

Hence

$$(\mathcal{N}_0 + 2\mathcal{Y})(PQ) = [(\mathcal{N}_0 + 2\mathcal{Y})P]Q = [(\mathcal{N}_0 + 2\mathcal{Y})P]Q + P[(\mathcal{N}_0 + 2\mathcal{Y})Q].$$

Since $\mathcal{N}_1 + 2\mathcal{Z}$ satisfies the product rule on $\mathscr{P}$ in general, this proves (1.31); (1.32) follows thence from the standard power series argument. □

$\mathcal{S}$

We conclude this section with the proof of Proposition 1.4.

**Lemma 3.7.** *For $n \geq 0$, let $P \in \mathscr{P}_n$. Then for $t \in \mathbb{R}$, $e^{\frac{t}{2}\mathcal{D}_N} P \in \mathscr{P}_n$.*

*Proof.* Lemma 1.19 gives that $\mathcal{D}_N$ preserves trace degree, and thus $\mathscr{P}_n$ is an invariant subspace for $\mathcal{D}_N$. The result follows. □

*Proof of Proposition 1.4.* Any trace Laurent polynomial function on $U(N)$ has the form $P_N$ for some $P \in \mathscr{P}$. Let $n = \deg(P)$, so that $P \in \mathscr{P}_n$. Theorem 1.20 shows that, for $t \in \mathbb{R}$,

$$e^{\frac{t}{2}\Delta_{U(N)}} P_N = [e^{\frac{t}{2}\mathcal{D}_N} P]_N$$

and Lemma 3.7 asserts that $e^{\frac{t}{2}\mathcal{D}_N} P \in \mathscr{P}_n$. Hence, $e^{\frac{t}{2}\Delta_{U(N)}} P_N$ is a trace Laurent polynomial function on $U(N)$. Since $\mathbf{B}_{s,t}^N P_N$ is the analytic continuation of this trace Laurent polynomial function to $GL(N,\mathbb{C})$ (which is therefore given by the same formula), the result is proved. □

### 3.4 Intertwining Formulas II

This section is devoted to proving an intertwining formula for $GL(N,\mathbb{C})$ (cf. Theorem 3.13) which is analogous to the intertwining formula for $U(N)$ in Theorem 1.20. This result is only needed in order to prove concentration of measures on $GL(N,\mathbb{C})$ (Eq. (1.35) of Theorem 1.26) and hence we do not need as much detailed information about the operators involved. On the other, hand we will now have to consider scalar trace Laurent polynomials in both $A$ and $A^*$, which complicates the notation somewhat.

**Notation 3.8.** *For $n \in \mathbb{N}$, let $\Omega_n$ denote the set of functions (words) $\varepsilon\colon \{1,\ldots,n\} \to \{\pm 1, \pm *\}$. For $\varepsilon \in \Omega_n$, we denote $|\varepsilon| = n$. Set $\Omega = \bigcup_n \Omega_n$. We define the* **word polynomial space** *$\mathscr{W}$ as*

$$\mathscr{W} = \mathbb{C}\left[\{v_\varepsilon\}_{\varepsilon\in\Omega}\right]$$

*the space of polynomials in the indeterminates $\{v_\varepsilon\}_{\varepsilon\in\Omega}$. Of frequent use will be the words*

$$\varepsilon(j,k) = (\overbrace{\pm 1,\ldots,\pm 1}^{|j| \text{ times}}, \overbrace{\pm *,\ldots,\pm *}^{|k| \text{ times}}) \in \Omega_{j+k}, \tag{3.27}$$

*where we use $+1$ in the first slots if $j > 0$ and $-1$ if $j < 0$, and similarly we use $+*$ in the last slots if $k > 0$ and $-*$ if $k < 0$.*

**Notation 3.9.** *For $\varepsilon \in \Omega_n$ and $A \in GL(N,\mathbb{C})$ we define $A^\varepsilon = A^{\varepsilon_1}A^{\varepsilon_2}\cdots A^{\varepsilon_n}$, where $A^{+*} \equiv A^*$ and $A^{-*} \equiv (A^*)^{-1} = (A^{-1})^*$. Given $P \in \mathscr{W}$, we let $P_N\colon GL(N,\mathbb{C}) \to \mathbb{C}$ be the function*

$$P_N(A) = P\left(\mathbf{V}(A)\right)$$

*where*

$$\mathbf{V}(A) = \{V_\varepsilon(A)\colon \varepsilon \in \Omega\}$$

*and*

$$V_\varepsilon(A) = \mathrm{tr}(A^\varepsilon) = \mathrm{tr}\left(A^{\varepsilon_1}A^{\varepsilon_2}\cdots A^{\varepsilon_n}\right).$$

The notation $\mathbf{V}$ here collides with Notation 3.6, but there should be no confusion as to which is being used. As in that case, we should technically write $V_\varepsilon = V_\varepsilon^N$ and $\mathbf{V} = \mathbf{V}^N$, but we suppress the $N$ throughout. Also, in terms of Notation 3.6, note that $V_{\varepsilon(k,0)}(A) = \mathrm{tr}(A^k) = V_k(A)$, while $V_{\varepsilon(0,k)}(A) = \mathrm{tr}((A^*)^k) = V_k(A^*)$. It is therefore natural to think of $\mathscr{P}^0$ as included in $\mathscr{W}$, in the following way.

**Notation 3.10.** *We can identify $\mathscr{P}^0$ as a subalgebra of $\mathscr{W}$ in two ways: $\iota, \iota^*\colon \mathscr{P}^0 \hookrightarrow \mathscr{W}$, with $\iota$ linear and $\iota^*$ conjugate linear, are determined by*

$$\iota(v_k) = v_{\varepsilon(k,0)} \qquad \iota^*(v_k) = v_{\varepsilon(0,k)}. \tag{3.28}$$

*The inclusions $\iota$ and $\iota^*$ intertwine with the evaluation maps as follows: for $Q \in \mathscr{P}^0$,*

$$[\iota(Q)]_N(A) = Q_N(A) \qquad [\iota^*(Q)]_N(A) = Q_N(A)^*. \tag{3.29}$$

The trace degree on $\mathscr{P}^0$ extends consistently to the larger space $\mathscr{W}$.

**Definition 3.11.** *The* **trace degree** *of a monomial $\prod_{i=1}^m v_{\varepsilon_j}^{k_j} \in \mathscr{W}$ is given by*

$$\deg\left(\prod_{j=1}^m v_{\varepsilon_j}^{k_j}\right) = \sum_{j=1}^m |k_j||\varepsilon_j|,$$

*and the trace degree of any element in $\mathscr{W}$ is the highest trace degree of any of its monomial terms. Since $|\varepsilon(k,0)| = |\varepsilon(0,k)| = k$, we have*

$$\deg \iota(Q) = \deg \iota^*(Q) = \deg Q \tag{3.30}$$

*for $Q \in \mathscr{P}^0$. Note, moreover, that $\deg(RS) = \deg(R) + \deg(S)$ for $R, S \in \mathscr{W}$. Finally, for $n \in \mathbb{N}$ we set*

$$\mathscr{W}_n = \{P \in \mathscr{W} : \deg(P) \le n\}.$$

*Note that $\mathscr{W}_n$ is finite dimensional, $\mathscr{W}_n \subset \mathbb{C}[\{v_\varepsilon\}_{|\varepsilon|\le n}]$, and $\mathscr{W} = \bigcup_n \mathscr{W}_n$.*

We now proceed to describe the action of $A^N_{s,t}$ on functions on $U(N)$ or $GL(N,\mathbb{C})$ of the form $R_N$ for some $R \in \mathscr{W}$; recall from (1.11) that

$$A^N_{s,t} \equiv \left(s - \frac{t}{2}\right) \sum_{X\in\beta_N} \partial_X^2 + \frac{t}{2} \sum_{X\in\beta_N} \partial_{iX}^2$$

where $\beta_N$ is an orthonormal basis for $\mathfrak{u}(N)$.

**Theorem 3.12.** *Fix $s, t \in \mathbb{R}$. There are collections $\left\{q_\varepsilon^{s,t} : \varepsilon \in \Omega\right\}$ and $\left\{r_{\varepsilon,\delta}^{s,t} : \varepsilon, \delta \in \Omega\right\}$ in $\mathscr{W}$ with the following properties:*

*(1) for each $\varepsilon \in \Omega$, $q_\varepsilon^{s,t}$ is a certain finite sum of monomials of trace degree $|\varepsilon|$ such that*

$$A^N_{s,t} V_\varepsilon = [q_\varepsilon^{s,t}]_N = q_\varepsilon^{s,t}(\mathbf{V}), \tag{3.31}$$

*(2) for $\varepsilon, \delta \in \Omega$, $r_{\varepsilon,\delta}^{s,t}$ is a certain finite sum of monomials of trace degree $|\varepsilon| + |\delta|$ such that*

$$\left(s - \frac{t}{2}\right) \sum_{X\in\beta_N} (\partial_X V_\varepsilon)(\partial_X V_\delta) + \frac{t}{2} \sum_{X\in\beta_N} (\partial_{iX} V_\varepsilon)(\partial_{iX} V_\delta) = \frac{1}{N^2}[r_{\varepsilon,\delta}^{s,t}]_N = \frac{1}{N^2} r_{\varepsilon,\delta}^{s,t}(\mathbf{V}). \tag{3.32}$$

Please note that the polynomials $q_\varepsilon^{s,t}$ and $r_{\varepsilon,\delta}^{s,t}$ *do not depend on $N$*. The $1/N^2$ in (3.32) comes from the magic formula (3.4), as we will see in the proof.

*Proof.* Fix $\varepsilon \in \Omega$, and let $n = |\varepsilon|$. Let $\beta_N$ denote an orthonormal basis for $\mathfrak{u}(N)$, and let $\beta_+ = \beta_N$ while $\beta_- = i\beta_N$. For any $\xi \in \mathfrak{u}(N) \oplus i\mathfrak{u}(N) = \mathfrak{gl}(N,\mathbb{C}) = M_N(\mathbb{C})$, we make the following conventions:

$$(A\xi)^1 \equiv A\xi, \qquad (A\xi)^{-1} \equiv -\xi A^{-1}, \qquad (A\xi)^* \equiv \xi^* A^*, \qquad (A\xi)^{-*} \equiv -A^*\xi^*. \tag{3.33}$$

Note that, for $\xi \in \beta_\pm$, $\xi^* = \mp\xi$. In the proof to follow, we do not precisely track all of the signs, and so $\pm$ denotes a sign that may be different in different terms and on different sides of an equation. Thus, we have

$$(\partial_\xi V_\varepsilon)(A) = \sum_{j=1}^n \operatorname{tr}\left(A^{\varepsilon_1} A^{\varepsilon_2} \ldots (A\xi)^{\varepsilon_j} \ldots A^{\varepsilon_n}\right)$$

and so

$$\left(\partial_\xi^2 V_\varepsilon\right)(A) = \sum_{j=1}^n \operatorname{tr}\left(A^{\varepsilon_1} A^{\varepsilon_2} \ldots \left(A\xi^2\right)^{\varepsilon_j} \ldots A^{\varepsilon_n}\right) \tag{3.34}$$

$$+ 2 \sum_{1\le j<k\le n} \operatorname{tr}\left(A^{\varepsilon_1} A^{\varepsilon_2} \ldots (A\xi)^{\varepsilon_j} \ldots (A\xi)^{\varepsilon_k} \ldots A^{\varepsilon_n}\right). \tag{3.35}$$

We must now sum over $\xi \in \beta_\pm$. It follows from magic formula (3.2) and convention (3.33) that each term in (3.34) simplifies to

$$\sum_{\xi\in\beta_\pm} \operatorname{tr}\left(A^{\varepsilon_1}A^{\varepsilon_2}\dots\left(A\xi^2\right)^{\varepsilon_j}\dots A^{\varepsilon_n}\right) = \pm\operatorname{tr}\left(A^{\varepsilon_1}A^{\varepsilon_2}\dots A^{\varepsilon_j}\dots A^{\varepsilon_n}\right) = \pm V_\varepsilon(A).$$

To be clear: the $\pm$ on the right varies with $j$ and whether the sum is over $\beta_+$ or $\beta_-$. Summing each of these terms over $1 \le j \le n$ shows that (3.34) summed over $\beta_\pm$ is

$$\sum_{\xi\in\beta_\pm}\sum_{j=1}^{n} \operatorname{tr}\left(A^{\varepsilon_1}A^{\varepsilon_2}\dots\left(A\xi^2\right)^{\varepsilon_j}\dots A^{\varepsilon_n}\right) = n_\pm(\varepsilon)V_\varepsilon(A) \tag{3.36}$$

for some $n_\pm(\varepsilon) \in \mathbb{Z}$. For the terms in (3.35), applying (3.33) shows that

$$\operatorname{tr}\left(A^{\varepsilon_1}A^{\varepsilon_2}\dots(A\xi)^{\varepsilon_j}\dots(A\xi)^{\varepsilon_k}\dots A^{\varepsilon_n}\right) = \pm\operatorname{tr}(A^{\varepsilon^0_{j,k}}\xi A^{\varepsilon^1_{j,k}}\xi A^{\varepsilon^2_{j,k}}) \tag{3.37}$$

where $\{\varepsilon^\ell_{j,k}\}_{\ell=0,1,2}$ are certain substrings of $\varepsilon$, whose concatenation is all of $\varepsilon$: $\varepsilon^0_{j,k}\varepsilon^1_{j,k}\varepsilon^2_{j,k} = \varepsilon$. Applying magic formula (3.3) to (3.37) gives

$$\sum_{\xi\in\beta_\pm} \operatorname{tr}(A^{\varepsilon^0_{j,k}}\xi A^{\varepsilon^1_{j,k}}\xi A^{\varepsilon^2_{j,k}}) = \pm\operatorname{tr}(A^{\varepsilon^0_{j,k}}A^{\varepsilon^2_{j,k}})\operatorname{tr}(A^{\varepsilon^1_{j,k}}) = \pm\operatorname{tr}(A^{\varepsilon_{j,k}})\operatorname{tr}(A^{\varepsilon^1_{j,k}})$$

where $\varepsilon_{j,k} = \varepsilon^0_{j,k}\varepsilon^2_{j,k}$. Note that $|\varepsilon_{j,k}| + |\varepsilon^1_{j,k}| = |\varepsilon|$. Hence, the sum in (3.35) summed over $\beta_\pm$ is equal to

$$\sum_{1\le j<k\le n} \pm\operatorname{tr}(A^{\varepsilon_{j,k}})\operatorname{tr}(A^{\varepsilon^1_{j,k}}) = \sum_{1\le j<k\le n} \pm V_{\varepsilon_{j,k}}(A)V_{\varepsilon^1_{j,k}}(A). \tag{3.38}$$

Hence, if we define

$$q^\pm_\varepsilon = n_\pm(\varepsilon)v_\varepsilon + 2\sum_{1\le j<k\le|\varepsilon|} \pm v_{\varepsilon_{j,k}}v_{\varepsilon^1_{j,k}}, \tag{3.39}$$

which have homogeneous trace degree $|\varepsilon|$, then (3.34)-(3.38) show that

$$q^{s,t}_\varepsilon = \left(s - \frac{t}{2}\right)q^+_\varepsilon + \frac{t}{2}q^-_\varepsilon$$

satisfies (3.31), proving item (1) of the theorem.

For item (2), fix $\delta \in \Omega$ and let $m = |\delta|$. We calculate for each $\xi \in M_N(\mathbb{C})$

$$(\partial_\xi V_\varepsilon)(A)(\partial_\xi V_\delta)(A) = \sum_{j=1}^{n}\sum_{k=1}^{m} \operatorname{tr}(A^{\varepsilon_1}A^{\varepsilon_2}\cdots(A\xi)^{\varepsilon_j}\cdots A^{\varepsilon_n})\cdot\operatorname{tr}(A^{\delta_1}A^{\delta_2}\cdots(A\xi)^{\delta_k}\cdots A^{\delta_m}),$$

again making use of convention (3.33). Using the cyclic property of the trace, we can write the terms in this sum in the form

$$\pm\operatorname{tr}(\xi A^{\varepsilon^{(j)}})\operatorname{tr}(\xi A^{\delta^{(k)}})$$

where $\varepsilon^{(j)}$ is a certain cyclic permutation of $\varepsilon$, and $\delta^{(k)}$ is a certain cyclic permutation of $\delta$. Summing over $\xi \in \beta_\pm$ and using magic formula (3.5), we then have

$$\sum_{\xi\in\beta_\pm} (\partial_\xi V_\varepsilon)(A)(\partial_\xi V_\delta)(A) = \frac{1}{N^2}\sum_{j=1}^{n}\sum_{k=1}^{m} \pm\operatorname{tr}(A^{\varepsilon^{(j)}}A^{\delta^{(k)}}) = \frac{1}{N^2}\sum_{j=1}^{n}\sum_{k=1}^{m} \pm V_{\varepsilon^{(j)}\delta^{(k)}}(A). \tag{3.40}$$

Since $\varepsilon^{(j)}\delta^{(k)}$ has length $|\varepsilon|+|\delta|$, the $\mathscr{W}$ elements

$$r^{\pm}_{\varepsilon,\delta}=\sum_{j=1}^{|\varepsilon|}\sum_{k=1}^{|\delta|}\pm v_{\varepsilon^{(j)}\delta^{(k)}} \tag{3.41}$$

have homogeneous trace degree $|\varepsilon|+|\delta|$, and (3.40) therefore shows that

$$r^{s,t}_{\varepsilon,\delta}=\left(s-\frac{t}{2}\right)r^{+}_{\varepsilon,\delta}+\frac{t}{2}r^{-}_{\varepsilon,\delta} \tag{3.42}$$

satisfies (3.32), proving item (2) of the theorem. □

**Theorem 3.13** (Intertwining Formual II)**.** *Fix $s,t\in\mathbb{R}$. Let $\left\{q^{s,t}_{\varepsilon}\colon \varepsilon\in\Omega\right\}$ and $\left\{r^{s,t}_{\varepsilon,\delta}\colon \varepsilon,\delta\in\Omega\right\}$ be the polynomials from Theorem 3.12 and define*

$$\widetilde{\mathcal{D}}_{s,t}=\frac{1}{2}\sum_{\varepsilon\in\Omega}q^{s,t}_{\varepsilon}\frac{\partial}{\partial w_{\varepsilon}}\qquad \text{and}\qquad \widetilde{\mathcal{L}}_{s,t}=\frac{1}{2}\sum_{\varepsilon,\delta\in\Omega}r^{s,t}_{\varepsilon,\delta}\frac{\partial^{2}}{\partial w_{\varepsilon}\partial w_{\delta}} \tag{3.43}$$

*which are first and second order differential operators on $\mathscr{W}$ which preserve trace degree. Then for all $N\in\mathbb{N}$ and $P\in\mathscr{W}$,*

$$\frac{1}{2}A^{N}_{s,t}P_{N}=\left[\widetilde{\mathcal{D}}_{s,t}P+\frac{1}{N^{2}}\widetilde{\mathcal{L}}_{s,t}P\right]_{N}. \tag{3.44}$$

*Remark* 3.14. Definition 1.6 of $A^{N}_{s,t}$ is stated for $s,t>0$ and $s>t/2$; it is only in this regime that the operator $A^{N}_{s,t}$ is negative-definite and the tools of heat kernel analysis apply. The operator itself is well-defined-for any $s,t\in\mathbb{R}$, however, and it will be convenient to utilize this in some of what follows.

*Proof.* By the chain rule, if $\xi\in M_{N}(\mathbb{C})$ then

$$\begin{aligned}\partial_{\xi}^{2}P_{N}&=\sum_{\varepsilon\in\Omega}\partial_{\xi}\left[\left(\frac{\partial P}{\partial v_{\varepsilon}}\right)(\mathbf{V})\cdot\partial_{\xi}V_{\varepsilon}\right]\\&=\sum_{\varepsilon\in\Omega}\left(\frac{\partial P}{\partial v_{\varepsilon}}\right)(\mathbf{V})\partial_{\xi}^{2}V_{\varepsilon}+\sum_{\varepsilon,\delta\in\Omega}\left(\frac{\partial^{2}P}{\partial v_{\varepsilon}\partial v_{\delta}}\right)(\mathbf{V})\cdot\left(\partial_{\xi}V_{\varepsilon}\right)\left(\partial_{\xi}V_{\delta}\right)\end{aligned}$$

from which it follows that

$$\begin{aligned}A^{N}_{s,t}P_{N}&=\sum_{\varepsilon\in\Omega}\left(\frac{\partial P}{\partial v_{\varepsilon}}\right)(\mathbf{V})\cdot A^{N}_{s,t}V_{\varepsilon}\\&\quad+\sum_{\varepsilon,\delta\in\Omega}\left(\frac{\partial^{2}P}{\partial v_{\varepsilon}\partial v_{\delta}}\right)(\mathbf{V})\left[\left(s-\frac{t}{2}\right)\sum_{\xi\in\beta}\partial_{\xi}V_{\varepsilon}\cdot\partial_{\xi}V_{\delta}+\frac{t}{2}\sum_{\xi\in i\beta}\left(\partial_{\xi}V_{\varepsilon}\right)\left(\partial_{\xi}V_{\delta}\right)\right].\end{aligned}$$

Combining this equation with the results of Theorem 3.12 completes the proof. □

We record one further intertwining formula that will be useful in the proofs of Theorems 1.26 and 1.30.

**Lemma 3.15.** *There exists a sequilinear (conjugate linear in the second variable) form $\mathcal{B}:\mathscr{P}\times\mathscr{P}\to\mathscr{W}$ such that, for all $P,Q\in\mathscr{P}$, we have $\deg\left(\mathcal{B}(P,Q)\right)=\deg(P)+\deg(Q)$ and*

$$[\mathcal{B}(P,Q)]_{N}(A)=\operatorname{tr}[P_{N}(A)Q_{N}(A)^{*}]\qquad \text{for all } A\in GL(N,\mathbb{C}).$$

*Proof.* By sesquilinearity, it suffices to define $\mathcal{B}$ on $P, Q \in \mathscr{P}$ of the form $P(u, \mathbf{v}) = u^k p(\mathbf{v})$ and $Q(u, \mathbf{v}) = u^\ell q(\mathbf{v})$ for $k, \ell \in \mathbb{Z}$ and $p, q \in \mathscr{P}^0$. We compute, for $A \in GL(N, \mathbb{C})$, that

$$\begin{aligned}\operatorname{tr}[P_N(A)Q_N(A)^*] = \operatorname{tr}[A^k p_N(A) A^{*\ell} q_N(A)^*] &= \operatorname{tr}(A^k A^{*\ell}) p_N(A) q_N(A)^* \\ &= [v_{\varepsilon(k,\ell)}]_N(A)[\iota(p)]_N(A)[\iota^*(q)]_N(A)\end{aligned}$$

by (3.29), where $\varepsilon(k, \ell)$ is defined in (3.27). Thus, we take $\mathscr{B}\colon \mathscr{P} \times \mathscr{P} \to \mathscr{W}$ to be the unique sesquilinear form such that, for $p, q \in \mathscr{P}^0$,

$$\mathcal{B}(u^k p, u^\ell q) = w_{\varepsilon(k,\ell)} \iota(p) \iota^*(q).$$

This is trace degree additive by (3.30). This concludes the proof. □

# 4 Limit Theorems

In this section, we prove that the heat kernel measures $\rho_s^N$ on $U(N)$ and $\mu_{s,t}^N$ on $GL(N, \mathbb{C})$ each concentrate all their mass in such a way that the space of trace Laurent polynomial functions collapses onto the space of single variable Laurent polynomial functions. To motivate this, consider the scalar-valued case: if $Q \in \mathscr{P}^0$, then Theorem 1.20 shows that

$$e^{\frac{s}{2}\Delta_{U(N)}}(Q_N) = \left[e^{\frac{s}{2}(\mathcal{D} - \frac{1}{N^2}\mathcal{L})} Q\right]_N = \left[e^{\frac{s}{2}\mathcal{D}} Q\right]_N + O\left(\frac{1}{N^2}\right), \tag{4.1}$$

where the second equality will be made precise in Lemma 4.1 below. Evaluating (4.1) at $I_N$ and using (1.10) shows that

$$\mathbb{E}_{\rho_s^N}(Q_N) = \left(e^{\frac{s}{2}\Delta_{U(N)}} Q_N\right)(I_N) = \left(e^{\frac{s}{2}\mathcal{D}} Q\right)(\mathbf{1}) + O\left(\frac{1}{N^2}\right),$$

where $Q(\mathbf{1}) = Q(1, 1, 1, \ldots)$ is the evaluation of $Q$ at all variables $= 1$. Thus, from Definition 1.24 and Remark 1.25, $\lim_{N\to\infty} \mathbb{E}_{\rho_s^N}(Q_N) = \pi_s Q$. At the same time, the restriction of $\mathcal{D}$ to $\mathscr{P}^0$ is $-(\mathcal{N}_0 + 2\mathcal{Z})$ (cf. Definition 1.16), which is a first-order differential operator. Thus, $e^{\frac{s}{2}\mathcal{D}}$ is an algebra homomorphism on $\mathscr{P}^0$, and so

$$\left[e^{\frac{s}{2}\mathcal{D}} Q^2\right]_N = \left(\left[e^{\frac{s}{2}\mathcal{D}} Q\right]_N\right)^2. \tag{4.2}$$

If $Q$ has real coefficients, then $Q^2 = |Q|^2$, and (1.10) together with (4.1) applied to $Q^2$ and (4.2) evaluated at $\mathbf{1}$ show that

$$\operatorname{Var}_{\rho_s^N}(Q_N) = \int_{U(N)} |Q_N(U)|^2 \, \rho_s^N(dU) - \left|\int_{U(N)} Q_N(U) \, \rho_s^N(dU)\right|^2 = O\left(\frac{1}{N^2}\right).$$

Thus, the random variables $Q_N$ concentrate on their limit mean $\pi_s Q$, summably fast. Section 4.1 fleshes out this argument in the general case; Sections 4.2 and 4.3 then use these ideas to prove Theorems 1.30 and 1.29.

## 4.1 Concentration of Measures

We begin with an abstract result that will be the gist of all our concentration of measure theorems.

**Lemma 4.1.** *Let $V$ be a finite dimensional normed $\mathbb{C}$-space and supposed that $D$ and $L$ are two operators on $V$. Then there exists a constant $C = C(D, L, \|\cdot\|_V) < \infty$ such that*

$$\left\|e^{D+\epsilon L} - e^D\right\|_{\operatorname{End}(V)} \le C\,|\epsilon| \text{ for all } |\epsilon| \le 1, \tag{4.3}$$

*where* $\|\cdot\|_{\mathrm{End}(V)}$ *is the operator norm on* $V$*. It follows that if* $\varphi \in V^*$ *is a linear functional, then*

$$\left|\varphi(e^{D+\epsilon L}x) - \varphi(e^D x)\right| \le C\|\varphi\|_{V^*}\|x\|_V|\epsilon|, \quad x \in V,\ |\epsilon| \le 1, \tag{4.4}$$

*where* $\|\cdot\|_{V^*}$ *is the dual norm on* $V^*$*.*

*Proof.* Using the well known differential of the exponential map (see for example [11, Theorem 1.5.3, p. 23], [19, Theorem 3.5, p. 70], or [27, Lemma 3.4, p. 35]),

$$\begin{aligned}\frac{d}{ds}e^{D+sL} &= e^{D+sL}\int_0^1 e^{-t(D+sL)}Le^{t(D+sL)}dt \\ &= \int_0^1 e^{(1-t)(D+sL)}Le^{t(D+sL)}dt,\end{aligned}$$

we may write

$$e^{D+\epsilon L} - e^D = \int_0^\epsilon \frac{d}{ds}e^{D+sL}ds = \int_0^\varepsilon \left[\int_0^1 e^{(1-t)(D+sL)}Le^{t(D+sL)}dt\right]ds.$$

Crude bounds now show

$$\left\|e^{D+\epsilon L} - e^D\right\|_{\mathrm{End}(V)} \le \int_0^{|\epsilon|}\left[\int_0^1 \left\|e^{(1-t)(D+sL)}Le^{t(D+sL)}\right\|_{\mathrm{End}(V)}dt\right]ds \le C(D,L,\|\cdot\|_V)|\epsilon|.$$

Proving (4.3); (4.4) follows immediately. □

The next lemma identifies the evaluation functional $\pi_{s-t}$ in terms of the operator $\widetilde{\mathcal{D}}_{s,t}$; it demonstrates the way in which Lemma 4.1 will be used throughout the remainder of this section. Recall the evaluation map $\pi_s$ of Definition 1.24, and the inclusion maps $\iota, \iota^*\colon \mathscr{P}^0 \hookrightarrow \mathscr{W}$ of Notation 3.10.

**Lemma 4.2.** *Let* $s,t > 0$ *with* $s > t/2$*. Let* $\widetilde{\mathcal{D}}_{s,t}$ *be given as in (3.43). Then, for any* $Q \in \mathscr{P}^0$*,*

$$[e^{\widetilde{\mathcal{D}}_{s,t}}\iota(Q)](\mathbf{1}) = \pi_{s-t}Q. \tag{4.5}$$

*Proof.* If $f\colon GL(N,\mathbb{C}) \to M_N(\mathbb{C})$ is holomorphic, then $\partial_{iX}f = i\partial_X f$ for all $X \in \mathfrak{u}(N)$, which then implies

$$A^N_{s,t}f\big|_{U(N)} = \left(s - \frac{t}{2}\right)\sum_{X\in\beta_N}\partial_X^2 f - \frac{t}{2}\sum_{X\in\beta_N}\partial_X^2 f = (s-t)\,\Delta_{U(N)}f.$$

Since the scalar trace Laurent polynomial function $Q_N$ is holomorphic, it follows that

$$e^{\frac{1}{2}A^N_{s,t}}Q_N = e^{\frac{1}{2}(s-t)\Delta_{U(N)}}Q_N. \tag{4.6}$$

Using intertwining formulas (3.29) and (3.44) on the left-hand-side of (4.6) and intertwining formula (1.30) on the right-hand-side, we have

$$\left[e^{\widetilde{\mathcal{D}}_{s,t}+\frac{1}{N^2}\widetilde{\mathcal{L}}_{s,t}}\iota(Q)\right]_N = e^{\frac{1}{2}A^N_{s,t}}Q_N = e^{\frac{1}{2}(s-t)\Delta_{U(N)}}Q_N = \left[e^{\frac{1}{2}(s-t)\mathcal{D}_N}Q\right]_N,$$

and evaluating both sides at $I_N$ and using $\mathcal{D}_N = \mathcal{D} - \frac{1}{N^2}\mathcal{L}$ (cf. Lemma 1.19), we have

$$\left(e^{\widetilde{\mathcal{D}}_{s,t}+\frac{1}{N^2}\widetilde{\mathcal{L}}_{s,t}}\iota(Q)\right)(\mathbf{1}) = \left(e^{\frac{1}{2}(s-t)(\mathcal{D}-\frac{1}{N^2}\mathcal{L})}Q\right)(\mathbf{1}). \tag{4.7}$$

Let $n = \deg(Q)$. The linear functional $\varphi(R) = R(\mathbf{1})$ is in continuous on $\mathscr{P}^0_n$ and $\mathscr{W}_n$, and hence Lemma 4.1 allows us to take the limit as $N \to \infty$ in (4.7), yielding

$$\left(e^{\widetilde{\mathcal{D}}_{s,t}}\iota(Q)\right)(\mathbf{1}) = \left(e^{\frac{1}{2}(s-t)\mathcal{D}}Q\right)(\mathbf{1}). \tag{4.8}$$

Finally, since $Q \in \mathscr{P}^0$, the left-hand-side of (4.8) is $\pi_{s-t}Q$; cf. Remark 1.25. This concludes the proof. □

*Remark* 4.3. (1) A similar calculation shows that $\left(e^{\widetilde{\mathcal{D}}_{s,t}}\iota^*(Q)\right)(\mathbf{1}) = \pi_{s-t}(\overline{Q})$.

(2) Properly speaking, the "heat operator" $e^{\frac{1}{2}(s-t)\Delta_{U(N)}}$ in (4.6) is only defined when $s > t$. However, since $Q_N$ is a trace Laurent polynomial function, this expression can be made sense of as a convergent power-series, and the intertwining formulas still apply.

We now proceed with the proof of Theorem 1.26.

*Proof of Theorem 1.26.* We begin with the proof of (1.35). By the triangle inequality, it suffices to prove the theorem for polynomails of the form $P(u, \mathbf{v}) = u^k Q(\mathbf{v})$ for $k \in \mathbb{Z}$ and $Q \in \mathscr{P}^0$ scalar. Therefore

$$P(u, \mathbf{v}) - \pi_{s-t}P(u, \mathbf{v}) = u^k[Q(\mathbf{v}) - \pi_{s-t}Q] = u^k R_{s-t}(\mathbf{v})$$

where $R_{s-t} = Q - \pi_{s-t}Q$. Note that $\pi_{s-t}R_{s-t} = 0$. Now, for $A \in GL(N, \mathbb{C})$,

$$\begin{aligned}\|P_N(A) - (\pi_s P)_N(A)\|^2_{M_N} &= \mathrm{tr}(A^k[R_{s-t}]_N(A)[R_{s-t}]_N(A)^* A^{*k}) \\ &= \mathrm{tr}(A^k A^{*k})[R_{s-t}]_N(A)[R_{s-t}]_N(A)^*. \end{aligned} \tag{4.9}$$

Thus

$$\|[P]_N(A) - (\pi_{s-t}P)_N(A)\|^2_{M_N} = [v_{\varepsilon(k,k)}\iota(R_{s-t})\iota^*(R_{s-t})]_N(A) \tag{4.10}$$

where, in the case $k = 0$, we interpret $v_{\varepsilon(0,0)} = 1$. We calculate the $L^2(\mu^N_{s,t})$-norm of $P - \pi_{s-t}P = u^k R_{s-t}$ using (1.10). Thus, using the intertwining formula (3.44) and (4.10), we have

$$\begin{aligned}\|P_N - (\pi_s P)_N\|^2_{L^2(\mu^N_{s,t})} &= e^{\frac{1}{2}A^N_{s,t}}\left(\|P_N - (\pi_{s-t}P)_N\|^2_{M_N}\right)(I_N) \\ &= \left(e^{\widetilde{\mathcal{D}}_{s,t}+\frac{1}{N^2}\widetilde{\mathcal{L}}_{s,t}}\left(v_{\varepsilon(k,k)}\iota(R_{s-t})\iota^*(R_{s-t})\right)\right)(\mathbf{1}). \end{aligned} \tag{4.11}$$

Now, let $n = \deg Q = \deg R_{s-t}$. Using the linear functional $\varphi(R) = R(\mathbf{1})$ on $\mathscr{W}_{2n}$, Lemma 4.1 then yields

$$\left|\left(e^{\widetilde{\mathcal{D}}_{s,t}+\frac{1}{N^2}\widetilde{\mathcal{L}}_{s,t}}\left(v_{\varepsilon(k,k)}\iota(R_{s-t})\iota^*(R_{s-t})\right)\right)(\mathbf{1}) - \left(e^{\widetilde{\mathcal{D}}_{s,t}}\left(v_{\varepsilon(k,k)}\iota(R_{s-t})\iota^*(R_{s-t})\right)\right)(\mathbf{1})\right| = O\left(\frac{1}{N^2}\right). \tag{4.12}$$

But, since $\widetilde{\mathcal{D}}_{s,t}$ is a first-order differential operator acting on $\mathscr{W}_{2n}$, $e^{\widetilde{\mathcal{D}}_{s,t}}$ is an algebra homomorphism, and we have

$$e^{\widetilde{\mathcal{D}}_{s,t}}\left(v_{\varepsilon(k,k)}\iota(R_{s-t})\iota^*(R_{s-t})\right) = e^{\widetilde{\mathcal{D}}_{s,t}}v_{\varepsilon(k,k)} \cdot e^{\widetilde{\mathcal{D}}_{s,t}}\iota(R_{s-t}) \cdot e^{\widetilde{\mathcal{D}}_{s,t}}\iota^*(R_{s-t}) = 0 \tag{4.13}$$

since $e^{\widetilde{\mathcal{D}}_{s,t}}\iota(R_{s-t}) = \pi_{s-t}R_{s-t} = 0$ by Lemma 4.2. Thus, (4.11)-(4.13) prove (1.35).

Note that $\frac{s}{2}\Delta_{U(N)} = \frac{1}{2}A^N_{s,0}$; thus taking $t = 0$ in (4.12) and restricting the function to $U(N)$ also proves (1.34), concluding the proof. □

## 4.2 Proof of Main Limit Theorem 1.30

*Proof of Theorem 1.30.* Let $f \in \mathscr{P}^1$; then by the intertwining formula (1.30),

$$e^{\frac{t}{2}\Delta_{U(N)}}f_N = [e^{\frac{t}{2}\mathcal{D}_N}f]_N,$$

where $\mathcal{D}_N$ is defined in (1.29).

The function on the right is a trace Laurent polynomial function of $U \in U(N)$ (with no $U^*$s), and therefore its analytic continuation to $GL(N, \mathbb{C})$ is given by *the same* trace polynomial function in $A \in GL(N, \mathbb{C})$. Thus

$$[\mathbf{B}^N_{s,t}f_N](A) = [e^{\frac{t}{2}\mathcal{D}_N}f]_N(A), \qquad A \in GL(N, \mathbb{C}).$$

Hence

$$\|\mathbf{B}_{s,t}^N f_N - [\mathbf{B}_{s,t} f]_N\|_{L^2(\mu_{s,t}^N)} = \|[e^{\frac{t}{2}\mathcal{D}_N} f]_N - [\pi_{s-t}\circ e^{\frac{t}{2}\mathcal{D}} f]_N\|_{L^2(\mu_{s,t}^N)}.$$

By the triangle inequality, the last quantity is

$$\leq \|[e^{\frac{t}{2}\mathcal{D}_N} f]_N - [e^{\frac{t}{2}\mathcal{D}} f]_N\|_{L^2(\mu_{s,t}^N)} + \|[e^{\frac{t}{2}\mathcal{D}} f]_N - [\pi_{s-t}\circ e^{\frac{t}{2}\mathcal{D}} f]_N\|_{L^2(\mu_{s,t}^N)}. \tag{4.14}$$

The second term in (4.14) is $O(1/N)$ by (1.35) (Theorem 1.26). Thus, to complete the proof of (1.37), it suffices to show that

$$\|[e^{\frac{t}{2}\mathcal{D}_N} f]_N - [e^{\frac{t}{2}\mathcal{D}} f]_N\|^2_{L^2(\mu_{s,t}^N)} = O\left(\frac{1}{N^2}\right) \tag{4.15}$$

for each $f \in \mathscr{P}^1 = \mathbb{C}[u,u^{-1}]$. Let $n = \deg f$, let $\mathcal{B} : \mathscr{P}\times\mathscr{P}\to\mathscr{W}$ be the sesquilinear form in Lemma 3.15, and let $R^{(N)} = e^{\frac{t}{2}\mathcal{D}_N} f - e^{\frac{t}{2}\mathcal{D}} f$. Then by (1.16) and (3.44), the left side of (4.15) is given by

$$\|[R^{(N)}]_N\|^2_{L^2(\mu_{s,t}^N)} = e^{\frac{1}{2}A_{s,t}^N}\left(\|[R^{(N)}]_N\|^2_{M_N}\right) = \left(e^{\widetilde{\mathcal{D}}_{s,t}+\frac{1}{N^2}\widetilde{\mathcal{L}}_{s,t}}\mathcal{B}(R^{(N)},R^{(N)})\right)(\mathbf{1}). \tag{4.16}$$

Using the linear functional $\varphi(P) = P(\mathbf{1})$ on $\mathscr{W}_{2n}$ and any norm $\|\cdot\|_{\mathscr{W}_{2n}}$, Lemma 4.1 ensures there is a constant $C$ (depending on $n,s,t$ but *not* on $N$) such that

$$\left|\left(e^{\widetilde{\mathcal{D}}_{s,t}+\frac{1}{N^2}\widetilde{\mathcal{L}}_{s,t}}\mathcal{B}(R^{(N)},R^{(N)})\right)(\mathbf{1}) - \left(e^{\widetilde{\mathcal{D}}_{s,t}}\mathcal{B}(R^{(N)},R^{(N)})\right)(\mathbf{1})\right| \leq \frac{C}{N^2}\|\mathcal{B}(R^{(N)},R^{(N)})\|_{\mathscr{W}_{2n}}. \tag{4.17}$$

Let $\psi(P) = \left(e^{\widetilde{\mathcal{D}}_{s,t}}P\right)(\mathbf{1})$, another linear functional on $\mathscr{W}_{2n}$; then

$$\left|\left(e^{\widetilde{\mathcal{D}}_{s,t}}\mathcal{B}(R^{(N)},R^{(N)})\right)(\mathbf{1})\right| \leq \|\psi\|^*_{2n}\|\mathcal{B}(R^{(N)},R^{(N)})\|_{\mathscr{W}_{2n}}.$$

This, in conjunction with (4.16) and (4.17), shows that

$$\|[R^{(N)}]_N\|^2_{L^2(\mu_{s,t}^N)} \leq \left(\|\psi\|^*_{2n} + \frac{C}{N^2}\right)\|\mathcal{B}(R^{(N)},R^{(N)})\|_{\mathscr{W}_{2n}}. \tag{4.18}$$

Since $\mathcal{B} : \mathscr{P}_n\times\mathscr{P}_n\to\mathscr{W}_{2n}$ is sesquilinear with finite dimensional domain and range, it is bounded with any choice of norms; in particular, given any norm $\|\cdot\|_{\mathscr{P}_n}$ on $\mathscr{P}_n$, there is a constant $C'$ (depending on $n$ but not $N$) so that

$$\|\mathcal{B}(P,Q)\|_{\mathscr{W}_{2n}} \leq C'\|P\|_{\mathscr{P}_n}\|Q\|_{\mathscr{P}_n} \qquad \text{for all } P,Q\in\mathscr{P}_n.$$

Together with (4.18), this yields

$$\|[R^{(N)}]_N\|^2_{L^2(\mu_{s,t}^N)} \leq C'\left(\|\psi\|^*_{2n} + \frac{C}{N^2}\right)\|R^{(N)}\|^2_{\mathscr{P}_n}. \tag{4.19}$$

Finally, Lemma 4.1 gives

$$\|R^{(N)}\|_{\mathscr{P}_n} = \|e^{\frac{t}{2}[\mathcal{D}-\frac{1}{N^2}\mathcal{L}]} f - e^{\frac{t}{2}\mathcal{D}} f\|_{\mathscr{P}_n} = O\left(\frac{1}{N^2}\right)$$

which proves (4.15). (In fact it shows this term is $O(1/N^4)$; however, since the square of the second term in (4.14) is $O(1/N^2)$, this faster convergence doesn't improve matters.)

The proof of (1.38) is similar: the restriction of $(\mathbf{B}_{s,t}^N)^{-1} f_N$ to $U(N)$ is simply $e^{-\frac{t}{2}\Delta_{U(N)}} f_N$, and a similar triangle inequality argument now using (1.34) shows that it suffices to prove

$$\|[e^{-\frac{t}{2}\mathcal{D}_N} f]_N - [e^{-\frac{t}{2}\mathcal{D}} f]_N\|^2_{L^2(\rho_s^N)} = O\left(\frac{1}{N^2}\right). \tag{4.20}$$

The argument now proceeds identically to above, by redefining $R^{(N)}$ with the substitution $t\mapsto -t$, and taking all norms with the substitution $(s,t)\mapsto(s,0)$ in all formulas from (4.16) onward. □

### 4.3 Limit Norms and the Proof of Theorems 1.29

We begin by proving that the transforms $\mathbf{B}_{s,t}$ and $\mathbf{H}_{s,t}$ are invertible on $\mathscr{P}^1$. (This will be subsumed by Theorem 1.29, but it will be useful to have this fact in the proof.)

**Lemma 4.4.** *$\mathbf{B}_{s,t}$ and $\mathbf{H}_{s,t}$ are invertible operators on $\mathscr{P}^1_n$ for each $n > 0$, and hence on $\mathscr{P}^1$.*

*Proof.* Consider $e^{\pm\frac{t}{2}\mathcal{D}}$ restricted to $\mathscr{P}_n$. Expanding as power-series, a straightforward induction using the forms of the composite operators $\mathcal{N}$, $\mathcal{Z}$, and $\mathcal{Y}$ shows that there exist $q_k^t \in \mathscr{P}^0$ with

$$e^{\pm\frac{t}{2}\mathcal{D}}u^n = e^{\mp\frac{n}{2}t}u^n + \sum_{k=0}^{n-1} q_k^t(\mathbf{v})u^k,$$

$$e^{\pm\frac{t}{2}\mathcal{D}}u^{-n} = e^{\pm\frac{n}{2}t}u^{-n} + \sum_{k=-n+1}^{0} q_k^t(\mathbf{v})u^k.$$

This shows that $e^{\pm\frac{t}{2}\mathcal{D}}$ preserves $\mathscr{P}_n^{1,+}$ and $\mathscr{P}_n^{1,-} \oplus \mathbb{C}$. Incorporating the evaluation maps $\pi_s$ or $\pi_{s-t}$, we find that

$$\mathbf{B}_{s,t}(u^{\pm n}), \mathbf{H}_{s,t}(u^{\pm n}) \in e^{\pm\frac{n}{2}t}u^{\pm n} + \mathscr{P}^1_{n-1}$$

Consider, then, the standard basis $\{1, u^1, \ldots, u^n\}$ of $\mathscr{P}_n^{1,+}$; it follows that, in this basis, $\mathbf{B}_{s,t}|_{\mathscr{P}_n^{1,+}}$ and $\mathbf{H}_{s,t}|_{\mathscr{P}_n^{1,+}}$ are upper-triangular, with diagonal entries $e^{\mp\frac{k}{2}t}$ for $0 \le k \le n$. Thus the restrictions of $\mathbf{B}_{s,t}$ and $\mathbf{H}_{s,t}$ to $\mathscr{P}_n^{1,+}$ are invertible. A similar argument shows the invertibility on $\mathscr{P}_n^{1,-}$, thus yielding the result on $\mathscr{P}_n$. Since $\mathscr{P}^1 = \bigcup_n \mathscr{P}^1_n$, the proof is complete. □

We now introduce two seminorms on $\mathscr{P}$.

**Definition 4.5.** *Let $s, t > 0$ with $s > t/2$. For each $N$, define the seminorms $\|\cdot\|_{s,N}$ and $\|\cdot\|^{s,t,N}$ on $\mathscr{P}$ by*

$$\|P\|_{s,N} = \|P_N\|_{L^2(U(N),\rho_s^N;M_N(\mathbb{C}))} \tag{4.21}$$

$$\|P\|^{s,t,N} = \|P_N\|_{L^2(GL(N,\mathbb{C}),\mu_{s,t}^N;M_N(\mathbb{C}))}. \tag{4.22}$$

In fact, for any $n > 0$ and sufficiently large $N$, seminorms (4.21) and (4.22) are actually *norms* when restricted to $\mathscr{P}_n$. Indeed, if $\|P\|_{s,N} = 0$ then $P_N = 0$ in $L^2(U(N), \rho_s^N; M_N(\mathbb{C}))$, and since $P_N$ is a smooth function and $\rho_s^N$ has a strictly positive density, this means $P_N$ is identically 0. By Proposition 2.10, when $N$ is sufficiently large (relative to $n$) it follows that $P = 0$.

For $P \in \mathscr{P}$, define

$$\|P\|_s = \lim_{N\to\infty} \|P\|_{s,N} \tag{4.23}$$

$$\|P\|^{s,t} = \lim_{N\to\infty} \|P\|^{s,t,N}. \tag{4.24}$$

These are also seminorms on $\mathscr{P}$, but they are *not* norms on all of $\mathscr{P}$, or even on $\mathscr{P}_n$ for any $n > 1$. However, restricted to *single-variable Laurent polynomials* $\mathscr{P}^1$, they are in fact norms. To prove this, we look to the free unitary Brownian motion distribution $\nu_s$; cf. Remark 1.27. The measure $\nu_s$ is the weak limit of $\nu_s^N$ of (2.1) (which exists by the Lévy continuity theorem). In [4, Proposition 10, p. 270], it is shown that $\nu_s$ is absolutely continuous with respect to Lebesgue measure on $\mathbb{T}$, with a continuous density that is strictly positive in a neighborhood of $1 \in \mathbb{T}$; we will need this result (in particular the fact that $\mathrm{supp}(\nu_s)$ is not a finite set) in the following.

**Lemma 4.6.** *The seminorms (4.23) and (4.24) are norms on the subspace $\mathscr{P}^1 \subset \mathscr{P}$ of single-variable Laurent polynomials.*

*Proof.* We begin with norm (4.23). Identify the Laurent polynomial $P \in \mathscr{P}^1$ as a trigonometric polynomial function $P_{\mathbb{T}}$ on the unit circle $\mathbb{T}$. Then (2.2) shows that

$$\|P\|_{s,N} = \|P_N\|_{L^2(U(N),\rho_s^N;M_N(\mathbb{C}))} = \|P_{\mathbb{T}}\|_{L^2(\mathbb{T},\nu_s^N)}.$$

Thus, since $\nu_s^N \rightharpoonup \nu_s$,

$$\|P\|_s = \lim_{N\to\infty} \|P_{\mathbb{T}}\|_{L^2(\mathbb{T},\nu_s^N)} = \|P_{\mathbb{T}}\|_{L^2(\mathbb{T},\nu_s)}. \tag{4.25}$$

Since the support of $\nu_s$ is infinite, (4.25) shows that seminorm (4.23) is indeed a norm on $\mathscr{P}^1$.

For seminorm (4.24), we will utilize the isometry property of the finite dimensional Segal–Bargmann transform $\mathbf{B}_{s,t}^N$. Fix $Q \in \mathscr{P}^1$; then there is some finite $n$ for which $Q \in \mathscr{P}_n^1$. By Lemma 4.4, there is a unique Laurent polynomial $P \in \mathscr{P}_n^1$ so that $\mathbf{B}_{s,t}(P) = Q$. Thus

$$\|Q\|^{s,t} = \|\mathbf{B}_{s,t}P\|^{s,t} = \lim_{N\to\infty} \|\mathbf{B}_{s,t}P\|^{s,t,N}.$$

By Theorem 1.30 and (4.22) we have

$$\lim_{N\to\infty} \|\mathbf{B}_{s,t}P\|^{s,t,N} = \lim_{N\to\infty} \|[\mathbf{B}_{s,t}^N P]_N\|_{L^2(GL(N,\mathbb{C}),\mu_{s,t}^N;M_N(\mathbb{C}))}$$

and by the isometry property of the Segal–Bargmann transform, we therefore have

$$\|Q\|^{s,t} = \lim_{N\to\infty} \|P_N\|_{L^2(U(N)),\rho_s^N;M_N(\mathbb{C}))} = \|P\|_s.$$

Thus, if $\|Q\|^{s,t} = 0$ then $\|P\|_s = 0$, so $Q = \mathbf{B}_{s,t}(0) = 0$. This concludes the proof. □

*Remark* 4.7. Eq. (4.25) shows that norm (4.23) is just an $L^2$-norm, with respect to a well-understood measure. Norm (4.24) is, at present, much more mysterious. In [4], a great deal of work is spent trying to understand this norm in the case $s = t$. It can, in that case, be identified as the norm of a certain reproducing kernel Hilbert space, built out of holomorphic functions on a bounded region $\Sigma_t \subset \mathbb{C} \setminus \{0\}$ that has no obvious symmetries, and which becomes non-simply-connected when $t \geq 4$. Understanding the norm (4.24) in general is a goal for future research of the present authors.

This leads us to the proof of Theorem 1.29.

*Proof of Theorem 1.29.* Fix $P \in \mathscr{P}^1$, and consider the Laurent polynomial $\mathbf{B}_{s,t}\mathbf{H}_{s,t}P \in \mathscr{P}^1$. By definition

$$\begin{aligned}\|\mathbf{B}_{s,t}\mathbf{H}_{s,t}P - P\|^{s,t} &= \lim_{N\to\infty} \|\mathbf{B}_{s,t}\mathbf{H}_{s,t}P - P\|^{s,t,N} \\ &= \lim_{N\to\infty} \|[\mathbf{B}_{s,t}\mathbf{H}_{s,t}P]_N - P_N\|_{L^2(\mu_{s,t}^N)}.\end{aligned} \tag{4.26}$$

The triangle inequality yields

$$\|[\mathbf{B}_{s,t}\mathbf{H}_{s,t}P]_N - P_N\|_{L^2(\mu_{s,t}^N)} \leq \|[\mathbf{B}_{s,t}\mathbf{H}_{s,t}P]_N - \mathbf{B}_{s,t}^N[\mathbf{H}_{s,t}P]_N\|_{L^2(\mu_{s,t}^N)} + \|\mathbf{B}_{s,t}^N[\mathbf{H}_{s,t}P]_N - P_N\|_{L^2(\mu_{s,t}^N)}.$$

Applying (1.37) with $f = \mathbf{H}_{s,t}P$ shows that the first term is $O(1/N)$. For the second term, we use the isometry property of the Segal–Bargmann transform: since $P_N$ is a single-variable (i.e. holomorphic) Laurent polynomial, it is in the range of $\mathbf{B}_{s,t}^N$, and so

$$\begin{aligned}\|\mathbf{B}_{s,t}^N[\mathbf{H}_{s,t}P]_N - P_N\|_{L^2(\mu_{s,t}^N)} &= \|\mathbf{B}_{s,t}^N\left([\mathbf{H}_{s,t}P]_N - (\mathbf{B}_{s,t}^N)^{-1}P_N\right)\|_{L^2(\mu_{s,t}^N)} \\ &= \|[\mathbf{H}_{s,t}P]_N - (\mathbf{B}_{s,t}^N)^{-1}P_N\|_{L^2(\rho_s^N)} = O\left(\frac{1}{N}\right),\end{aligned}$$

by (1.38). Hence, the quantity in the limit on the right-hand-side of (4.26) is $O(1/N)$, so its limit is 0. We therefore have $\|\mathbf{B}_{s,t}\mathbf{H}_{s,t}P - P\|^{s,t} = 0$. Lemma 4.6 shows that $\|\cdot\|^{s,t}$ is a norm on $\mathscr{P}^1$, and so it follows that $\mathbf{B}_{s,t}\mathbf{H}_{s,t}P - P = 0$. Hence, since $\mathbf{B}_{s,t}$ and $\mathbf{H}_{s,t}$ are known to be invertible (Lemma 4.4), it follows that $\mathbf{H}_{s,t} = \mathbf{B}_{s,t}^{-1}$ as desired. □

# 5 The Free Unitary Segal–Bargmann Transform

In this final section, we identify the limit Segal–Bargmann transform $\mathbf{B}_{s,t}$, which has been constructed (Definition 1.28) as a linear operator on the space $\mathscr{P}^1$ of single-variable Laurent polynomials. We will characterize the *Biane polynomials* for $\mathbf{B}_{s,t}$:

$$p_k^{s,t} = \mathbf{H}_{s,t}((\,\cdot\,)^k) = \pi_s \circ e^{-\frac{t}{2}\mathcal{D}}(\,\cdot\,)^k, \qquad k \in \mathbb{Z} \tag{5.1}$$

defined so that

$$\mathbf{B}_{s,t}(p_k^{s,t})(z) = z^k$$

when $s > t/2 > 0$. We call them Biane polynomials since, as we will prove, they match the polynomials that Biane introduced in [4, Lemma 18] to characterize the free unitary Segal–Bargmann transform $\mathscr{G}^t$, in the specal case $s = t$. There is classical motivation to understand these polynomials. Consider the 1-dimensional classical Segal–Bargmann transform $S_t^1$. Since polynomials are dense in the Gaussian $L^2$-spaces forming the domain and image of $S_t^1$, its action is completely determined by the polynomials $H_k(t, \cdot)$ satusfying $S_t^1(H_k(t,\cdot))(z) = z^k$. In this case, since the measure $\gamma_{t/2}^2$ is rotationally-invariant, the monomails $z \mapsto z^k$ are orthogonal, and since $S_t^1$ is an isometry, it follows that $H_k(t,\cdot)$ are the orthogonal polynomials of the Gaussian measure $\gamma_t^1$: the *Hermite polynomials* of (variance $t/2$). Hence, the Biane polynomials are the unitary version of the Hermite polynomials. We will determine the generating function $\Pi$ of these polynomials; cf. (1.39). In the case $s = t$, this precisely matches the generating function in [4, Lemma 18]; in this way, we verify that our limit Segal–Bargmann transform *is* the aforementioned free unitary Segal–Bargmann transform $\mathscr{G}^t$.

Before proceeding, we make an observation. It is immediate from the form of the operators $\mathcal{N}$, $\mathcal{Z}$, and $\mathcal{Y}$ in Definition 1.16 that $\mathcal{D} = -\mathcal{N} - 2\mathcal{Z} - 2\mathcal{Y}$ satisfies

$$\mathcal{D}(u^{-k}) = \left(\mathcal{D}(\,\cdot\,)^k\right)(u^{-1}), \qquad k \in \mathbb{Z}.$$

Expanding $e^{-\frac{t}{2}\mathcal{D}}$ as a power series shows that the semigroup also commutes with the reciprocal map, and applying the algebra homomorphism $\pi_s$ then shows that

$$p_{-k}^{s,t}(u) = p_k^{s,t}(u^{-1}), \qquad k \in \mathbb{Z}. \tag{5.2}$$

Note also that $\mathcal{D}$ preserves the subspaces $\mathscr{P}^{1,\pm} \otimes \mathscr{P}^0$, and hence $p_k^{s,t}(u)$ is a polynomial in $u$ for $k \geq 0$, while $p_k^{s,t}(u)$ is a polynomial in $u^{-1}$ for $k < 0$. Hence, since $p_0^{s,t} = 1$, it will suffice to identify $p_k^{s,t}$ only for $k \geq 1$.

## 5.1 Biane Polynomials and Differential Recursion

It will be convenient to look at the related family of "unprojected" polynomials.

**Definition 5.1.** *For $t \in \mathbb{R}$ and $k \in \mathbb{N}$, define $B_k^t \in \mathscr{P}$ and $C_k^t \in \mathscr{P}^0$ by*

$$B_k^t(u, \mathbf{v}) = e^{-\frac{k}{2}t} e^{-\frac{t}{2}\mathcal{D}} u^k \qquad \textit{and} \qquad C_k^t(\mathbf{v}) = \mathcal{T}(B_k^t)(\mathbf{v}), \tag{5.3}$$

*where $\mathcal{T}$ is the tracing map of (1.21). For $s \in \mathbb{R}$, define $b_k(s,t,\,\cdot\,) \in \mathscr{P}^1$ and $c_k(s,t) \in \mathbb{C}$ by*

$$b_k(s,t,u) = \pi_s(B_k^t)(u) \qquad \textit{and} \qquad c_k(s,t) = \pi_s(C_k^t). \tag{5.4}$$

*Note, by (5.1) and the linearity of $\pi_s$, that*

$$b_k(s,t,u) = e^{-\frac{k}{2}t} p_k^{s,t}(u). \tag{5.5}$$

It is useful to note the following alternative expression for $c_k(s,t)$. From (5.3),

$$C_k^t(\mathbf{v}) = e^{-\frac{k}{2}t}\mathcal{T}(e^{-\frac{t}{2}\mathcal{D}}u^k) = e^{-\frac{k}{2}t}e^{-\frac{t}{2}\mathcal{D}}v_k \tag{5.6}$$

since $\mathcal{T}$ commutes with $\mathcal{D}$. Thus, from Definition 1.24 and Remark 1.25,

$$c_k(s,t) = e^{-\frac{k}{2}t}\pi_s(e^{-\frac{t}{2}\mathcal{D}}v_k) = e^{-\frac{k}{2}t}\left(e^{\frac{1}{2}(s-t)\mathcal{D}}v_k\right)\Big|_{\mathbf{v}=\mathbf{1}} = e^{-\frac{k}{2}t}\nu_k(s-t). \tag{5.7}$$

The main computational tool that will lead to the identification of the Biane polynomials $p_k^{s,t}$ is the following recursion.

**Proposition 5.2.** *Let $s,t\in\mathbb{R}$, $u\in\mathbb{C}$, and $k\geq 1$. Let $c_k(s,t)$ and $b_k(s,t,u)$ be given as in Definiton 5.1. Then*

$$c_k(s,t) = \nu_k(s) + \sum_{m=1}^{k-1}\int_0^t mc_{k-m}(s,\tau)c_m(s,\tau)\,d\tau, \qquad k\geq 2 \tag{5.8}$$

*with $c_1(s,t)=\nu_1(s)$; and*

$$b_k(s,t,u) = u^k + \sum_{m=1}^{k-1}\int_0^t mc_{k-m}(s,\tau)b_m(s,\tau,u)\,d\tau, \qquad k\geq 2 \tag{5.9}$$

*with $b_1(s,t,u)=u$.*

*Proof.* First note that $B_k^0(u;\mathbf{v}) = u^k$ and $C_k^0(u;\mathbf{v}) = v_k$ by definition, and thus $b_k(s,0,u) = \pi_s(u^k) = u^k$, while $c_k(s,0) = \pi_s(v_k) = \nu_k(s)$ by (1.33). For $k=1$, we have

$$B_1^t(u) = e^{-\frac{t}{2}}e^{-\frac{t}{2}\mathcal{D}}u = u$$

because $\mathcal{D}u=-u$. For $k\geq 2$,

$$\frac{d}{dt}B_k^t = \frac{d}{dt}e^{-\frac{t}{2}(k+\mathcal{D})}u^k = -\frac{1}{2}e^{-\frac{t}{2}(k+\mathcal{D})}(k+\mathcal{D})u^k.$$

Recall (1.28) that $\mathcal{D} = -\mathcal{N}-2\mathcal{Z}-2\mathcal{Y}$. Eq. (1.24) shows that $\mathcal{N}(u^k)=ku^k$; (1.26) shows that $\mathcal{Z}$ annihilates $u^k$; and Example 1.18 works out that $\mathcal{Y}(u^k) = \sum_{j=1}^{k-1}(k-j)v_ju^{k-j} = \sum_{m=1}^{k-1}mu^mv_{k-m}$. Thus

$$(k+\mathcal{D})(u^k) = ku^k - ku^k - 2\sum_{m=1}^{k-1}mu^mv_{k-m} = -2\sum_{m=1}^{k-1}mu^mv_{k-m}.$$

Hence

$$\frac{d}{dt}B_k^t = e^{-\frac{k}{2}t}e^{-\frac{t}{2}\mathcal{D}}\left(\sum_{m=1}^{k-1}mu^mv_{k-m}\right) = e^{-\frac{k}{2}t}\sum_{m=1}^{k-1}me^{-\frac{t}{2}\mathcal{D}}(u^mv_{k-m}). \tag{5.10}$$

We now use the partial homomorphism property of (1.32) at time $-t$, which yields (since $v_{k-m}\in\mathscr{P}^0$) that

$$e^{-\frac{t}{2}\mathcal{D}}(u^mv_{k-m}) = (e^{-\frac{t}{2}\mathcal{D}}u^m)(e^{-\frac{t}{2}\mathcal{D}}v_{k-m}). \tag{5.11}$$

Now, $v_{k-m}=\mathcal{T}(u^{k-m})$, and by Lemma 1.19 $\mathcal{T}$ and $\mathcal{D}$ commute. We may rewrite (5.11) as

$$e^{-\frac{t}{2}\mathcal{D}}(u^mv_{k-m}) = (e^{-\frac{t}{2}\mathcal{D}}u^m)\mathcal{T}(e^{-\frac{t}{2}\mathcal{D}}u^{k-m}) \tag{5.12}$$

Eq. (5.3) gives

$$e^{-\frac{t}{2}\mathcal{D}}(\,\cdot\,)^m = e^{\frac{m}{2}t}B_m^t \qquad \text{and} \qquad \mathcal{T}[e^{-\frac{t}{2}\mathcal{D}}(\,\cdot\,)^{k-m}] = e^{\frac{k-m}{2}t}C_{k-m}^t.$$

Thus, (5.10) and (5.12) combine to give

$$\frac{d}{dt}B_k^t = e^{-\frac{k}{2}t}\sum_{m=1}^{k-1} m e^{\frac{m}{2}t}B_m^t e^{\frac{k-m}{2}t}C_{k-m}^t = \sum_{m=1}^{k-1} mC_{k-m}^t B_m^t. \tag{5.13}$$

Integrating both sides of (5.13) from 0 to $t$, and using the initial condition $B_k^t(u;\mathbf{v}) = u^k$, gives

$$B_k^t = u^k + \sum_{m=1}^{k-1} m\int_0^t C_{k-m}^\tau B_m^\tau\,d\tau. \tag{5.14}$$

The tracing map $\mathcal{T}$ is linear, and commutes with the integral (easily verified since all terms are polynomials); moreover, if $C \in \mathscr{P}^0$, then $\mathcal{T}(CB) = C\mathcal{T}(B)$. Thus

$$C_k^t = \mathcal{T}(B_k^t) = \mathcal{T}(u^k) + \sum_{m=1}^{k-1} m\int_0^t \mathcal{T}[C_{k-m}^\tau B_m^\tau]\,d\tau = v_k + \sum_{m=1}^{k-1} m\int_0^t C_{k-m}^\tau C_m^\tau\,d\tau. \tag{5.15}$$

Finally, the evaluation map $\pi_s$ is an algebra homomorphism, and (as with $\mathcal{T}$) commutes with the integral; applying $\pi_s$ to (5.14) and (5.15) yields the desired equations (5.8) and (5.9), concluding the proof. □

*Remark* 5.3. By changing the index $m \mapsto k-m$ in (5.8) and averaging the results, we may alternatively state the recursion for $c_k$ as

$$c_k(s,t) = \nu_k(s) + \frac{k}{2}\sum_{m=1}^{k-1}\int_0^t c_{k-m}(s,\tau)c_m(s,\tau)\,d\tau. \tag{5.16}$$

A transformation of this form is not possible for the $b_k(s,t,u)$ recursion, however.

## 5.2 Exponential Growth Bounds

In Section 5.3, we will study the generating functions of the quantities $\nu_k(s)$, $c_k(s,t)$, and $b_k(s,t,u)$. As such, we will need a priori exponential growth bounds.

**Lemma 5.4.** *For $s,t \in \mathbb{R}$ and $k \geq 2$,*

$$|\nu_k(t)| \leq C_{k-1}(1+|t|)^{k-1}e^{-\frac{k}{2}t}, \qquad \textit{and} \tag{5.17}$$

$$|c_k(s,t)| \leq C_{k-1}(1+|s-t|)^{k-1}e^{-\frac{k}{2}s}, \tag{5.18}$$

*where $C_k = \frac{1}{k+1}\binom{2k}{k}$ are the Catalan numbers.*

*Remark* 5.5. When $t > 0$, $\nu_k(t)$ is the $k$th moment of the probability measure $\nu_t$ on the unit circle $\mathbb{T}$, and we therefore have the much better bound $|\nu_k(t)| \leq 1$; similarly, if $s \geq t$, $|c_k(s,t)| \leq e^{-\frac{k}{2}t}$. It is necessary to have a priori bounds for negative $t$ and $s-t$ as well, however. While (5.17) is by no means sharp, the known exact formula (1.36) for $\nu_k(t)$ shows that, when $t < 0$, $\nu_k(t)$ does grown exponentially with $k$ (at least for small $|t|$).

In the proof of Lemma 5.4, we will use the well-known fact that the Catalan numbers satisfy Segner's recurrence relation

$$C_k = \sum_{m=1}^{k} C_{m-1}C_{k-m}, \qquad k \geq 1.$$

*Proof.* Taking $s = 0$ in (5.16), and noting that $\nu_k(0) = 1$ for all $k$, we have

$$c_k(0,t) = 1 + \frac{k}{2}\sum_{m=1}^{k-1}\int_0^t c_m(0,\tau)c_{k-m}(0,\tau)\,d\tau, \qquad k \geq 2. \tag{5.19}$$

We claim that

$$|c_k(0,t)| \leq C_{k-1}(1+|t|)^{k-1}, \qquad k \geq 1. \tag{5.20}$$

Since $c_1(0,t) = 1 = C_1$, we proceed by induction. Let $k \geq 2$, and assume that (5.20) holds below level $k$; then (5.19) yields

$$\begin{aligned}|c_k(0,t)| &\leq 1 + \frac{k}{2}\int_0^{|t|}\sum_{m=1}^{k-1} C_{m-1}C_{k-m-1}\,(1+\tau)^{k-2}\,d\tau \\ &= 1 + \frac{k}{2\,(k-1)}\left((1+|t|)^{k-1} - 1\right)\sum_{m=1}^{k-1} C_{m-1}C_{k-m-1} \\ &= 1 - \frac{k}{2\,(k-1)}C_{k-1} + (1+|t|)^{k-1}\,C_{k-1} \leq C_{k-1}\,(1+|t|)^{k-1}\end{aligned} \tag{5.21}$$

wherein we have used $\frac{k}{2(k-1)}C_{k-1} \geq 1$ for all $k \geq 2$. This completes the induction argument, proving (5.20) holds. Now, taking $s = 0$ in (5.7) yields

$$c_k(0,t) = e^{-\frac{k}{2}t}\nu_k(-t) \tag{5.22}$$

meaning that $\nu_k(t) = e^{-\frac{k}{2}t}c_k(0,-t)$, and this together with (5.20) proves (5.17). Then, using (5.7) once more, (5.17) implies that

$$|c_k(s,t)| = e^{-\frac{k}{2}t}|\nu_k(s-t)| \leq e^{-\frac{k}{2}t}e^{-\frac{k}{2}(s-t)}\cdot C_{k-1}(1+|s-t|)^{k-1}$$

which prove (5.18). □

*Remark* 5.6. Equations (5.19) and (5.22) together yield a recursion for the coefficients $\varrho_k(t) = e^{\frac{k}{2}t}\nu_k(t) = c_k(0,-t)$:

$$\varrho_k(t) = 1 - \frac{k}{2}\sum_{m=1}^{k-1}\int_0^t \varrho_m(\tau)\varrho_{k-m}(\tau)\,d\tau. \tag{5.23}$$

This same recursion was derived in [4, Lemma 11], using free stochastic calculus, with $\nu_k(s)$ being identified as the limit moments of the free unitary Brownian motion distribution. It is interesting that we can derive it directly from derivative formulas on the unitary group.

**Lemma 5.7.** *Let $s,t > 0$ and $u \in \mathbb{C}$. For $k \geq 2$, the $b_k(s,t,u)$ of (5.9) satisfy*

$$|b_k(s,t,u)| \leq [5(1+s)(1+t)]^{k-1}|u|^k. \tag{5.24}$$

*Proof.* Since $b_1(s,t,u) = u$, (5.24) holds for $k = 1$. We proceed by induction, assuming (5.24) holds below level $k$. Then (5.9) gives us

$$|b_k(s,t,u)| \leq |u|^k + \sum_{m=1}^{k-1}\int_0^t m|c_{k-m}(s,\tau)||b_m(s,\tau,u)|\,d\tau. \tag{5.25}$$

The Catalan number $C_k$ is $\le 4^k$ (in fact it is $\sim 4^k/k^{3/2}\sqrt{\pi}$). Note that, for $s,t>0$, $1+|s-t|\le(1+s)(1+t)$. Hence (5.18) implies that $|c_k(s,t)|\le[4(1+s)(1+t)]^{k-1}$. Thus (5.25) and the inductive hypothesis give us, for $k\ge 2$,

$$\begin{aligned}|b_k(s,t,u)| &\le |u|^k+\sum_{m=1}^{k-1}\int_0^t m[4(1+s)(1+\tau)]^{k-m-1}\cdot[5(1+s)(1+\tau)]^{m-1}|u|^k\,d\tau\\ &= |u|^k+|u|^k\cdot(1+s)^{k-2}\int_0^t(1+\tau)^{k-2}\,d\tau\cdot\sum_{m=1}^{k-1}m4^{k-m-1}5^{m-1}.\end{aligned}\tag{5.26}$$

Summing the geometric series, we may estimate

$$5^{k-1}-4^{k-1}\le\sum_{m=1}^{k-1}m4^{k-m-1}5^{m-1}\le(k-1)5^{k-1}.$$

Substituting this into (5.26) we have

$$\begin{aligned}|b_k(s,t,u)| &\le |u|^k+|u|^k(1+s)^{k-2}[(1+t)^{k-1}-1]\frac{1}{k-1}\sum_{m=1}^{k-1}m4^{k-m-1}5^{m-1}\\ &\le |u|^k\left(1-(1+s)^{k-2}\frac{5^{k-1}-4^{k-1}}{k-1}\right)+5^{k-1}(1+s)^{k-2}(1+t)^{k-1}|u|^k\\ &\le [5(1+s)(1+t)]^{k-1}|u|^k\end{aligned}$$

where we have used that $1+s\ge 1$ and $\frac{5^{k-1}-4^{k-1}}{k-1}\ge 1$ for $k\ge 2$. This concludes the inductive proof. □

### 5.3 Holomorphic PDE

The double recursion of Proposition 5.2 can be written in the form of coupled holomorphic PDEs for the generating functions of $c_k(s,t)$ and $b_k(s,t,u)$.

**Definition 5.8.** *Let $s,t\in\mathbb{R}$. For $z\in\mathbb{C}$, define*

$$\psi^s(t,z)=\sum_{k=1}^{\infty}c_k(s,t)z^k.$$

*Additionally, for $u\in\mathbb{C}$ define*

$$\phi^{s,u}(t,z)=\sum_{k=1}^{\infty}b_k(s,t,u)z^k.$$

By (5.18) and the Catalan bound $C_k\le 4^k$, the power series $z\mapsto\psi^s(t,z)$ is convergent whenever $|z|<e^{s/2}/4(1+|s-t|)$; similarly, by (5.24), the power series $z\mapsto\phi^{s,u}(t,z)$ is convergent whenever $s,t>0$ and $|z|<[5(1+s)(1+t)|u|]^{-1}$. Hence, $\psi^s(t,\cdot)$ and $\phi^{s,u}(t,\cdot)$ are holomorphic in a small disk (with radius that depends continuously on $s,t>0$). Note that, by (5.5),

$$\Pi(s,t,u,z)=\sum_{k\ge1}p_k^{s,t}(u)z^k=\sum_{k\ge1}e^{\frac{k}{2}t}b_k(s,t,u)z^k=\phi^{s,u}(t,e^{\frac{t}{2}}z).\tag{5.27}$$

So, identifying $\phi^{s,u}(t,z)$ will also identify the sought-after generating function $\Pi(s,t,u,z)$.

**Proposition 5.9.** *For fixed $s > 0$, the functions $\mathbb{R} \ni t \mapsto \psi^s(t,z)$ and $\mathbb{R}_+ \ni t \mapsto \phi^{s,u}(t,z)$ are differentiable for all sufficiently small $|z|$ and $|u|$. Their derivatives are given by*

$$\frac{\partial}{\partial t}\psi^s(t,z) = \sum_{k=1}^{\infty} \frac{\partial}{\partial t} c_k(s,t) z^k \qquad \textit{and} \qquad \frac{\partial}{\partial t}\phi^{s,u}(t,z) = \sum_{k=1}^{\infty} \frac{\partial}{\partial t} b_k(s,t,u) z^k.$$

*Proof.* From (5.8), $\frac{\partial}{\partial t}c_1(s,t) = 0$, while for $k \geq 2$ we have

$$\frac{\partial}{\partial t}c_k(s,t) = k \sum_{m=1}^{k-1} c_{k-m}(s,t) c_m(s,t).$$

Hence, from (5.18) and the Catalan bound $C_k \leq \frac{4^k}{k}$,

$$\left|\frac{\partial}{\partial t}c_k(s,t)\right| \leq \sum_{m=1}^{k-1} m|c_{k-m}(s,t)||c_m(s,t)| \leq (k-1)4^k e^{-\frac{k}{2}s}(1+|s-t|)^k$$

for $k \geq 2$. It follows that $\sum_{k=1}^{\infty} \frac{\partial}{\partial t}c_k(s,t)z^k$ converges to an analytic function of $z$ on the domain $|z| < e^{s/2}/4(1+|s-t|)$. Integrating this series term-by-term over the interval $[0,t]$ shows that it is the derivative of $\psi^s(t,z)$, as claimed. A completely analogous argument applies to $\phi^{s,u}(t,z)$. ☐

We will shortly write down coupled PDEs satisfies by $\psi^s$ and $\phi^{s,u}$. First, we remark on their initial conditions. From Proposition 5.2, we have

$$c_k(s,0) = \nu_k(s) \qquad \text{and} \qquad b_k(s,0,u) = u^k.$$

Thus

$$\psi^s(0,z) = \sum_{k\geq 1} \nu_k(s) z^k, \tag{5.28}$$

$$\phi^{s,u}(0,z) = \sum_{k\geq 1} u^k z^k = \frac{uz}{1-uz}. \tag{5.29}$$

It will be convenient to express $\psi^s(0,z)$ in terms of the shifted coefficients $\varrho_k(s) = e^{\frac{k}{2}s}\nu_k(s)$ considered in Remark 5.6. Define

$$\varrho(s,z) = \sum_{k\geq 1} \varrho_k(s) z^k = \psi^s(0, e^{\frac{s}{2}} z). \tag{5.30}$$

Note that, since $\nu_k(0) = 1$ for all $k$, $\varrho(0,z) = \frac{z}{1-z}$.

**Proposition 5.10.** *For $s,t > 0$ and $|z|$ and $|u|$ sufficiently small, the functions $\varrho$, $\psi^s$, and $\phi^{s,u}$ satisfy the following holomorphic PDEs:*

$$\frac{\partial \varrho}{\partial s} = -z\varrho \frac{\partial \varrho}{\partial z}, \qquad \varrho(0,z) = \frac{z}{1-z}, \tag{5.31}$$

$$\frac{\partial \psi^s}{\partial t} = z\psi^s \frac{\partial \psi^s}{\partial z}, \qquad \psi^s(0,z) = \varrho(s, e^{-\frac{s}{2}} z), \tag{5.32}$$

$$\frac{\partial \phi^{s,u}}{\partial t} = z\psi^s \frac{\partial \phi^{s,u}}{\partial z}, \qquad \phi^{s,u}(0,z) = \frac{uz}{1-uz}. \tag{5.33}$$

*Remark* 5.11. (1) PDE (5.31) was proved in [3, Lemma 1], using the recursion (5.23). We reprove it here, as a special case of (5.32).

(2) The requirement that $s,t>0$ is only needed to verify the exponential growth bounds of the coefficients $b_k(s,t,u)$; cf. Lemma 5.7. Lemma 5.4, on the other hand, is valid for all $s,t\in\mathbb{R}$, and so (5.31) and (5.32) are valid for $s,t\in\mathbb{R}$.

*Proof.* First, Remark 5.6 and (5.22) show that $\varrho_k(t)=c_k(0,-t)$, and hence $\varrho(t,z)=\psi^0(-t,z)$. Hence, (5.31) follows immediately from (5.32). Now, Proposition 5.9 yields that $\psi^s(t,z)$ is differentiable in $t$, and so by Proposition 5.2

$$\frac{\partial}{\partial t}\psi^s(t,z)=\sum_{k=2}^{\infty}\frac{\partial}{\partial t}c_k(s,t)\,z^k=\sum_{k=2}^{\infty}\sum_{m=1}^{k-1}mc_m(s,t)c_{k-m}(s,t)\,z^k. \tag{5.34}$$

On the other hand, $\psi^s(t,z)$ is analytic in $z$, and

$$z\frac{\partial}{\partial z}\psi^s(t,z)=\sum_{k=1}^{\infty}c_k(s,t)\cdot z\frac{\partial}{\partial z}z^k=\sum_{k=1}^{\infty}kc_k(s,t)z^k,$$

and so

$$\begin{aligned}z\psi^s(t,z)\frac{\partial}{\partial z}\psi^s(t,z)&=\sum_{k_1=1}^{\infty}c_{k_1}(s,t)z^{k_1}\cdot\sum_{k_2=1}^{\infty}k_2c_{k_2}(s,t)z^{k_2}\\&=\sum_{k=2}^{\infty}z^k\sum_{\substack{k_1+k_2=k\\k_1,k_2\ge1}}k_2c_{k_1}(s,t)c_{k_2}(s,t).\end{aligned}$$

Reindexing the internal sum and comparing with (5.34) proves (5.32). The proof of (5.33) is entirely analogous. □

### 5.4 Generating Function

We now proceed to prove the implicit formula (1.39), by solving the coupled PDEs (5.31)–(5.33). We do this essentially by the method of characteristics. These quasilinear PDEs have a fairly simple form; as a result, the characteristic curves are the same as the level curves in this case. As we will see, all three equations have the same level curves.

**Lemma 5.12.** *Fix $s_0\ge0$ and $w_0\in\mathbb{C}$ with $|w_0|<[4(1+s_0)]^{-1}$. Consider the exponential curve*

$$\mathbf{w}(s)=w_0\,e^{\varrho(0,w_0)s}.$$

*Then $s\mapsto\varrho(s,\mathbf{w}(s))$ is constant. In particular, $\varrho(s,\mathbf{w}(s))=\varrho(0,w_0)$ for all $s\in[0,s_0)$.*

*Proof.* Lemma 5.4 shows that $e^{\frac{k}{2}s}|\nu_k(s)|\le[4(1+s)]^k$; thus $\varrho(s,w)=\psi_{\nu_s}(e^{\frac{s}{2}}w)=\sum_{k\ge1}e^{\frac{k}{2}s}\nu_k(s)w^k$ converges to an analytic function of $w$ for $|w|<[4(1+s)]^{-1}$. Thus, since $s\mapsto[4(1+s)]^{-1}$ is decreasing, $\varrho(s,w)$ is differentiable in $s$ and analytic in $w$ for $|w|<[4(1+s_0)]^{-1}$ and $0\le s<s_0$. Since $4(1+s_0)>1$, the initial condition $\varrho(0,w)=\frac{w}{1-w}$ is also analytic on this domain. Thus, subject to these constraints, we can simply differentiate. To avoid confusion, we denote $\dot\varrho(s,w)=\frac{\partial\varrho}{\partial s}(s,w)$ and $\varrho'(s,w)=\frac{\partial\varrho}{\partial w}(s,w)$.

$$\frac{d}{ds}\varrho(s,\mathbf{w}(s))=\dot\varrho(s,\mathbf{w}(s))+\varrho'(s,\mathbf{w}(s))\dot{\mathbf{w}}(s). \tag{5.35}$$

We now use (5.31), which asserts that $\dot\varrho(s,w)=-w\varrho(s,w)\varrho'(s,w)$; hence

$$\dot\varrho(s,\mathbf{w}(s))=-\mathbf{w}(s)\varrho(s,\mathbf{w}(s))\,\varrho'(s,\mathbf{w}(s)).$$

Plugging this into (5.35) yields

$$\frac{d}{ds}\varrho(s,\mathbf{w}(s)) = \varrho'(s,\mathbf{w}(s))\left[-\mathbf{w}(s)\varrho(s,\mathbf{w}(s)) + \dot{\mathbf{w}}(s)\right]. \tag{5.36}$$

Note that $\mathbf{w}$ satisfies the ODE

$$\dot{\mathbf{w}}(s) = \frac{d}{ds}w_0\, e^{\varrho(0,w_0)s} = \varrho(0,w_0)w_0\, e^{\varrho(0,w_0)s} = \varrho(0,w_0)\mathbf{w}(s).$$

Substituting this into (5.35) yields

$$\begin{aligned}&\frac{d}{ds}\varrho(s,\mathbf{w}(s)) = \varrho'(s,\mathbf{w}(s))\mathbf{w}(s)\left[\varrho(0,w_0) - \varrho(s,\mathbf{w}(s))\right] \\ &\varrho\left(s,\mathbf{w}(s)\right)|_{s=0} = \varrho(0,w_0).\end{aligned} \tag{5.37}$$

We now easily see that $\varrho\left(s,\mathbf{w}(s)\right) \equiv \varrho(0,w_0) = \frac{w_0}{1-w_0}$ is indeed the (unique) solution to this this ODE. □

**Corollary 5.13.** *Subject to the constraints on $s, w$ in Lemma 5.12, the function $\psi^s(0,w) = \varrho(s, e^{-\frac{s}{2}}w)$ is constant along the curves $s \mapsto e^{\frac{s}{2}}\mathbf{w}(s) = w_0 e^{[\varrho(0,w_0)+\frac{1}{2}]s}$. Note that*

$$\varrho(0,w_0) + \tfrac{1}{2} = \tfrac{w_0}{1-w_0} + \tfrac{1}{2} = \tfrac{1}{2}\tfrac{1+w_0}{1-w_0}.$$

*Thus, for all sufficiently small $w$ and $s$,*

$$\psi^s(0, w\, e^{\frac{s}{2}\frac{1+w}{1-w}}) = \upsilon(0,w) = \varrho(0,w) = \frac{w}{1-w}. \tag{5.38}$$

Differentiation shows that the function $w \mapsto we^{\frac{s}{2}\frac{1+w}{1-w}}$ is strictly increasing for all $w \in \mathbb{R}$ (provided $s < 4$); and in general for all $w > 0$ for all $s$; hence, (5.38) actually uniquely determines $\psi^s(0,z)$ for $z$ (by analytic continuation) when $s < 4$; moreoever, by the inverse function theorem, it is analytic in $z$.

Following the idea of Lemma 5.12, we now show that the level-curves of the functions $\psi^s$ and $\phi^{s,u}$ are also exponentials.

**Lemma 5.14.** *For $z_0 \in \mathbb{C}$, consider the exponential curve*

$$\mathbf{z}(t) = z_0\, e^{-\psi^s(0,z_0)t}.$$

*Then for $z_0$ and $t$ sufficiently small, $t \mapsto \psi^s(t,\mathbf{z}(t))$ and $t \mapsto \phi^{s,u}(t,\mathbf{z}(t))$ are constant. In particular,*

$$\psi^s(t,\mathbf{z}(t)) = \psi^s(0,z_0), \quad \textit{and} \quad \phi^{s,u}(t,\mathbf{z}(t)) = \phi^{s,u}(0,z_0).$$

*Proof.* To improve readability, through this proof we suppress the parameters $s, u$ and simply write $\phi^{s,u}(t,z) = \phi(t,z)$ and $\psi^s(t,z) = \psi(t,z)$. As per the discussion following Definition 5.8, these functions are differentiable in $t$ and analytic in $z$ for sufficiently small $z$. As in the proof of Lemma 5.12, we set $\dot{\psi}(t,z) = \frac{\partial}{\partial t}\psi(t,z)$, and $\psi'(t,z) = \frac{\partial}{\partial z}\psi(t,z)$, and similarly with $\dot{\phi}$ and $\phi'$. Differentiating, we have

$$\begin{aligned}\frac{d}{dt}\psi(t,\mathbf{z}(t)) &= \dot{\psi}(t,\mathbf{z}(t)) + \psi'(t,\mathbf{z}(t))\dot{\mathbf{z}}(t) \\ \frac{d}{dt}\phi(t,\mathbf{z}(t)) &= \dot{\phi}(t,\mathbf{z}(t)) + \phi'(t,\mathbf{z}(t))\dot{\mathbf{z}}(t).\end{aligned}$$

PDEs (5.32) and (5.33) say $\dot{\psi}(t,z) = z\psi(t,z)\psi'(t,z)$ and $\dot{\phi}(t,z) = z\psi(t,z)\psi'(t,z)$, and so

$$\frac{d}{dt}\psi(t,\mathbf{z}(t)) = [\mathbf{z}(t)\psi(t,\mathbf{z}(t)) + \dot{\mathbf{z}}(t)]\,\psi'(t,\mathbf{z}(t)) \tag{5.39}$$

$$\frac{d}{dt}\phi(t,\mathbf{z}(t)) = [\mathbf{z}(t)\psi(t,\mathbf{z}(t)) + \dot{\mathbf{z}}(t)]\,\phi'(t,\mathbf{z}(t)) \tag{5.40}$$

As in the proof of Lemma 5.12, we note that $\mathbf{z}$ satisfies the ODE

$$\dot{\mathbf{z}}(t) - z_0\psi(0,z_0)e^{-\psi(0,z_0)t} = -\psi(0,z_0)\mathbf{z}(t).$$

Substituting this into (5.39) and (5.40) yields

$$\frac{d}{dt}\psi(t,\mathbf{z}(t)) = [\psi(t,\mathbf{z}(t)) - \psi(0,z_0)]\,\mathbf{z}(t)\psi'(t,\mathbf{z}(t)) \tag{5.41}$$

$$\frac{d}{dt}\phi(t,\mathbf{z}(t)) = [\psi(t,\mathbf{z}(t)) - \psi(0,z_0)]\,\mathbf{z}(t)\phi'(t,\mathbf{z}(t)). \tag{5.42}$$

The initial condition for (5.41) is $\psi(t,\mathbf{z}(t))|_{t=0} = \psi(0,z_0)$, and it follows immediately that $\psi(t,\mathbf{z}(t)) = \psi(0,z_0)$ is the unique solution of this ODE. Hence, (5.42) reduces to the equation $\frac{d}{dt}\phi(t,\mathbf{z}(t)) = 0$, and since its initial condition is $\phi(t,\mathbf{z}(t))|_{t=0} = \phi(0,z_0)$, it follows that $\phi(t,\mathbf{z}(t)) = \phi(0,z_0)$ as well. □

This brings us to the proof of (1.39). First, Lemma 5.14, together with the initial condition in (5.33), yields

$$\phi^{s,u}(t, ze^{-\psi^s(0,z)t}) = \phi^{s,u}(0,z) = \frac{uz}{1-uz} = \frac{1}{1-uz} - 1. \tag{5.43}$$

Next, Corollary 5.13 describes $(s,z) \mapsto \psi^s(0,z)$ in terms of its level curves; (5.38) states that

$$\psi^s(0, w\,e^{\frac{s}{2}\frac{1+w}{1-w}}) = \varrho(0,w) = \frac{w}{1-w}. \tag{5.44}$$

So set $z = we^{\frac{s}{2}\frac{1+w}{1-w}}$; then (5.43) and (5.44) say

$$\phi^{s,u}(t, e^{-\frac{w}{1-w}t}we^{\frac{s}{2}\frac{1+w}{1-w}}) = \phi^{s,u}(t, e^{-\psi^s(0,z)t}z) = \left(1 - uwe^{\frac{s}{2}\frac{1+w}{1-w}}\right)^{-1} - 1. \tag{5.45}$$

Finally, note that

$$-\tfrac{w}{1-w} = -\tfrac{1}{2}\tfrac{1+w}{1-w} + \tfrac{1}{2}$$

and so (5.45) may be written in the form

$$\phi^{s,u}(t, e^{\frac{t}{2}}we^{\frac{1}{2}(s-t)\frac{1+w}{1-w}}) = \left(1 - uwe^{\frac{s}{2}\frac{1+w}{1-w}}\right)^{-1} - 1. \tag{5.46}$$

Finally, recall (5.27), which (in this language) says that

$$\Pi(s,t,u,\zeta) = \phi^{s,u}(t, e^{\frac{t}{2}}\zeta). \tag{5.47}$$

Setting $\zeta = we^{\frac{1}{2}(s-t)\frac{1+w}{1-w}}$, (5.46) and (5.47) combine to yield

$$\left(1 - uwe^{\frac{s}{2}\frac{1+w}{1-w}}\right)^{-1} - 1 = \phi^{s,u}(t, e^{\frac{t}{2}}we^{\frac{1}{2}(s-t)\frac{1+w}{1-w}}) = \phi^{s,u}(t, e^{\frac{t}{2}}\zeta) = \Pi(s,t,u,\zeta)$$

which is precisely the statement of (1.39).

### 5.5 Proof of Theorem 1.31 ($\mathrm{B}_{t,t} = \mathscr{G}_t$)

We are now in a position to complete the proof of Theorem 1.31, modulo a small error in [4].

*Remark* 5.15. In [4, Lemma 18], there is a typographical error that is propagated through the remainder of that paper. In the second line of the proof of that lemma, the function $\iota(t,\cdot)$ should be the inverse of $z \mapsto ze^{\frac{t}{2}\frac{1+z}{1-z}}$ rather than the inverse of $z \mapsto \frac{z}{1+z}e^{\frac{t}{2}(1+2z)}$ as stated. That $\iota(t,\cdot)$ has this different form follows from [4, Lemma 11], which defines the kernel function $\kappa(t,z)$ (formula 4.2.2.a) implicitly by $\frac{\kappa(t,z)-1}{\kappa(t,z)+1}e^{\frac{t}{2}\kappa(t,z)} = z$; then $\iota(t,z) = \frac{\kappa(t,1/z)+1}{\kappa(t,1/z)-1}$ yields the result. Hence, the correct generating function for the Biane polynomials in [4] is the one in (1.39). The presence of this error, and the tracking of its source, were confirmed by Philippe Biane in a private communication on October 27, 2011.

*Proof of Theorem 1.31.* By the density of trigonometric polynomials in $L^2(\mathbb{T},\nu_t)$ for any measure $\nu_t$, the transform $\mathscr{G}^t$ is determined by its action on Laurent polynomial functions. Hence, to verify that $\mathbf{B}_{t,t} = \mathscr{G}^t$, it suffices to verify that $(\mathscr{G}^t)^{-1}$ agrees with $\mathbf{H}_{t,t}$ on monomials $z \mapsto z^k$ for $k \in \mathbb{Z}$. Eq. (5.2) is consistent with [4, Lemma 18], and so it suffices to prove this result for $k \geq 1$. Eq. (1.39) verifies that the Biane polynomials $p_k^{t,t}$ for $\mathbf{H}_{t,t}$ have the same generating function as the Biane polynomials of $\mathscr{G}^t$ (cf. Remark 5.15), and this concludes the proof. □

## A Heat Kernel Measures on Lie Groups

Suppose that $G$ is a connected Lie group and $\beta$ is a basis for $\mathrm{Lie}\,(G)$. Then $A = \sum_{X\in\beta}\partial_X^2$ is a left-invariant non-positive elliptic differential operator which is essentially self adjoint on $C_c^\infty(G)$ as an operator on $L^2(G,dg)$ where $dg$ is a right Haar measure on $G$. Associated to the contraction semigroup $\left\{e^{tA/2}\right\}_{t>0}$ is a convolution semigroup of probability (heat kernel) densities $\{h_t\}_{t>0}$. In more detail, $\mathbb{R}_+ \times G \ni (t,g) \to h_t(g) \in \mathbb{R}_+$ is a smooth function such that

$$\partial_t h_t(g) = \frac{1}{2}Ah_t(g) \text{ for } t > 0$$

and

$$\lim_{t\downarrow 0}\int_G f(g)h_t(g)\,dg = f(e) \text{ for all } f \in C_c(G).$$

(Throughout, $e = 1_G$.) Basic properties of these heat kernels are summarized in [8, Proposition 3.1] and [9, Section 3.]. For an exhaustive treatment of heat kernels on Lie groups see [24] and [31]. For our present purposes, we need to know that, if $G = U(N)$ or $G = GL(N,\mathbb{C})$ (and so $h_t$ is the density of $\rho_t$ or $\mu_{s,t}$, respectively), then

$$\int_G f(g)h_t(g)\,dg = \sum_{n=0}^{\infty}\frac{1}{n!}\left(\frac{t}{2}\right)^n (A^n f)(I) \text{ for all } t \geq 0 \tag{A.1}$$

whenever $f$ is a trace Laurent polynomial. This result can be seen as a consequence of Langland's theorem; see, for example, [24, Theorem 2.1 (p. 152)]. As it is a bit heavy to get to Langland's theorem in Robinson we will, for the reader's convenience, sketch a proof of (A.1); see Theorem A.2 below. For the rest of this section let $d$ denote the left-invariant metric on $G$ such that $\{\partial_X\}_{X\in\beta}$ is an orthonormal frame on $G$ and set $|g| = d(e,g)$. Also let us use the abbreviation $h_t(f)$ for $\int_G f(g)h_t(g)\,dg$.

**Lemma A.1.** *Suppose $f\colon [0,T]\times G \to \mathbb{C}$ is a $C^2$ function such that $|h(t,g)| \leq Ce^{C|g|}$ for some $C < \infty$, where $h$ is any of the functions $f$, $\partial_t f$, or $\partial_X f$ for any $X \in \mathrm{Lie}(A)$, or $Af$. Then*

$$\partial_t h_t\left(f(t,\cdot)\right) = h_t\left(\partial_t f(t,\cdot) + \frac{1}{2}Af(t,\cdot)\right) \textit{ for } t \in (0,T] \tag{A.2}$$

*and*

$$\lim_{t\downarrow 0} h_t\left(f(t,\cdot)\right) = f(0,\cdot). \tag{A.3}$$

*Proof.* Let $\{h_n\} \subset C_c^\infty(G,[0,1])$ be smooth cutoff functions as in [8, Lemma 3.6] and set $f_n(t,g) \equiv h_n(g)f(t,g)$. Then it is easy to verify that it is now permissible to differentiate past the integrals and perform the required integration by parts in order to show that

$$\frac{d}{dt}\left[h_t(f_n(t,\cdot))\right] = h_t\left(\partial_t f(t,\cdot) + \frac{1}{2}Af(t,\cdot)\right).$$

Let $F(t,\cdot) = \partial_t f(t,\cdot) + \frac{1}{2}Af(t,\cdot)$ and

$$\begin{aligned} F_n(t,\cdot) &= \partial_t f_n(t,\cdot) + \frac{1}{2}Af_n(t,\cdot) \\ &= F(t,\cdot)h_n + \frac{1}{2}f(t,\cdot)Ah_n + \sum_{X\in\beta} \partial_X f(t,\cdot)\partial_X h_n. \end{aligned}$$

From the properties of $h_n$ and the assumed bounds on $f$, given $\epsilon \in (0,T)$ there exist $C < \infty$ independent of $n$ such that

$$\sup_{\epsilon\le t\le T} |F_n(t,g) - F(t,g)| \le \mathbb{1}_{|g|\ge n} Ce^{C|g|}.$$

It then follows by the standard heat kernel bounds (see for example [31] or [24, page 286]) that

$$\sup_{\epsilon\le t\le T} \left|h_t\left(F_n(t,\cdot)\right) - h_t\left(F(t,\cdot)\right)\right| \to 0 \text{ as } n\to\infty.$$

Hence we may conclude that $\frac{d}{dt}\left[h_t(f(t,\cdot))\right]$ exists and

$$\begin{aligned} \frac{d}{dt}\left[h_t\left(f(t,\cdot)\right)\right] &= \lim_{n\to\infty} \frac{d}{dt}\left[h_t\left(f_n(t,\cdot)\right)\right] \\ &= h_t\left(\partial_t f_n(t,\cdot) + \frac{1}{2}Af_n(t,\cdot)\right) \text{ for } \varepsilon < t \le T \end{aligned}$$

which proves (A.2). To prove (A.3) we start with the estimate

$$\begin{aligned} \left|h_t\left(f(t,\cdot)\right) - f(0,e)\right| &= \left|\int_G \left[f(t,y) - f(0,e)\right] h_t(y)\,dy\right| \\ &\le \int_G \left|f(t,y) - f(0,e)\right| h_t(y)\,dy \\ &\le \delta(\epsilon,t) + C\int_{|y|>\epsilon} e^{C|y|}h_t(y)\,dy \end{aligned}$$

where

$$\delta(\epsilon,t) = \int_{|y|\le\varepsilon} \left|f(t,y) - f(0,e)\right| h_t(y)\,dy \le \sup_{|y|\le\epsilon} \left|f(t,y) - f(0,e)\right|.$$

From [8, Lemma 4.3] modified in a trivial way from its original form where $\epsilon$ was take to be 1, we know that

$$\limsup_{t\downarrow 0} \int_{|y|>\epsilon} e^{c|y|}h_t(y)\,dy = 0 \text{ for all } \epsilon > 0 \text{ and } c < \infty.$$

Therefore, we conclude that

$$\limsup_{t\downarrow 0} \left|h_t\left(f(t,\cdot)\right) - f(0,e)\right| \le \limsup_{t\downarrow 0} \delta(\epsilon,t) \to 0 \text{ as } \epsilon\downarrow 0$$

as claimed. □

**Theorem A.2.** *Suppose now that $G = U(N)$ or $G = GL(N, \mathbb{C})$ and $P_N$ is a trace Laurent polynomial function on $G$. Then for $T > 0$,*

$$h_T(P_N) = \left( \sum_{n=0}^{\infty} \frac{1}{n!} \left( \frac{T}{2} \right)^n A^n P_N \right)(I). \tag{A.4}$$

*Proof.* Fix $T > 0$, and for $0 < t < T$ let

$$f(t, \cdot) = \sum_{n=0}^{\infty} \frac{1}{n!} \left( \frac{T-t}{2} \right)^n A^n P_N$$

where the sum is convergent as $A$ is a bounded operator on the finite dimensional subspace of trace Laurent polynomials of trace degree $\deg P$ or less. Moreover, $f(t, \cdot)$ is again a trace Laurent polynomial with time dependent coefficients and $f$ satisfies

$$\partial_t f(t, \cdot) + \frac{1}{2} A f(t, \cdot) = 0 \text{ with } f(T, \cdot) = P_N.$$

From Lemma A.1 we may now conclude,

$$\frac{d}{dt} \left[ h_t \left( f(t, \cdot) \right) \right] = h_t \left( \partial_t f(t, \cdot) + \frac{1}{2} A f(t, \cdot) \right) = 0.$$

Therefore $t \to h_t\left(f(t, \cdot)\right)$ is constant for $t > 0$ and hence, using Lemma A.1 again,

$$h_T(P_N) = h_T\left(f(T, \cdot)\right) = \lim_{t \downarrow 0} h_t\left(f(t, \cdot)\right) = f(0, I) = \left( \sum_{n=0}^{\infty} \frac{1}{n!} \left( \frac{T}{2} \right)^n A^n P_N \right)(I).$$

□


## Acknowledgments

The authors wish to thank Matt Dyer and Brendon Rhoades, who provided very useful algebraic insights to the authors for the proofs of Propositions 2.7 and 2.10.


# References


[1] BARGMANN, V. On a Hilbert space of analytic functions and an associated integral transform. *Comm. Pure Appl. Math. 14* (1961), 187–214.

[2] BARGMANN, V. Remarks on a Hilbert space of analytic functions. *Proc. Nat. Acad. Sci. U.S.A. 48* (1962), 199–204.

[3] 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.

[4] 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.

[5] BIANE, P., AND SPEICHER, R. Stochastic calculus with respect to free Brownian motion and analysis on Wigner space. *Probab. Theory Related Fields 112*, 3 (1998), 373–409.

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

[7] DRIVER, B. K. On the Kakutani-Itô-Segal-Gross and Segal-Bargmann-Hall isomorphisms. *J. Funct. Anal. 133*, 1 (1995), 69–128.

[8] DRIVER, B. K., AND GROSS, L. Hilbert spaces of holomorphic functions on complex Lie groups. In *New trends in stochastic analysis (Charingworth, 1994)*. World Sci. Publ., River Edge, NJ, 1997, pp. 76–106.

[9] DRIVER, B. K., GROSS, L., AND SALOFF-COSTE, L. Holomorphic functions and subelliptic heat kernels over Lie groups. *J. Eur. Math. Soc. (JEMS) 11*, 5 (2009), 941–978.

[10] DRIVER, B. K., AND HALL, B. C. Yang-Mills theory and the Segal-Bargmann transform. *Comm. Math. Phys. 201*, 2 (1999), 249–290.

[11] DUISTERMAAT, J. J., AND KOLK, J. A. C. *Lie groups*. Universitext. Springer-Verlag, Berlin, 2000.

[12] GORDINA, M. Heat kernel analysis and Cameron-Martin subgroup for infinite dimensional groups. *J. Funct. Anal. 171*, 1 (2000), 192–232.

[13] GORDINA, M. Holomorphic functions and the heat kernel measure on an infinite-dimensional complex orthogonal group. *Potential Anal. 12*, 4 (2000), 325–357.

[14] GROSS, L. Uniqueness of ground states for Schrödinger operators over loop groups. *J. Funct. Anal. 112*, 2 (1993), 373–441.

[15] GROSS, L., AND MALLIAVIN, P. Hall's transform and the Segal-Bargmann map. In *Itô's stochastic calculus and probability theory*. Springer, Tokyo, 1996, pp. 73–116.

[16] GUILLAUME, C. Free convolution operators and free Hall transform. *Preprint* (April 2013). arXiv:1304.1713.

[17] HALL, B. C. The Segal-Bargmann "coherent state" transform for compact Lie groups. *J. Funct. Anal. 122*, 1 (1994), 103–151.

[18] 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.

[19] HALL, B. C. *Lie groups, Lie algebras, and representations*, vol. 222 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2003. An elementary introduction.

[20] HALL, B. C., AND SENGUPTA, A. N. The Segal-Bargmann transform for path-groups. *J. Funct. Anal. 152*, 1 (1998), 220–254.

[21] KEMP, T., NOURDIN, I., PECCATI, G., AND SPEICHER, R. Wigner chaos and the fourth moment. *Ann. Probab. 40*, 4 (2012), 1577–1635.

[22] LAGERSTROM, P. A proof of a theorem on commutative matrices. *Bull. Amer. Math. Soc. 51* (1945), 535–536.

[23] LÉVY, T. Schur-Weyl duality and the heat kernel measure on the unitary group. *Adv. Math. 218*, 2 (2008), 537–575.

[24] ROBINSON, D. W. *Elliptic operators and Lie groups*. Oxford Mathematical Monographs. The Clarendon Press Oxford University Press, New York, 1991. Oxford Science Publications.

[25] RUDIN, W. *Real and complex analysis*, third ed. McGraw-Hill Book Co., New York, 1987.

[26] SAGAN, B. E. *The symmetric group*, second ed., vol. 203 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2001. Representations, combinatorial algorithms, and symmetric functions.

[27] SATTINGER, D. H., AND WEAVER, O. L. *Lie groups and algebras with applications to physics, geometry, and mechanics*, vol. 61 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1993. Corrected reprint of the 1986 original.

[28] SEGAL, I. E. Mathematical characterization of the physical vacuum for a linear Bose-Einstein field. (Foundations of the dynamics of infinite systems. III). *Illinois J. Math. 6* (1962), 500–523.

[29] SEGAL, I. E. *Mathematical problems of relativistic physics*, vol. 1960 of *With an appendix by George W. Mackey. Lectures in Applied Mathematics (proceedings of the Summer Seminar, Boulder, Colorado*. American Mathematical Society, Providence, R.I., 1963.

[30] SEGAL, I. E. The complex-wave representation of the free boson field. In *Topics in functional analysis (essays dedicated to M. G. Krein on the occasion of his 70th birthday)*, vol. 3 of *Adv. in Math. Suppl. Stud.* Academic Press, New York, 1978, pp. 321–343.

[31] VAROPOULOS, N. T., SALOFF-COSTE, L., AND COULHON, T. *Analysis and geometry on groups*, vol. 100 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1992.