The Spectral Edge of Unitary Brownian Motion

Benoît Collins*
Kyoto University
Department of Mathematics
Kyoto 606-8502, Japan & CNRS
collins@math.kyoto-u.ac.jp

Antoine Dahlqvist†
Statistical Laboratory
Centre for Mathematical Sciences
Wilberforce Road, Cambridge, CB3 0WA
ad814@maths.cam.ac.uk

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

December 16, 2016

Abstract

The Brownian motion $(U_t^N)_{t\geq 0}$ on the unitary group converges, as a process, to the free unitary Brownian motion $(u_t)_{t\geq 0}$ as $N\to\infty$. In this paper, we prove that it converges *strongly* as a process: not only in distribution but also in operator norm. In particular, for a fixed time $t>0$, we prove that the unitary Brownian motion has a *spectral edge*: there are no outlier eigenvalues in the limit. We also prove an extension theorem: any strongly convergent collection of random matrix ensembles independent from a unitary Brownian motion also converge strongly *jointly* with the Brownian motion. We give an application of this strong convergence to the Jacobi process.

Contents

*Supported in part by Kakenhi, an NSERC discovery grant, Ontario's ERA and ANR-14-CE25-0003
†Supported in part by RTG 1845 and EPSRC grant EP/I03372X/1.
‡Supported in part by NSF CAREER Award DMS-1254807

1 Introduction

This paper is concerned with convergence of the noncommutative distribution of the standard Brownian motion on unitary groups. Let $\mathbb{M}_N$ denote the space of $N\times N$ complex matrices, and let $\mathrm{Tr}(A) = \sum_{j=1}^N A_{jj}$ denote the (usual) trace. It will be convenient throughout to use the *normalized* trace, and so we use the symbol $\mathrm{tr} = \frac{1}{N}\mathrm{Tr}$ (with a lower-case t) for this purpose. We denote the unitary group in $\mathbb{M}_N$ as $\mathbb{U}_N$. The Brownian motion on $\mathbb{U}_N$ is the diffusion process $(U_t^N)_{t\geq 0}$ started at the identity with infinitesimal generator $\frac{1}{2}\Delta_{\mathbb{U}_N}$, where $\Delta_{\mathbb{U}_N}$ is the left-invariant Laplacian on $\mathbb{U}_N$. (This is uniquely defined up to a choice of N-dependent scale; see Section 2.1 for precise definitions, notation, and discussion.)

For each fixed $t\geq 0$, U_t^N is a random unitary matrix, whose spectrum $\mathrm{spec}(U_t^N)$ consists of N eigenvalues $\lambda_1(U_t^N),\ldots,\lambda_N(U_t^N)$. The *empirical spectral distribution*, also known as the *empirical law of eigenvalues*, of U_t^N (for a fixed $t\geq 0$) is the random probability measure $\mathrm{Law}_{U_t^N}$ on the unit circle $\mathbb{U}_1$ that puts equal mass on each eigenvalue (counted according to multiplicity):

$$\mathrm{Law}_{U_t^N} = \frac{1}{N}\sum_{j=1}^N \delta_{\lambda_j(U_t^N)}.$$

In other words: $\mathrm{Law}_{\mathbb{U}_t^N}$ is the random measure determined by the characterization that its integral against a test function $f\in C(\mathbb{U}_1)$ is given by

$$\int_{\mathbb{U}_1} f\, d\mathrm{Law}_{U_t^N} = \frac{1}{N}\sum_{j=1}^N f(\lambda_j(U_t^N)). \tag{1.1}$$

It is customary to realize all of the processes $\{U_t^N\colon N\in\mathbb{N}\}$ on a single probability space in order to talk about almost sure convergence of $\mathrm{Law}_{U_t^N}$ as $N\to\infty$. The standard realization is to declare that U_t^N and U_s^M are independent for all $s,t\geq 0$ and all $N\neq M$. (To be clear, though, none of the results stated below depend on this particular realization, and indeed hold for any coupling.)

In [4], Biane showed that the random measure $\mathrm{Law}_{U_t^N}$ converges weakly almost surely to a deterministic limit probability measure ν_t,

$$\lim_{N\to\infty}\int_{\mathbb{U}_1} f\, d\mathrm{Law}_{U_t^N} = \int_{\mathbb{U}_1} f\, d\nu_t \;\; a.s. \qquad f\in C(\mathbb{U}_1). \tag{1.2}$$

The measure ν_t can be described as the spectral measure of a *free unitary Brownian motion* (cf. Section 2.3). For $t>0$, ν_t possesses a continuous density that is symmetric about $1\in\mathbb{U}_1$, and is supported on an arc strictly contained in the circle for $0<t<4$; for $t\geq 4$, $\mathrm{supp}\,\nu_t = \mathbb{U}_1$.

The result of (1.2) is a bulk result: it does not constrain the behavior of eigenvalues near the edge. The additive counterpart is the classical Wigner's semicircle law. Let X^N be a Gaussian unitary ensemble (GUE^N), meaning that the joint density of entries of X^N is proportional to $\exp\left(-\frac{N}{2}\mathrm{Tr}(X^2)\right)$. Alternatively, X^N may be described as a *Gaussian Wigner matrix*, meaning it is Hermitian, and otherwise has i.i.d. centered Gaussian entries of variance $\frac{1}{N}$. Wigner's law states that the empirical spectral distribution converges weakly almost surely to a limit: the semicircle distribution $\frac{1}{2\pi}\sqrt{(4-x^2)_+}\,dx$, supported on $[-2,2]$ (cf. [50]). This holds for all Wigner matrices, independent of the distribution of the entries, cf. [2]. But this does not imply that the spectrum of X^N converges almost surely to $[-2,2]$; indeed, it is known that this spectral edge phenomenon occurs iff the fourth moments of the entries of X^N are finite (cf. [3]).

Our first major theorem is a spectral edge result for the empirical law of eigenvalues of the Brownian motion U_t^N. Since the spectrum is contained in the circle $\mathbb{U}_1$, instead of discussing the ill-defined "largest" eigenvalue, we characterize convergence in terms of Hausdorff distance d_H: the Hausdorff distance between two compact subsets A,B of a metric space is defined to be

$$d_H(A,B) = \inf\{\epsilon \geq 0\colon\, A \subseteq B_\epsilon \;\&\; B \subseteq A_\epsilon\},$$

where A_ϵ is the set of points within distance ϵ of A. It is easy to check that the spectral edge theorem for Wigner ensembles is equivalent to the statement that $d_H(\mathrm{spec}(X^N),[-2,2]) \to 0$ $a.s.$ as $N\to\infty$; for a related discussion, see Corollary 3.3 and Remark 3.4 below.

Theorem 1.1. *Let $N\in\mathbb{N}$, and let $(U_t^N)_{t\geq 0}$ be a Brownian motion on $\mathbb{U}_N$. Fix $t\geq 0$. Denote by ν_t the law of the free unitary Brownian motion, cf. Theorem 2.5. Then*

$$d_H\left(\mathrm{spec}(U_t^N), \mathrm{supp}\,\nu_t\right) \to 0 \;\; a.s. \text{ as } N\to\infty.$$

Remark 1.2. When $t\geq 4$, $\mathrm{supp}\,\nu_t = \mathbb{U}_1$, and Theorem 1.1 is immediate; the content here is that, for $0\leq t<4$, for large N *all* the eigenvalues are very close to the arc defined in (2.7).

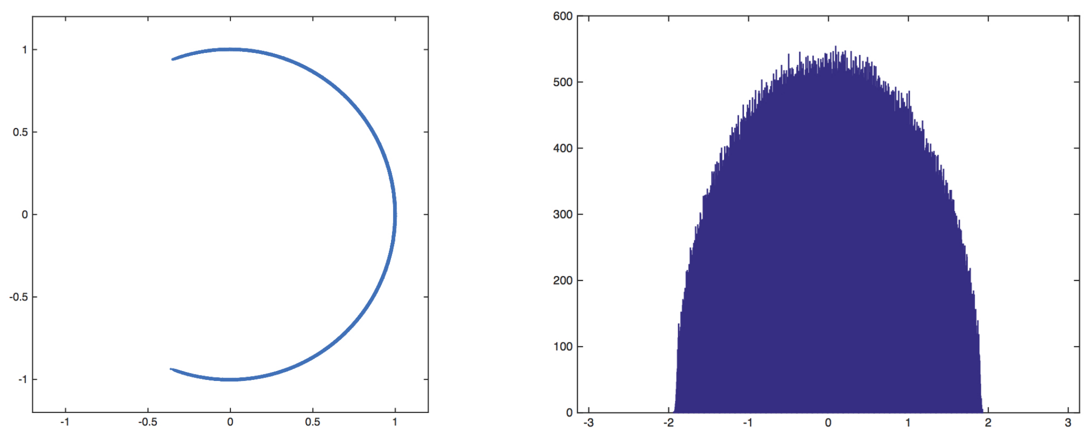

Figure 1: The spectrum of the unitary Brownian motion U_t^N with $N=400$ and $t=1$. These figures were produced from 1000 trials. On the left is a plot of the eigenvalues, while on the right is a 1000-bin histogram of their complex arguments. The argument range of the data is $[-1.9392, 1.9291]$, as compared to the predicted large-N limit range (to four digits) $[-1.9132, 1.9132]$, cf. (2.7).

To prove Theorem 1.1, our method is to prove sufficiently tight estimates on the rate of convergence of the moments of U_t^N. We record the main estimate here, since it is of independent interest.

Theorem 1.3. *Let $N,n\in\mathbb{N}$, and fix $t\geq 0$. Then*

$$\left|\mathbb{E}\mathrm{tr}\big[(U_t^N)^n\big] - \int_{\mathbb{U}_1} w^n\,\nu_t(dw)\right| \leq \frac{t^2n^4}{N^2}. \tag{1.3}$$

Theorems 1.1 and 1.3 are proved in Section 3.

The second half of this paper is devoted to a multi-time, multi-matrix extension of this result. Biane's main theorem in [4] states that the *process* $(U_t^N)_{t\geq 0}$ converges (in the sense of finite-dimensional noncommutative distributions) to a free unitary Brownian motion $(u_t)_{t\geq 0}$. To be precise: for any $k \in \mathbb{N}$ and times $t_1, \ldots, t_k \geq 0$, and any noncommutative polynomial $P \in \mathbb{C}\langle X_1, \ldots, X_{2k}\rangle$ in $2k$ indeterminates, the random trace moments of $(U_{t_j}^N)_{1\leq j\leq k}$ converge almost surely to the corresponding trace moments of $(u_{t_j})_{1\leq j\leq k}$:

$$\lim_{N\to\infty} \mathrm{tr}\big(P(U_{t_1}^N, (U_{t_1}^N)^*, \ldots, U_{t_k}^N, (U_{t_k}^N)^*)\big) = \tau\big(P(u_{t_1}, u_{t_1}^*, \ldots, u_{t_k}, u_{t_k}^*)\big) \quad a.s.$$

(Here τ is the tracial state on the noncommutative probability space where $(u_t)_{t\geq 0}$ lives; cf. Section 2.3.) This is the noncommutative extension of a.s. weak convergence of the empirical spectral distribution. The corresponding strengthening to the level of the spectral edge is *strong convergence*: instead of measuring moments with the linear functionals tr and τ, we insist on a.s. convergence of polynomials in *operator norm*. See Section 2.2 for a full definition and history.

Theorem 1.1 can be rephrased to say that, for any fixed $t \geq 0$, U_t^N converges strongly to u_t (cf. Corollary 3.3). Our second main theorem is the extension of this to any finite collection of times. In fact, we prove a more general extension theorem, as follows.

Theorem 1.4. *For each N, let $(U_t^N)_{t\geq 0}$ be a Brownian motion on $\mathbb{U}_N$. Let $A_1^N, \ldots, A_n^N$ be random matrix ensembles in $\mathbb{M}_N$ all independent from $(U_t^N)_{t\geq 0}$, and suppose that $(A_1^N, \ldots, A_n^N)$ converges strongly to $(a_1, \ldots, a_n)$. Let $(u_t)_{t\geq 0}$ be a free unitary Brownian motion freely independent from $\{a_1, \ldots, a_n\}$. Then, for any $k \in \mathbb{N}$, and any $t_1, \ldots, t_k \geq 0$,*

$$(A_1^N, \ldots, A_n^N, U_{t_1}^N, \ldots, U_{t_k}^N) \text{ converges strongly to } (a_1, \ldots, a_n, u_{t_1}, \ldots, u_{t_k}).$$

Theorem 1.4 is proved in Section 4.

We conclude the paper with an application of these strong convergence results to the empirical spectral distribution of the Jacobi process, in Theorem 5.7. We proceed now with Section 2, laying out the basic concepts, preceding results, and notation we will use throughout.

2 Background

Here we set notation and briefly recall some main ideas and results we will need to prove our main results. Section 2.1 introduces the Brownian motion $(U_t^N)_{t\geq 0}$ on $\mathbb{U}_N$. Section 2.2 discusses noncommutative distributions (which generalize empirical spectral distributions to collections of noncommuting random matrix ensembles, and beyond) and associated notions of convergence, including strong convergence. Finally, Section 2.3 reviews key ideas from free probability and free stochastic calculus, leading up to the definition of free unitary Brownian motion and its spectral measure ν_t.

2.1 Brownian Motion on $\mathbb{U}_N$

Throughout, $\mathbb{U}_N$ denotes the unitary group of rank N; its Lie algebra $\mathrm{Lie}(\mathbb{U}_N) = \mathfrak{u}_N$ consists of the skew-Hermitian matrices in $\mathbb{M}_N$, $\mathfrak{u}_N = \{X \in \mathbb{M}_N \colon X^* = -X\}$. We define a real inner product on $\mathfrak{u}_N$ by scaling the Hilbert-Schmidt inner product

$$\langle X, Y\rangle_N \equiv -N\mathrm{Tr}(XY), \qquad X, Y \in \mathfrak{u}_N.$$

As explained in [20], this is the unique scaling that gives a meaningful limit as $N \to \infty$.

Any vector $X \in \mathfrak{u}_N$ gives rise to a unique left-invariant vector field on $\mathbb{U}_N$; we denote this vector field as ∂_X (it is more commonly called $\widetilde{X}$ in the geometry literature). That is: ∂_X is a left-invariant derivation on $C^\infty(\mathbb{U}_N)$ whose action is

$$(\partial_X f)(U) = \frac{d}{dt}\bigg|_{t=0} f(Ue^{tX})$$

where e^{tX} denotes the usual matrix exponential (which is the exponential map for the matrix Lie group $\mathbb{U}_N$; in particular $e^{tX} \in \mathbb{U}_N$ whenever $X \in \mathfrak{u}_N$). The **Laplacian** $\Delta_{\mathbb{U}_N}$ on $\mathbb{U}_N$ (determined by the metric $\langle\cdot,\cdot\rangle_N$) is the second-order differential operator

$$\Delta_{\mathbb{U}_N} \equiv \sum_{X\in\beta_N} \partial_X^2$$

where β_N is any orthonormal basis for $\mathfrak{u}_N$; the operator does not depend on which orthonormal basis is used. The Laplacian is a negative definite elliptic operator; it is essentially self-adjoint in $L^2(\mathbb{U}_N)$ taken with respect to the Haar measure (cf. [41, 45]).

The **unitary Brownian motion** $U^N = (U_t^N)_{t\geq 0}$ is the Markov diffusion process on $\mathbb{U}_N$ with generator $\frac{1}{2}\Delta_{\mathbb{U}_N}$, with $U_0^N = I_N$. In particular, this means that the law of U_t^N at any fixed time $t \geq 0$ is the **heat kernel measure** on $\mathbb{U}_N$. This is essentially by definition: the heat kernel measure ρ_t^N is defined weakly by

$$\mathbb{E}_{\rho_t^N}(f) \equiv \int_{\mathbb{U}_N} f\, d\rho_t^N = \left(e^{\frac{t}{2}\Delta_{\mathbb{U}_N}} f\right)(I_N), \qquad f \in C(\mathbb{U}_N). \tag{2.1}$$

We mention here the fact that the heat kernel measure is symmetric: it is invariant under $U \mapsto U^{-1}$ (this is true on any Lie group).

There are (at least) two more constructive ways to understand the Brownian motion U^N directly. The first is as a Lévy process: U^N is uniquely defined by the following properties.

- CONTINUITY: The paths $t \mapsto U_t^N$ are a.s. continuous.
- INDEPENDENT MULTIPLICATIVE INCREMENTS: For $0 \leq s \leq t$, the multiplicative increment $(U_s^N)^{-1}U_t^N$ is independent from the filtration up to time s (i.e. from all random variables measurable with respect to the entires of U_r^N for $0 \leq r \leq s$).
- STATIONARY HEAT-KERNEL DISTRIBUTED INCREMENTS: For $0 \leq s \leq t$, the multiplicative increment $(U_s^N)^{-1}U_t^N$ has the distribution ρ_{t-s}^N.

In particular, since U_t^N is distributed according to ρ_t^N, we typically write expectations of functions on $\mathbb{U}_N$ with respect to ρ_t^N as

$$\mathbb{E}_{\rho_t^N}(f) = \mathbb{E}[f(U_t^N)].$$

For the purpose of computations, the best representation of U^N is as the solution to a stochastic differential equation. Let X^N be a GUE^N-valued Brownian motion: that is, X^N is Hermitian where the random variables $[X^N]_{jj}, \mathrm{Re}[X^N]_{jk}, \mathrm{Im}[X^N]_{jk}$ for $1 \leq j < k \leq N$ are all independent Brownian motions (of variance t/N on the main diagonal and $t/2N$ above it). Then U^N is the solution of the Itô stochastic differential equation

$$dU_t^N = iU_t^N\, dX_t^N - \frac{1}{2}U_t^N\, dt, \qquad U_0^N = I_N. \tag{2.2}$$

We will use this latter definition of U_t^N, via the SDE in terms of GUE^N-valued Brownian motion, almost exclusively throughout this paper.

2.2 Noncommutative distributions and convergence

Let $(\mathscr{A}, \tau)$ be a W^*-probability space: a von Neumann algebra $\mathscr{A}$ equipped with a faithful, normal, tracial state τ. Elements $a \in \mathscr{A}$ are referred to as (noncommutative) **random variables**. The **noncommutative distribution** of any finite collection $a_1, \ldots, a_k \in A$ is the linear functional $\mu_{(a_1,\ldots,a_k)}$ on noncommutative polynomials defined by

$$\begin{aligned} \mu_{(a_1,\ldots,a_k)} \colon \mathbb{C}\langle X_1, \ldots, X_k\rangle &\to \mathbb{C} \\ P &\mapsto \tau(P(a_1, \ldots, a_k)). \end{aligned} \tag{2.3}$$

Some authors explicitly include moments in a_j, a_j^* in the definition of the distribution; we will instead refer to the $*$-distribution as the noncommutative distribution $\mu_{(a_1,a_1^*,\ldots,a_k,a_k^*)}$ explicitly when needed. Note, when $a \in \mathscr{A}$ is normal, μ_{a,a^*} is determined by a unique probability measure Law_a, the spectral measure of a, on $\mathbb{C}$ in the usual way:

$$\int_{\mathbb{C}} f(z, \bar{z})\, \mathrm{Law}_a(dz d\bar{z}) = \mu_{a,a^*}(f), \quad f \in \mathbb{C}[X, X^*]$$

(i.e. when normal it suffices to restrict the noncommutative distribution to ordinary commuting polynomials). In this case, the support $\operatorname{supp}\mathrm{Law}_a$ is equal to the spectrum $\operatorname{spec}(a)$. If $u \in \mathscr{A}$ is unitary, Law_u is supported in the unit circle $\mathbb{U}_1$. For example: a **Haar unitary** is a unitary operator in $(\mathscr{A}, \tau)$ whose spectral measure is the uniform probability measure on $\mathbb{U}_1$ (equivalently $\tau(u^n) = \delta_{n0}$ for $n \in \mathbb{Z}$). In general, however, for a collection of elements $a_1, \ldots, a_k$ (normal or not) that do not commute, the noncommutative distribution is not determined by any measure on $\mathbb{C}$.

As a prominent example, let A^N be a normal random matrix ensemble in $\mathbb{M}_N$: i.e. A^N is a random variable defined on some probability space $(\Omega, \mathscr{F}, \mathbb{P})$, taking values in $\mathbb{M}_N$. The distribution of A^N as a random variable is a measure on $\mathbb{M}_N$; but for each instance $\omega \in \Omega$, the matrix $A^N(\omega)$ is a noncommutative random variable in the W^*-probability space $\mathbb{M}_N$, whose unique tracial state is tr. In this interpretation, the law $\mathrm{Law}_{A^N(\omega)}$ determined by its noncommutative distribution is precisely the empirical spectral distribution

$$\mathrm{Law}_{A^N(\omega)} = \frac{1}{N}\sum_{j=1}^{N} \delta_{\lambda_j(A^N(\omega))},$$

where $\lambda_1(A^N(\omega)), \ldots, \lambda_n(A^N(\omega))$ are the (random) eigenvalues of A^N.

Let $(A_1^N, \ldots, A_n^N)$ be a collection of random matrix ensembles, viewed as (random) noncommutative random variables in $(\mathbb{M}_N, \mathrm{tr})$. We will assume that the entries of A^N are in $L^{\infty-}(\Omega, \mathscr{F}, \mathbb{P})$, meaning that they have finite moments of all orders. The noncommutative distribution $\mu_{(A_1^N,\ldots,A_n^N)}$ is thus a random linear functional $\mathbb{C}\langle X_1, \ldots, X_n\rangle \to \mathbb{C}$; its value on a polynomial P is the (classical) random variable $\mathrm{tr}(P(A_1^N, \ldots, A_n^N))$, cf. (2.3). Now, let $(\mathscr{A}, \tau)$ be a W^*-probability space, and let $a_1, \ldots, a_n \in \mathscr{A}$. Say that $(A_1^N, \ldots, A_n^N)$ **converges in noncommutative distribution to** $a_1, \ldots, a_n$ **almost surely** if $\mu_{(A_1^N,\ldots,A_n^N)} \longrightarrow \mu_{(a_1,\ldots,a_n)}$ almost surely in the topology of pointwise convergence. That is to say: convergence in noncommutative distribution means that all (random) mixed tr moments of the ensembles A_j^N converge a.s. to the same mixed τ moments of the a_j. Later, a stronger notion of convergence emerged.

Definition 2.1. *Let* $\mathbf{A}^N = (A_1^N, \ldots, A_n^N)$ *be random matrix ensembles in* $(\mathbb{M}_N, \mathrm{tr})$*, and let* $\mathbf{a} = (a_1, \ldots, a_n)$ *be random variables in a* W^**-probability space* $(\mathscr{A}, \tau)$*. Say that* $\mathbf{A}^N$ **converges strongly to a** *if* $\mathbf{A}^N$ *converges to* $\mathbf{a}$ *almost surely in noncommutative distribution, and additionally*

$$\|P(A_1^N, \ldots, A_n^N)\|_{\mathbb{M}_N} \to \|P(a_1, \ldots, a_n)\|_{\mathscr{A}} \;\; a.s. \qquad \forall\, P \in \mathbb{C}\langle X_1, \ldots, X_n\rangle.$$

(Here $\|\cdot\|_{\mathbb{M}_N}$ *denotes the usual operator norm on* $\mathbb{M}_N$*, and* $\|\cdot\|_{\mathscr{A}}$ *denotes the operator norm on* $\mathscr{A}$*.)*

This notion first appeared in the seminal paper [24] of Haagerup and Thorbjørnsen, where they showed that if $X_1^N, \ldots, X_n^N$ are independent GUE^N random matrices, then they converge strongly to free semicircular random variables $(x_1, \ldots, x_n)$. The notion was formalized into Definition 2.1 in the dissertation of Male (cf. [35]).

Remark 2.2. It should be noted that the choice of terminology *strong convergence* is at odds with the standard notion of strong topology in functional analysis, which certainly does not involve the operator norm! While this may be jarring to some readers, the terminology is now standard in free probability circles.

Male's paper [35] also proved the following generalization: an extension property of strong convergence).

Theorem 2.3 (Male, 2012)**.** *Let $\mathbf{A}^N = (A_1^N, \ldots, A_n^N)$ be a collection of random matrix ensembles that converges strongly to some $\mathbf{a} = (a_1, \ldots, a_n)$ in a W^*-probability space $(\mathscr{A}, \tau)$. Let $\mathbf{X}^N = (X_1^N, \ldots, X_k^N)$ be independent Gaussian unitary ensembles independent from $\mathbf{A}^N$, and let $\mathbf{x} = (x_1, \ldots, x_k)$ be freely independent semicircular random variables in $\mathscr{A}$ all free from $\mathbf{a}$. Then $(\mathbf{A}^N, \mathbf{X}^N)$ converges strongly to $(\mathbf{a}, \mathbf{x})$.*

(For a brief definition and discussion of free independence, see Section 2.3 below.) Later, together with the present first author in [13], Male proved a strong convergence result for Haar distributed random unitary matrices (which can be realized as $\lim_{t\to\infty} U_t^N$).

Theorem 2.4 (Collins, Male, 2013)**.** *Let $\mathbf{A}^N = (A_1^N, \ldots, A_n^N)$ be a collection of random matrix ensembles that converges strongly to some $\mathbf{a} = (a_1, \ldots, a_n)$ in a W^*-probability space $(\mathscr{A}, \tau)$. Let U^N be a Haar-distributed random unitary matrix independent from $\mathbf{A}^N$, and let u be a Haar unitary operator in $\mathscr{A}$ freely independent from $\mathbf{a}$. Then $(\mathbf{A}^N, U^N, (U^N)^*)$ converges strongly to $(\mathbf{a}, u, u^*)$.*

(The convergence in distribution in Theorem 2.4 is originally due to Voiculescu [47]; a simpler proof of this result was given in [10].) Note that, for any matrix $A \in \mathbb{M}_N$ and any operator $a \in \mathscr{A}$,

$$\|A\|_{\mathbb{M}_N} = \lim_{p\to\infty} \left(\mathrm{tr}\left[(AA^*)^{p/2}\right]\right)^{1/p}, \quad \text{and} \quad \|a\|_{\mathscr{A}} = \lim_{p\to\infty} \left(\tau\left[(aa^*)^{p/2}\right]\right)^{1/p}.$$

These hold because the states tr and τ are faithful. These are the noncommutative L^p-norms on $L^p(\mathbb{M}_N, \mathrm{tr})$ and $L^p(\mathscr{A}, \tau)$ respectively. The norm-convergence statement of strong convergence can thus be rephrased as an almost sure interchange of limits: if $\mathbf{A}^N$ converges a.s. to $\mathbf{a}$ in noncommutative distribution, then $\mathbf{A}^N$ converges to $\mathbf{a}$ strongly if and only if

$$\mathbb{P}\left(\lim_{N\to\infty}\lim_{p\to\infty} \|P(\mathbf{A}^N)\|_{L^p(\mathbb{M}_N,\mathrm{tr})} = \lim_{p\to\infty} \|P(\mathbf{a})\|_{L^p(\mathscr{A},\tau)}\right) = 1, \quad \forall\, P \in \mathbb{C}\langle X_1, \ldots, X_n\rangle. \tag{2.4}$$

If we fix p instead of sending $p \to \infty$, the corresponding notion of "L^p-strong convergence" of the unitary Brownian motion $(U_t^N)_{t\geq 0}$ to the free unitary Brownian motion $(u_t)_{t\geq 0}$ was proved in the third author's paper [30]. This weaker notion of strong convergence does not have the same important applications as strong convergence, however. As a demonstration of the power of true strong convergence, we give an application to the eigenvalues of the Jacobi process in Section 5: the principal angles between subspaces randomly rotated by U_t^N evolve a.s. with finite speed for all large N.

2.3 Free probability, free stochastics, and free unitary Brownian motion

We briefly recall basic definitions and constructions here, mostly for the sake of fixing notation. The uninitiated reader is referred to the monographs [38, 49], and the introductions of the authors' previous papers [12, 30, 32] for more details.

Let $(\mathscr{A}, \tau)$ be a W^*-probability space. Unital subalgebras $\mathscr{A}_1, \ldots, \mathscr{A}_m \subset \mathscr{A}$ are called **free** or **freely independent** if the following property holds: given any sequence of indices $k_1, \ldots, k_n \in \{1, \ldots, m\}$ that are consecutively-distinct (meaning $k_{j-1} \neq k_j$ for $1 < j \leq n$) and random variables $a_j \in \mathscr{A}_{k_j}$, if $\tau(a_j) = 0$ for $1 \leq j \leq n$ then $\tau(a_1 \cdots a_n) = 0$. We say random variables $a_1, \ldots, a_m$ are freely independent if the unital $*$-subalgebras $\mathscr{A}_j \equiv \langle a_j, a_j^* \rangle \subset \mathscr{A}$ they generate are freely independent. Freeness is a moment factorization property: by centering random variables $a \to a - \tau(a)1_{\mathscr{A}}$, freeness allows the (recursive) computation of any joint moment in free variables as a polynomial in the moments of the separate random variables. In other words: the distribution $\mu_{(a_1,\ldots,a_k)}$ of a collection of free random variables is determined by the distributions $\mu_{a_1}, \ldots, \mu_{a_k}$ separately.

A **noncommutative stochastic process** is simply a one-parameter family $a = (a_t)_{t \geq 0}$ of random variables in some W^*-probability space $(\mathscr{A}, \tau)$. It defines a **filtration**: an increasing (by inclusion) collection $\mathscr{A}_t$ of subalgebras of $\mathscr{A}$ defined by $\mathscr{A}_t \equiv W^*(a_s\colon 0 \leq s \leq t)$, the von Neumann algebras generated by all the random variables a_s for $s \leq t$. Given such a filtration $(\mathscr{A}_t)_{t \geq 0}$, we call a process $b = (b_t)_{t \geq 0}$ **adapted** if $b_t \in \mathscr{A}_t$ for all $t \geq 0$.

A **free additive Brownian motion** is a selfadjoint noncommutative stochastic process $x = (x_t)_{t \geq 0}$ in a W^*-probability space $(\mathscr{A}, \tau)$ with the following properties:

- CONTINUITY: The map $\mathbb{R}_+ \to \mathscr{A} : t \mapsto x_t$ is weak*-continuous.
- FREE INCREMENTS: For $0 \leq s \leq t$, the additive increment $x_t - x_s$ is freely independent from $\mathscr{A}_s$ (the filtration generated by x up to time s).
- STATIONARY INCREMENTS: For $0 \leq s \leq t$, $\mu_{x_t - x_s} = \mu_{x_{t-s}}$.

It follows from the free central limit theorem that the increments must have the semicircular distribution: $\mathrm{Law}_{x_t} = \frac{1}{2\pi t}\sqrt{(4t - x^2)_+}\, dx$. Voiculescu (cf. [46, 47, 49]) showed that free additive Brownian motions exist: they can be constructed in any W^*-probability space rich enough to contain an infinite sequence of freely independent semicircular random variables (where x can be constructed in the usual way as an isonormal process).

In pioneering work of Biane and Speicher [6, 7] (and many subsequent works such as the third author's joint paper with Nourdin, Peccati, and Speicher [32]), a theory of stochastic analysis built on x was developed. **Free stochastic integrals** with respect to x are defined precisely as in the classical setting: as $L^2(\mathscr{A}, \tau)$-limits of integrals of simple processes, where for constant $a \in \mathscr{A}$, $\int_0^t \mathbb{1}_{[t_-, t_+]}(s) a\, dx_s$ is defined to be $a \cdot (x_{t_+} - x_{t_-})$. Using the standard Picard iteration techniques, it is known that free stochastic integral equations of the form

$$a_t = a_0 + \int_0^t \phi(s, a_s)\, ds + \int_0^t \sigma(s, a_s)\, dx_s \tag{2.5}$$

have unique adapted solutions for drift ϕ and diffusion σ coefficient functions that are globally Lipschitz. Note: due to the noncommutativity, the kinds of processes one should really use in the stochastic integral are *biprocesses*: $\beta_t \in \mathscr{A} \otimes \mathscr{A}$. The stochastic integral against a single free Brownian motion x_t is then denoted $\beta_t \# dx_t$, where this is defined so that, if $\beta_t = a_t \otimes b_t$ is a pure tensor state, then $\beta_t \# dx_t = a_t dx_t b_t$, allowing the process to act on both sides of the Brownian motion. (See [6, 32] for details.) A one-sided process like the one in (2.5) is typically not self-adjoint, which limits ϕ, σ to be polynomials, and ergo linear polynomials due to the Lipschitz constraint. That will suffice for our present purposes.) Equations like (2.5) are often written in "stochastic differential" form as

$$da_t = \phi(t, a_t)\, dt + \sigma(t, a_t)\, dx_t.$$

Given a free additive Brownian motion x, the associated **free unitary Brownian motion** $u = (u_t)_{t\geq 0}$ is the solution to the free SDE

$$du_t = iu_t\,dx_t - \frac{1}{2}u_t\,dt, \qquad u_0 = 1. \tag{2.6}$$

This precisely mirrors the (classical) Itô SDE (2.2) that determines the Brownian motion $(U_t^N)_{t\geq 0}$ on $\mathbb{U}_N$.

The free unitary Brownian motion $(u_t)_{t\geq 0}$ was introduced by Biane in [4] via the above definition. In that paper, with more details in Biane's subsequent [5], together with independent statements of the same type in [40], Law_{u_t} was computed. Since u_t is unitary, this distribution is determined by a measure ν_t that is supported on the unit circle $\mathbb{U}_1$. This measure is described as follows.

Theorem 2.5 (Biane 1997). *For $t > 0$, ν_t has a continuous density ϱ_t with respect to the normalized Haar measure on $\mathbb{U}_1$. For $0 < t < 4$, its support is the connected arc*

$$\operatorname{supp}\nu_t = \left\{e^{i\theta}\colon |\theta| \leq \frac{1}{2}\sqrt{t(4-t)} + \arccos\left(1 - \frac{t}{2}\right)\right\}, \tag{2.7}$$

while $\operatorname{supp}\nu_t = \mathbb{U}_1$ *for $t \geq 4$. The density ϱ_t is real analytic on the interior of the arc. It is symmetric about 1, and is determined by $\varrho_t(e^{i\theta}) = \operatorname{Re}\kappa_t(e^{i\theta})$ where $z = \kappa_t(e^{i\theta})$ is the unique solution (with positive real part) to*

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

Note that the arc (2.7) is the spectrum $\operatorname{spec}(u_t)$ for $0 < t < 4$; for $t \geq 4$, $\operatorname{spec}(u_t) = \mathbb{U}_1$.

With this description, one can also give a characterization of the free unitary Brownian motion similar to the invariant characterization of the Brownian motion $(U_t^N)_{t\geq 0}$ on page 5. That is, $(u_t)_{t\geq 0}$ is the unique unitary-valued process that satisfies:

- Continuity: The map $\mathbb{R}_+ \to \mathscr{A}\colon t \mapsto u_t$ is weak* continuous.
- Freely Independent Multiplicative Increments: For $0 \leq s \leq t$, the multiplicative increment $u_s^{-1}u_t$ is independent from the filtration up to time s (i.e. from the von Neumann algebra $\mathscr{A}_s$ generated by $\{u_r\colon 0 \leq r \leq s\}$).
- Stationary Increments with Distribution ν: For $0 \leq s \leq t$, the multiplicative increment $u_s^{-1}u_t$ has distribution given by the law ν_{t-s}.

3 The Edge of the Spectrum

This section is devoted to the proof of our spectral edge theorem for a single time marginal U_t^N. We begin by showing how Theorem 1.1 follows from Theorem 1.3, and recast the conclusion as a strong convergence statement in Corollary 3.3. Section 3.2 is then devoted to the proof of the moment growth bound of Theorem 1.3.

3.1 Strong convergence and the proof of Theorem 1.1

We begin by briefly recalling some basic Fourier analysis on the circle $\mathbb{U}_1$. For $f \in L^2(\mathbb{U}_1)$, its Fourier expansion is

$$f(w) = \sum_{n\in\mathbb{Z}} \hat{f}(n)w^n, \quad \text{where} \quad \hat{f}(n) = \int_{\mathbb{U}_1} f(w)w^{-n}\,dw,$$

where dw is the normalized Lebesgue measure on $\mathbb{U}_1$. For $p > 0$, the **Sobolev space** $H_p(\mathbb{U}_1)$ is defined to be

$$H_p(\mathbb{U}_1) = \left\{ f \in L^2(\mathbb{U}_1)\colon \|f\|_{H_p}^2 \equiv \sum_{n\in\mathbb{Z}} (1+n^2)^p |\hat{f}(n)|^2 < \infty \right\}. \tag{3.1}$$

If $\ell > k \geq 1$ are integers, and $\ell \geq p \geq k + \frac{1}{2}$, then $C^\ell(\mathbb{U}_1) \subset H_p(\mathbb{U}) \subset C^k(\mathbb{U}_1)$; it follows that $H_\infty(\mathbb{U}_1) \equiv \bigcap_{p\geq 0} H_p(\mathbb{U}_1) = C^\infty(\mathbb{U}_1)$. These are standard Sobolev imbedding theorems (that hold for smooth manifolds); for reference, see [22, Chapter 5.6] and [42, Chapter 3.2].

Theorem 1.3 yields the following estimate on moment growth tested against Sobolev functions disjoint from the limit support.

Proposition 3.1. *Fix $0 \leq t < 4$. Let $f \in H_5(\mathbb{U}_1)$ have support disjoint from $\operatorname{supp}\nu_t$. There is a constant $C(f) > 0$ such that, for all $N \in \mathbb{N}$,*

$$\left|\mathbb{E}\mathrm{tr}[f(U_t^N)]\right| \leq \frac{t^2 C(f)}{N^2}. \tag{3.2}$$

Proof. Denote by $\nu_t^N(n) \equiv \mathbb{E}\mathrm{tr}[(U_t^N)^n]$ and $\nu_t(n) \equiv \int_{\mathbb{U}_1} w^n\, \nu_t(dw) = \lim_{N\to\infty} \nu_t^N(n)$. Expanding f as a Fourier series, we have

$$\mathbb{E}\mathrm{tr}[f(U_t^N)] = \sum_{n\in\mathbb{Z}} \hat{f}(n)\mathbb{E}\mathrm{tr}[(U_t^N)^n] = \sum_{n\in\mathbb{Z}} \hat{f}(n)\nu_t^N(n). \tag{3.3}$$

By the assumption that $\operatorname{supp} f$ is disjoint from $\operatorname{supp}\nu_t$, we have

$$0 = \int_{\mathbb{U}_1} f\, d\nu_t = \sum_{n\in\mathbb{Z}} \hat{f}(n) \int_{\mathbb{U}_1} w^n\, \nu_t(dw) = \sum_{n\in\mathbb{Z}} \hat{f}(n)\nu_t(n). \tag{3.4}$$

Combining (3.3) and (3.4) with Theorem 1.3 yields

$$\left|\mathbb{E}\mathrm{tr}[f(U_t^N)]\right| \leq \sum_{n\in\mathbb{Z}} |\hat{f}(n)||\nu_t^N(n) - \nu_t(n)| \leq \sum_{n\in\mathbb{Z}} |\hat{f}(n)| \cdot \frac{t^2 n^4}{N^2}.$$

By assumption $f \in H_5(\mathbb{U}_1)$, and so

$$\sum_{n\in\mathbb{Z}} n^4|\hat{f}(n)| = \sum_{n\in\mathbb{Z}\setminus\{0\}} \frac{1}{n}\cdot n^5|\hat{f}(n)| \leq \left(\sum_{n\in\mathbb{Z}\setminus\{0\}} \frac{1}{n^2}\right)^{1/2} \left(\sum_{n\in\mathbb{Z}} n^{10}|\hat{f}(n)|^2\right)^{1/2} \leq \frac{\pi}{\sqrt{3}}\|f\|_{H_5} < \infty.$$

Taking $C(f) = \frac{\pi}{\sqrt{3}}\|f\|_{H_5}$ concludes the proof. □

We now use Proposition 3.1 to give an improved variance estimate related to [34, Propositions 6.1, 6.2].

Proposition 3.2. *Fix $0 \leq t < 4$. Let $f \in C^6(\mathbb{U}_1)$ with support disjoint from $\operatorname{supp}\nu_t$. There is a constant $C'(f) > 0$ such that, for all $N \in \mathbb{N}$,*

$$\mathrm{Var}[\mathrm{Tr}(f(U_t^N))] \leq \frac{t^3 C'(f)}{N^2}.$$

Proof. In the proof of [34, Proposition 3.1] (on p. 3179), and also in [9, Proposition 4.2 & Corollary 4.5], the desired variance is shown to have the form

$$\mathrm{Var}[\mathrm{Tr}(f(U_t^N))] = \int_0^t \mathbb{E}\mathrm{tr}[f'(U_s^N V_{t-s}^N) f'(U_s^N W_{t-s}^N)]\, ds \tag{3.5}$$

where U^N, V^N, W^N are three independent Brownian motions on $\mathbb{U}_N$. For fixed $s \in [0,t]$, we apply the Cauchy-Schwarz inequality twice and use the equidistribution of $U_s^N V_{t-s}^N$ and $U_s^N W_{t-s}^N$ to yield

$$\left|\mathbb{E}\mathrm{tr}[f'(U_s^N V_{t-s}^N) f'(U_s^N W_{t-s}^N)]\right| \le \mathbb{E}\left[\left|\mathrm{tr}[f'(U_s^N V_{t-s}^N)^2]\right|^{1/2} \cdot \left|\mathrm{tr}[f'(U_s^N W_{t-s}^N)^2]\right|^{1/2}\right] \le \mathbb{E}\mathrm{tr}[f'(U_s^N V_{t-s}^N)^2].$$

Since U^N and V^N are independent, (U_s^N, V_{t-s}^N) has the same distribution as $(U_s^N, (U_s^N)^{-1}U_t^N)$ (as the increments are independent and stationary). Thus $\mathbb{E}\mathrm{tr}[f'(U_s^N V_{t-s}^N)^2] = \mathbb{E}\mathrm{tr}[f'(U_t^N)^2]$, and so, integrating, we find

$$\mathrm{Var}[\mathrm{Tr}(f(U_t^N))] \le t\mathbb{E}\mathrm{tr}[f'(U_t^N)^2]. \tag{3.6}$$

Since $f \in C^6(\mathbb{U}_1)$, the function $(f')^2$ is $C^5 \subset H_5$, and the result now follows from Proposition 3.1, with $C'(f) = C((f')^2)$. $\square$

This brings us to the proof of the spectral edge theorem.

Proof of Theorem 1.1 assuming Theorem 1.3. Fix a closed arc $\alpha \subset \mathbb{U}_1$ that is disjoint from $\mathrm{supp}\,\nu_t$. Let f be a C^∞ bump function with values in $[0,1]$ such that $f|_\alpha = 1$ and $\mathrm{supp}\, f \cap \mathrm{supp}\,\nu_t = \varnothing$. Then

$$\mathbb{P}(\mathrm{spec}(U_t^N) \cap \alpha \neq \varnothing) \le \mathbb{P}(\mathrm{Tr}[f(U_t^N)] \ge 1). \tag{3.7}$$

We now apply Chebyshev's inequality, in the following form: let $Y = \mathrm{Tr}[f(U_t^N)]$. Then, assuming $1 - \mathbb{E}(Y) > 0$, we have

$$\mathbb{P}(Y \ge 1) = \mathbb{P}(Y - \mathbb{E}(Y) \ge 1 - \mathbb{E}(Y)) \le \frac{\mathrm{Var}(Y)}{(1-\mathbb{E}(Y))^2}.$$

In our case, we have $|\mathbb{E}(Y)| = |\mathbb{E}\mathrm{Tr}[f(U_t^N)]| = N|\mathbb{E}\mathrm{tr}[f(U_t^N)]| \le \frac{t^2 C(f)}{N}$ by Proposition 3.1. Thus, there is N_0 (depending only on f and t) so that $(1 - \mathbb{E}\mathrm{Tr}[f(U_t^N)])^2 \ge \frac{1}{2}$ for $N \ge N_0$. Combining this with (3.7) yields

$$\mathbb{P}(\mathrm{spec}(U_t^N) \cap \alpha \neq \varnothing) \le 2\mathrm{Var}[\mathrm{Tr}(f(U_t^N))] \quad \text{for} \quad N \ge N_0.$$

Now invoking Proposition 3.2, we find that this is $\le \frac{2t^3 C'(f)}{N^2}$ whenever $N \ge N_0$. It thus follows from the Borel-Cantelli lemma that the probability that $\mathrm{spec}(U_t^N) \cap \alpha \neq \varnothing$ for infinitely many N is 0; i.e. with probability 1, for all sufficiently large N, U_t^N has no eigenvalues in α.

Thus, we have shown that, for any closed arc α disjoint from $\mathrm{supp}\,\nu_t$, with probability 1, $\mathrm{spec}(U_t^N)$ is contained in $\mathbb{U}_1 \setminus \alpha$ for all large N. In particular, fixing any open arc $\beta \subset \mathbb{U}_1$ containing $\mathrm{supp}\,\nu_t$, this applies to $\alpha = \mathbb{U}_1 \setminus \beta$. I.e. $\mathrm{spec}(U_t^N)$ is a.s. contained in any neighborhood of $\mathrm{supp}\,\nu_t$ for all sufficiently large N. This suffices to prove the theorem: because $\mathrm{Law}_{U_t^N}$ converges weakly almost surely to the measure ν_t which possesses a strictly positive continuous density on its support, any neighborhood of the spectrum of U_t^N eventually covers $\mathrm{supp}\,\nu_t$. $\square$

Thus, we have proved Theorem 1.1 under the assumption that Theorem 1.3 is true. Before turning to the proof of this latter result, let us recast Theorem 1.1 in the language of strong convergence, as we will proceed to generalize this to the fully noncommutative setting in Section 4.

Corollary 3.3. *For $N \in \mathbb{N}$, let $(U_t^N)_{t\geq 0}$ be a Brownian motion on $\mathbb{U}_N$. Let $(u_t)_{t\geq 0}$ be a free unitary Brownian motion. Then for any fixed $t \geq 0$, $(U_t^N, (U_t^N)^*)$ converges strongly to (u_t, u_t^*).*

Proof. Since $U_t^N \to u_t$ in noncommutative distribution, strong convergence is the statement that

$$\|P(U_t^N, (U_t^N)^*)\| \to \|P(u_t, u_t^*)\|$$

in operator norms. Fix a noncommutative polynomial P in two variables, and let p be the unique Laurent polynomial in one variable so that $P(U, U^*) = p(U)$ for every unitary operator U. Since U_t^N is normal, $\|p(U_t^N)\| = \max\{|\lambda| \colon \lambda \in p(\mathrm{spec}(U_t^N))\}$; similarly, $\|p(u_t)\| = \max\{|\lambda| \colon \lambda \in p(\mathrm{supp}\,\nu_t)\}$ where $\mathrm{supp}\,\nu_t$ is the arc in (2.7).

Let $\Lambda_p^N = |p|(\mathrm{spec}(U_t^N))$, and let $\Lambda_p = |p|(\mathrm{supp}\,\nu_t)$. (Here $|p|$ denotes the modulus of the polynomial function, $|p|(u) = |p(u)|$.) Since $\mathrm{spec}(U_t^N)$ converges to $\mathrm{supp}\,\nu_t$ in Hausdorff distance and all the sets are compact, it follows easily from the continuity of $|p|$ (on the unit circle) that Λ_p^N converges to Λ_p in Hausdorff distance as well. In particular, for any $\epsilon > 0$, there is some $N_0 \in \mathbb{N}$ so that, for all $N \geq N_0$, $(\Lambda_p^N)_\epsilon \subseteq \Lambda_p$ and $(\Lambda_p)_\epsilon \subseteq \Lambda_p^N$. It follows that $\max \Lambda_p - \epsilon \leq \max \Lambda_p^N \leq \max \Lambda_p + \epsilon$. This shows that $\|p(U_t^N)\| = \max \Lambda_p^N \to \max \Lambda_p = \|p(u_t)\|$, as desired. $\square$

Remark 3.4. In fact, the converse of Corollary 3.3 also holds: strong convergence of $U_t^N \to u_t$ (for a fixed $t < 4$) implies convergence of the spectrum in Hausdorff distance. Indeed, suppose we know strong convergence. In particular, taking the polynomial $P(U) = U - 1$, we then have $\|U_t^N - I_N\| \to \|u_t - 1\|$. Since the spectrum of u_t is an arc symmetric about 1, $\mathrm{spec}(u_t) = B_{\|u_t-1\|}(1) \cap \mathbb{U}_1$ (here $B_R(1)$ denotes the ball $\{z \in \mathbb{C} \colon |z-1| < R\}$). Note that $\|U_t^N - I_N\| = \max_k \|\lambda_k(t) - 1\|$; this shows that all eigenvalues $\lambda_k(t)$ are in an arc very close to $\mathrm{spec}(u_t)$ for large N. This shows that $\mathrm{spec}(U_t^N)$ is eventually contained in any neighborhood of $\mathrm{supp}\,\nu_t$; the other half of the convergence in Hausdorff distance follows from the convergence in distribution (and strict positivity of the limit density ν_t on $\mathrm{supp}\,\nu_t$).

When $t \geq 4$, $\mathrm{supp}\,\nu_t = \mathbb{U}_1$, and strong convergence becomes vacuously equivalent to the known convergence in distribution.

3.2 The proof of Theorem 1.3

This section is devoted to the proof of the main moment growth bound, i.e. (1.3). Before proceeding with our proof which is quite involved, it is worth noting a key feature. The bound (1.3) on the speed of convergence $\nu_t^N(n) \to \nu_t(n)$ depends polynomially on n; this is crucial to the proof of Theorem 1.1. If one only requires the $O(\frac{1}{N^2})$ bound of this estimate without much regard for the dependence on n, a much simpler (though still nontrivial) approach can be found in the third auhor's paper [30, Section 3.3], which provides a bound of the form form $K(t,n)/N^2$, where $K(t,n) \sim \frac{tn^2}{2}\exp(\frac{tn^2}{2})$. This growth in n is much too large to get control over test functions that are only in a Sobolev space, or even in $C^\infty(\mathbb{U}_1)$; the largest class of functions for which this Fourier series is summable is an ultra-analytic Gevrey class. That blunter estimate does not suffice for our present purposes; showing polynomial dependence on n turns out to require very careful analysis so as not to ignore the many cancellations in the complicated expansions for the moments.

Remark 3.5. We note that the blunter bound in [30] was proved not only for U^N, but for a family of diffusions on $\mathbb{GL}_N$ including both U^N and the Brownian motion on $\mathbb{GL}_N$. It remains open whether a polynomial bound like (1.3) holds for this wider class of diffusions.

The proof of the weaker bound described above relied on an explicit decomposition of the Laplacian on $\mathbb{U}_N$ in the form $D + \frac{1}{N^2}L$, which was exploited in the third author's papers [30, 31] and joint work with Driver and Hall [20] (and Cébron's independent work [8]), in the complementary work of the second author [14] and preceding paper of T. Lévy [33], and in some form also in the preceding work of Rains [40], Sengupta [43], and others. In that approach, the idea was to decompose the relevant space of functions (polynomials in traces of powers of the matrix) into a nested family of finite-dimensional subspaces all invariant under the Laplacian, and then appropriately estimate the distortion between the norm of $\exp(D)$ and $\exp(D+\frac{1}{N^2}L)$ on each such space. While an approach of that nature might conceivably yield bounds similar to (1.3), our present approach is quite different.

To begin: we will actually prove the following Cauchy sequence growth estimate. We again use the notation $\nu_t^N(n)=\mathbb{E}[\mathrm{tr}(U_t^N)^n]$.

Proposition 3.6. *Let $N,n\in\mathbb{N}$, and fix $t\geq 0$. Then*

$$\left|\nu_t^N(n)-\nu_t^{2N}(n)\right|\leq\frac{3t^2n^4}{4N^2}. \tag{3.8}$$

This is the main technical result of the first part of the paper, and its proof will occupy most of this section. Let us first show how Theorem 1.3 follows from Proposition 3.6.

Proof of Theorem 1.3 assuming Proposition 3.6. Since $\lim_{N\to\infty}\nu_t^N(n)=\nu_t(n)$, we have the following convergent telescoping series:

$$|\nu_t^N(n)-\nu_t(n)|=\left|\sum_{k=0}^{\infty}\left(\nu_t^{N2^k}(n)-\nu_t^{N2^{k+1}}(n)\right)\right|\leq\sum_{k=0}^{\infty}\left|\nu_t^{N2^k}(n)-\nu_t^{N2^{k+1}}(n)\right|.$$

Now apply (3.8) with N replaced by $N2^k$, we find

$$\left|\nu_t^{N2^k}(n)-\nu_t^{N2^{k+1}}(n)\right|\leq\frac{3}{4}\frac{t^2n^4}{(N2^k)^2}=\frac{3}{4}\frac{1}{4^k}\frac{t^2n^4}{N^2}.$$

Summing the geometric series proves the theorem. □

3.2.1 Outline of the Proof of Proposition 3.6

The proposition requires comparison of a moment $\nu_t^N(n)$ of an $N\times N$ unitary Brownian motion with a moment $\nu_t^{2N}(n)$ of a $2N\times 2N$ unitary Brownian motion. To compare them, we need to choose a coupling between the two processes in different spaces. As such, fix a Brownian motion U^{2N} on $\mathbb{U}_{2N}$, along with two Brownian motions $U^{N,1},U^{N,2}$ on $\mathbb{U}_N$, so that the processes $U^{2N},U^{N,1},U^{N,2}$ are all independent. For $t\geq 0$, let $B_t^{2N}\in\mathbb{U}_{2N}$ denote the block diagonal random matrix

$$B_t^{2N}=\begin{bmatrix} U_t^{N,1} & 0\\ 0 & U_t^{N,2}\end{bmatrix}\in\mathbb{U}_{2N}.$$

We will hold t fixed through; let us fix the notation

$$A_s^{2N}=U_{t-s}^{2N}B_s^{2N}.$$

This process will be used very often in what follows.

Setting $s=0$, we have $A_0^{2N}=U_t^{2N}$; on the other hand, setting $s=t$, we have $A_t^{2N}=B_t^{2N}$. Hence

$$\begin{aligned}&\mathbb{E}\mathrm{tr}[(A_0^{2N})^n]=\mathbb{E}\mathrm{tr}[(U_t^{2N})^n]=\nu_t^{2N}(n),\quad\text{while}\\ &\mathbb{E}\mathrm{tr}[(A_t^{2N})^n]=\mathbb{E}\mathrm{tr}[(B_t^{2N})^n]=\frac{1}{2N}\mathbb{E}\left(\mathrm{Tr}[(U_t^{N,1})^n]+\mathrm{Tr}[(U_t^{N,2})^n]\right)=\mathbb{E}\mathrm{tr}[(U_t^{N,1})^n]=\nu_t^N(n).\end{aligned}$$

Therefore, setting

$$F(s) = \mathbb{E}\mathrm{tr}[(A_s^{2N})^n] = \mathbb{E}\mathrm{tr}[(U_{t-s}^{2N} B_s^{2N})^n] \tag{3.9}$$

we see that the quantity $\nu_t^N(n) - \nu_t^{2N}(n)$ to be estimated in Proposition 3.6 is equal to

$$\nu_t^N(n) - \nu_t^{2N}(n) = F(t) - F(0).$$

The main approach of the proof will be to show that $F \in C^1[0,t]$, so that the Fundamental Theorem of Calculus allows us to express

$$\nu_t^N(n) - \nu_t^{2N}(n) = \int_0^t F'(s)\,ds \tag{3.10}$$

and then appropriately estimate $|F'(s)|$ to achieve the desired bound of (3.8).

Due to the nonlinearity (for $n > 1$) of the trace moment $F(s) = \mathbb{E}\mathrm{tr}[(A_s^{2N})^n]$ as a function of A_s^{2N}, it is useful to multilinearize this quantity in what follows. To do so, we will work in the n-fold tensor product of $\mathbb{M}_{2N}$, essentially replacing a power A^n with a tensor power $A^{\otimes n} = A \otimes A \otimes \cdots \otimes A \in (\mathbb{M}_{2N})^{\otimes n}$; since we could identify this space with $\mathbb{M}_{2Nn}$, we will still refer to such tensor product operators as matrices. To recover the original quantity (in terms of the trace of A^n), we cannot simply take the trace of $A^{\otimes n}$ (which would produce $[\mathrm{Tr}(A)]^n$); instead, we must involve an action of the symmetric group S_n. To any permutation $\sigma \in S_n$, we associate a matrix $[\sigma] \in (\mathbb{M}_{2N})^{\otimes n}$ by permuting the tensor indices: for any vector $v_1 \otimes \cdots \otimes v_n \in (\mathbb{C}^{2N})^{\otimes n}$,

$$[\sigma](v_1 \otimes \cdots \otimes v_n) = v_{\sigma^{-1}(1)} \otimes \cdots \otimes v_{\sigma^{-1}(n)}. \tag{3.11}$$

Remark 3.7. This representation of S_n on $(\mathbb{C}^{2N})^{\otimes n}$ is sometimes called the *Schur–Weyl representation*. In general, for any vector space V, the Schur–Weyl representation of S_n on $V^{\otimes n}$ is faithful whenever $\dim V \geq n$, and since we hold n fixed and send $N \to \infty$, we may always safely assume the representation is faithful; i.e. the map $\sigma \mapsto [\sigma]$ may be assumed to be one-to-one.

This representation is dual to the standard representation of $\mathrm{GL}(V)$ on $V^{\otimes n}$ — $G \cdot v_1 \otimes \cdots \otimes v_n = (Gv_1) \otimes \cdots \otimes (Gv_n)$ — in the sense that the two representations commute and are in fact mutual centralizers in $\mathrm{End}(V^{\otimes n})$. This is the so-called *Schur–Weyl duality*, which allows for a lot of useful computations by interchanging the two representations back and forth. The Schur–Weyl duality plays a central role in T. Lévy's approach to the heat kernel on unitary groups, cf. [33]; we will not need much of that detailed approach here, but we are motivated by it in our present constructions.

We may readily verify that, if $\sigma \in S_n$ and has cycle decomposition $\sigma = c_1 \cdots c_r$ with $c_k = (i_1^k \cdots i_{\ell_k}^k)$, then for $A_1, \ldots, A_n \in \mathbb{M}_{2N}$,

$$\mathrm{Tr}_{(\mathbb{M}_{2N})^{\otimes n}}([\sigma] A_1 \otimes \cdots \otimes A_n) = \mathrm{Tr}_{\mathbb{M}_{2N}}(A_{i_{\ell_1}^1} \cdots A_{i_1^1}) \cdots \mathrm{Tr}_{\mathbb{M}_{2N}}(A_{i_{\ell_r}^r} \cdots A_{i_1^r}) \tag{3.12}$$

where we have (in this one instance only) emphasized over which space each trace is taken. Note, in particular, that if $\sigma = (i_1 \cdots i_n)$ is a cycle, then $\mathrm{Tr}([\sigma] A_1 \otimes \cdots \otimes A_n) = \mathrm{Tr}(A_{i_n} \cdots A_{i_1})$ is a single trace. Hence

$$F(s) = \mathbb{E}\mathrm{tr}[(A_s^{2N})^n] = \frac{1}{N} \mathbb{E}\mathrm{Tr}\left([(1 \cdots n)](A_s^{2N})^{\otimes n}\right) \tag{3.13}$$

where $(1 \cdots n)$ denotes the standard full cycle in S_n.

Interchanging the trace and expectation, we can now work with the following simpler tensor-valued function:

$$G(s) = \mathbb{E}[(A_s^{2N})^{\otimes n}]. \tag{3.14}$$

Then we have $F(s) = \frac{1}{N}\mathrm{Tr}\left([(1\cdots n)]G(s)\right)$, and hence to prove F is C^1, it suffices to prove that G is C^1. Computing this derivative is greatly aided by the independence of the processes U^{2N} and B^{2N}: by definition $(A_s^{2N})^{\otimes n} = (U_{t-s}^{2N}B_s^{2N})^{\otimes n} = (U_{t-s}^{2N})^{\otimes n}(B_s^{2N})^{\otimes n}$, and by independence we have

$$G(s) = \mathbb{E}[(U_{t-s}^{2N})^{\otimes n}]\,\mathbb{E}[(B_s^{2N})^{\otimes n}]. \tag{3.15}$$

To concretely express the derivative of G, we introduce a little more notation. First, due to the block structure of the matrix B_s^{2N}, the two projections

$$P_1 = \begin{bmatrix} I_N & 0 \\ 0 & 0 \end{bmatrix}, \qquad P_2 = \begin{bmatrix} 0 & 0 \\ 0 & I_N \end{bmatrix} \tag{3.16}$$

will play a role in all of what follows. Because the diffusion term in the SDE (2.2) governing U_s^{2N} depends linearly on U_s^{2N}, there is mixing that can be expressed as a combination of all transpositions in the Schur–Weyl representation. Hence, we introduce the following notation: for any $A, B \in \mathbb{M}_{2N}$ and $1 \le i < j \le n$, denote by $[A\otimes B]_{i,j}$ the following matrix in $(\mathbb{M}_{2N})^{\otimes n}$:

$$[A\otimes B]_{i,j} = I^{\otimes i-1}\otimes A\otimes I^{\otimes j-i-1}\otimes B\otimes I^{\otimes n-j} \tag{3.17}$$

where $I = I_{2N}$ is the identity matrix in $\mathbb{M}_{2N}$.

With this notation in hand, we can now show that $G \in C^1[0,t]$, and compute that its derivative is given by

$$G'(s) = \frac{1}{2N}G(s)\sum_{1\le i<j\le n}[(i\,j)]\left(I - 2[P_1\otimes P_1]_{i,j} - 2[P_2\otimes P_2]_{i,j}\right) \tag{3.18}$$

where $[(i\,j)]$ denotes the Schur-Weyl representation of the transposition in S_n, cf. (3.11). This is proved in Lemma 3.8 in Section 3.2.2 below. The computation begins by writing down SDEs for $(U_s^{2N})^{\otimes n}$ and$(B_s^{2N})^{\otimes n}$ (following from (2.2)), and then using the product rule in conjunction with (3.15).

Since F is a linear functional of G, this shows that F is also $C^1[0,t]$. The next step in the proof is to plug (3.18) into (3.13) to derive an expression for the derivative $F'(s)$. The derivative $G'(s)$ is a sum of $\binom{n}{2}$ terms, but after contracting them in the trace, the derivative $F'(s)$ can be written as a sum of $n-1$ similar terms:

$$F'(s) = \frac{n}{2}\sum_{p=1}^{n-1}H_{p,n-p}(s)$$

where

$$H_{p,q}(s) = \frac{1}{4N^2}\mathbb{E}[\mathrm{Tr}((A_s^{2N})^p)\mathrm{Tr}((A_s^{2N})^q)] - \frac{1}{2N^2}\sum_{\ell=1}^{2}\mathbb{E}[\mathrm{Tr}((A_s^{2N})^pP_\ell)\mathrm{Tr}((A_s^{2N})^qP_\ell)]. \tag{3.19}$$

This is proved in Lemma 3.9 below. Combining this with (3.10), we therefore have

$$\nu_t^N(n) - \nu_t^{2N}(n) = \frac{n}{2}\sum_{p=1}^{n-1}\int_0^t H_{p,n-p}(s)\,ds. \tag{3.20}$$

It now behooves us to estimate the terms $H_{p,n-p}(s)$, cf. (3.19). These terms have an explicit $\frac{1}{N^2}$ factor, which appears to be good news for the desired estimate of Proposition 3.6. Note, however, that (3.19) involves expectations of products of pairs of *unnormalized* traces of powers of A_s^{2N}. Since A_s^{2N} possesses a large-N limit noncommutative distribution in terms of *normalized* traces, the leading order contribution of the first term in $H_{p,n-p}(s)$ is $O(1)$. In fact, there are many cancellations between the two terms; this is the key observation of the whole proof. The way we will encapsulate the cancellations is by first re-expressing $H_{p,q}(s)$ as a combination of covariances:

$$N^2 H_{p,q}(s) = \frac{1}{4}\mathrm{Cov}[\mathrm{Tr}((A_s^{2N})^p), \mathrm{Tr}((A_s^{2N})^q)] - \mathrm{Cov}[\mathrm{Tr}((A_s^{2N})^p P_1), \mathrm{Tr}((A_s^{2N})^q P_1)]. \tag{3.21}$$

This is proved in Lemma 3.10, and is a fairly straightforward computation: the P_1 and P_2 terms can be combined due to the rotational-invariance of the process A_s^{2N}, and the remaining product terms cancel between the two covariances.

While the two expectation terms in the expression for $N^2 H_{p,q}(s)$ in (3.19) are each $O(N^2)$, once the cancellations between them have been taken into account to produce the covariance expression of (3.21), we find that $N^2 H_{p,q}(s) = O(1)$. In fact, each of the covariance terms in (3.21) is $O(1)$: the precise estimates are

$$\begin{aligned} \left|\mathrm{Cov}[\mathrm{Tr}((A_s^{2N})^p), \mathrm{Tr}((A_s^{2N})^{n-p})]\right| &\le p(n-p)(t+3s), \\ \left|\mathrm{Cov}[\mathrm{Tr}((A_s^{2N})^p P_1), \mathrm{Tr}((A_s^{2N})^{n-p} P_1)]\right| &\le p(n-p)(t+3s) \end{aligned} \tag{3.22}$$

for $0 \le s \le t$. From here, the remainder of the proof is routine: (3.21) shows that $|H_{p,n-p}(s)| \le \frac{5p(n-p)(t+3s)}{4N^2}$; summing $0 \le p \le n$, and integrating $s \in [0,t]$ gives a cubic in n times a linear factor in t times $\frac{1}{N^2}$, and then (3.20) completes the proof. These details are written out more explicitly at the end of Section 3.2.2.

Thus, the proof is completed by proving the covariance estimates (3.22). This is the most technically challenging part of the proof, and requires several involved steps, which we outline here.

1. **Multilinearize the covariances**: as we did in replacing the moment $F(s)$ with the tensor-valued function $G(s)$, we compute that both covariances in (3.21) can be expressed as certain contractions of a common *covariance tensor* $\mathcal{C}_{p,n-p}(s)$, defined by

$$\mathcal{C}_{p,n-p}(s) \equiv \mathbb{E}((A_s^{2N})^{\otimes n}) - \mathbb{E}((A_s^{2N})^{\otimes p}) \otimes \mathbb{E}((A_s^{2N})^{\otimes (n-p)}). \tag{3.23}$$

   The two covariance terms in (3.21) can be expressed in terms of $\mathcal{C}_{p,n-p}(s)$ by tracing against certain p-dependent permutation and projection matrices. This computation is the content of Lemma 3.11.

2. **Explicitly compute the covariance tensor** $\mathcal{C}_{p,n-p}(s)$: using the tensorized SDE for the unitary Brownian motion (following from (2.2), as used in the computation (3.18) of $G'(s)$) and taking expectations, we compute two explicit matrices $\Phi_p, \Psi_p \in (\mathbb{M}_{2N})^{\otimes n}$ for $0 \le p \le n$ such that

$$\mathcal{C}_{p,n-p}(s) = e^{-\frac{nt}{2}}\left[e^{(t-s)\Phi_n}e^{s\Psi_n} - e^{(t-s)\Phi_p}e^{s\Psi_p}\right].$$

   The matrices Φ_p and Ψ_p are given by certain p-restricted sums of transposition matrices $[(i\,j)]$ and the tensorized projection matrices $[P_\ell \otimes P_\ell]_{i,j}$, and are explicitly defined below in (3.36). These computations are done in Lemma 3.12 and Corollary 3.13.

3. **Use Duhamel's formula to produce more cancellations**: for any complex matrices X and Y, Duhamel's formula asserts that

$$e^X - e^Y = \int_0^1 e^{(1-u)X}(X-Y)e^{uY}\,du. \tag{3.24}$$

   Applying this in telescoping fashion to the above expression for $\mathcal{C}_{p,n-p}(s)$ and simplifying yields the expression

$$\mathcal{C}_{p,n-p}(s) = \int_0^{t-s} e^{(t-s-u)\Phi_n}(\Phi_n - \Phi_p)e^{u\Phi_p}e^{s\Psi_n}\,du + \int_0^s e^{(t-s)\Phi_p}e^{(s-u)\Psi_n}(\Psi_n - \Psi_p)e^{u\Psi_p}\,du.$$

This is helpful because $\Phi_n - \Phi_p$ and $\Psi_n - \Psi_p$ yield yet more cancellations. The resulting expression for $\mathcal{C}_{p,n-p}(s)$ is given in Lemma 3.14.

4. **Reinterpret the terms as expectations of products of unitary and projection matrices**: since the matrices Φ_p and Ψ_p generate (via exponential) the expectations of tensor powers of U^{2N} and B^{2N}, it is possible to reinterpret the preceding expression for $\mathcal{C}_{p,n-p}(s)$ as a sum of expected traces of matrices that are tensor products of independent copies of U^{2N} and B^{2N} mixed with certain permutation matrices and projection matrices (given in the definition (3.36) of Φ_p and Ψ_p).

   To be more precise, for $0 \le p \le n$ and i, j with $1 \le i \le p$ and $p < j \le n$, there are random matrices $R^p_{i,j}(u;t,s), Q^p_{i,j}(u;t,s) \in (\mathbb{M}_{2N})^{\otimes n}$ such that

$$\mathcal{C}_{p,n-p}(s) = \frac{1}{2N} \sum_{1 \le i \le p < j \le n} \left(\int_0^{t-s} \mathbb{E}[R^p_{i,j}(u;t,s)]\, du + \int_0^s \mathbb{E}[Q^p_{i,j}(u;t,s)]\, du \right).$$

   The matrices $R^p_{i,j}(u;t,s)$ and $Q^p_{i,j}(u;t,s)$ are defined in (3.45) and (3.46); each is a pure tensor product of matrices built out of independent copies of the processes U^{2N} and B^{2N}, together with permutation and projection matrices. This is proved in Lemma 3.15.

5. **Contract $\mathcal{C}_{p,n-p}(s)$ to yield expressions for the desired covariance terms, and use soft estimates to prove** (3.22): the desired covariance terms are related to $\mathcal{C}_{p,n-p}(s)$ via contraction against certain permutation and projection matrices, as explained in Step 1. Performing these contractions with the integrand expressions in Step 4 yield a sum of terms all of which are given as an expected (unnormalized) trace Tr of a product of unitary and projection matrices. In particular, as each is a contraction, the modulus of the trace is bounded by $2N$ (by the Schatten–Hölder inequality), which cancels the $\frac{1}{2N}$ factor before the sum in the final expression for $\mathcal{C}_{p,n-p}(s)$. Integrating and summing then yields the desired estimates (3.22) to complete the proof. The calculations are performed in Lemmas 3.16 and 3.17.

3.2.2 Proof of Proposition 3.6

We now proceed to fill in the details of the proof outline given in the previous section. We succinctly restate terminology and notation throughout to make the proof more readable. We begin with the computation of the derivative of the tensor-valued function G.

Lemma 3.8. *Fix $t > 0$ and $N, n \in \mathbb{N}$, and let $G\colon [0,t] \to (\mathbb{M}_{2N})^{\otimes n}$ denote the function $G(s) = \mathbb{E}[(A^{2N}_s)^{\otimes n}]$. Then $G \in C^1[0,t]$, and its derivative is given by* (3.18)*:*

$$G'(s) = \frac{1}{2N} G(s) \sum_{1 \le i < j \le n} [(i\,j)] \left(I - 2[P_1 \otimes P_1]_{i,j} - 2[P_2 \otimes P_2]_{i,j} \right).$$

Proof. To begin, note that U^{2N} and B^{2N} are independent, and so following (3.15) we have

$$G(s) = \mathbb{E}[(U^{2N}_{t-s})^{\otimes n}]\mathbb{E}[(B^{2N}_s)^{\otimes n}] \equiv G_1(s)G_2(s), \tag{3.25}$$

where both factors G_1, G_2 are continuous (since they are expectations of polynomials in diffusions). Using the SDE (2.2) and applying Itô's formula to the diffusion B^{2N} shows that there is an L^2-martingale $(M^{2N}_s)_{s \ge 0}$ such that

$$d\left((B^{2N}_s)^{\otimes n}\right) = dM^{2N}_s - \frac{n}{2}(B^{2N}_s)^{\otimes n}\, ds - \frac{1}{N} \sum_{\substack{1 \le i < j \le n \\ 1 \le a,b \le N}} \sum_{\ell=1}^{2} (B^{2N}_s)^{\otimes n} \cdot (E_{a+\ell N, b+\ell N} \otimes E_{b+\ell N, a+\ell N})_{i,j}\, ds$$

where $E_{c,d} \in \mathbb{M}_{2N}$ is the standard matrix unit (all 0 entries except a 1 in entry (c,d)) with indices written modulo $2N$. Using the projection matrices P_1 and P_2 of (3.16) and notation (3.17), we can write this SDE in the form

$$d\left((B_s^{2N})^{\otimes n}\right) = dM_s^{2N} - \frac{n}{2}(B_s^{2N})^{\otimes n}\,ds - \frac{1}{N}\sum_{1\le i<j\le n}\sum_{\ell=1}^{2}(B_s^{2N})^{\otimes n}[(i\,j)][P_\ell\otimes P_\ell]_{i,j}\,ds. \tag{3.26}$$

It follows that $G_2 \in C^1[0,t]$, and

$$G_2'(s) = -\frac{n}{2}G_2(s) - \frac{1}{N}\sum_{1\le i<j\le n}\sum_{\ell=1}^{2}\mathbb{E}\left((B_s^{2N})^{\otimes n}[(i\,j)][P_\ell\otimes P_\ell]_{i,j}\right). \tag{3.27}$$

At the same time, a similar calculation with Itô's formula applied with (2.2) shows that there is an L^2-martingale $(\widetilde{M}_s^{2N})_{s\ge 0}$ such that

$$d((U_s^{2N})^{\otimes n}) = d\widetilde{M}_s^{2N} - \frac{n}{2}(U_s^{2N})^{\otimes n}\,ds - \frac{1}{2N}\sum_{1\le i<j\le n}(U_s^{2N})^{\otimes n}[(i\,j)]\,ds \tag{3.28}$$

which, changing $s\mapsto t-s$, implies that G_1 is $C^1[0,t]$ and

$$G_1'(s) = \frac{n}{2}G_1(s) + \frac{1}{2N}\sum_{1\le i<j\le n}\mathbb{E}\left((U_{t-s}^{2N})^{\otimes n}[(i\,j)]\right). \tag{3.29}$$

Combining (3.27) and (3.29), the product rule $G'(s) = G_1(s)G_2'(s) + G_2(s)G_1'(s)$ shows that $G \in C^1[0,t]$. Using $G = G_1\cdot G_2$ again when recombining, we see that the $\frac{n}{2}$ terms cancel; moreover, the same recombination due to independence yields

$$G'(s) = \frac{1}{2N}\sum_{1\le i<j\le n}\mathbb{E}\left((U_{t-s}^{2N})^{\otimes n}[(i\,j)](B_s^{2N})^{\otimes n}\right) - \frac{1}{N}\sum_{1\le i<j\le n}\sum_{\ell=1}^{2}\mathbb{E}\left((A_s^{2N})^{\otimes n}[(i\,j)][P_\ell\otimes P_\ell]_{i,j}\right).$$

Finally, in the first term, notice that $[(i\,j)](B_s^{2N})^{\otimes n} = (B_s^{2N})^{\otimes n}[(i\,j)]$ (since the Schur-Weyl representation of any permutation commutes with any matrix of the form $B^{\otimes n}$). Hence, we have

$$\begin{aligned}\mathbb{E}\left((U_{t-s}^{2N})^{\otimes n}[(i\,j)](B_s^{2N})^{\otimes n}\right) &= \mathbb{E}\left((U_{t-s}^{2N})^{\otimes n}\right)\mathbb{E}\left([(i\,j)](B_s^{2N})^{\otimes n}\right)\\ &= \mathbb{E}\left((U_{t-s}^{2N})^{\otimes n}\right)\mathbb{E}\left((B_s^{2N})^{\otimes n}[(i\,j)]\right) = G(s)[(i\,j)].\end{aligned}$$

Similarly factoring out the $G(s)$ from the second term yields the result. $\square$

This allows us to compute the derivative of $F(s) = \frac{1}{N}\mathrm{Tr}\left([(1\cdots n)]G(s)\right)$, in terms of the auxiliary functions $H_{p,q}(s)$ defined in (3.19).

Lemma 3.9. *For $0\le s\le t$,*

$$F'(s) = \frac{n}{2}\sum_{p=1}^{n-1}H_{p,n-p}(s) \tag{3.30}$$

where

$$H_{p,q}(s) = \frac{1}{4N^2}\mathbb{E}[\mathrm{Tr}((A_s^{2N})^p)\mathrm{Tr}((A_s^{2N})^q)] - \frac{1}{2N^2}\sum_{\ell=1}^{2}\mathbb{E}[\mathrm{Tr}((A_s^{2N})^pP_\ell)\mathrm{Tr}((A_s^{2N})^qP_\ell)].$$

Proof. Applying (3.18), we have

$$F'(s) = \frac{1}{2N}\mathrm{Tr}\left([(1\cdots n)]G'(s)\right)$$
$$= \frac{1}{4N^2}\sum_{1\le i<j\le n}\left\{\mathrm{Tr}\left([(1\cdots n)]G(s)[(i\,j)]\right) - 2\sum_{\ell=1}^{2}\mathrm{Tr}\left([(1\cdots n)]G(s)[(i\,j)][P_\ell\otimes P_\ell]_{i,j}\right)\right\}.$$

We now use the trace property in the first term to cyclically reorder the matrices; thus we must compute $[(i\,j)][(1\cdots n)] = [(i\,j)(1\cdots n)]$. Noting that $(i\,j)(1\,\cdots\,n) = (1,\cdots,i-1,j,\cdots,n)(i\cdots j-1)$, it follows from (3.12) that

$$\mathrm{Tr}\left([(1,\cdots,i-1,j,\cdots,n)(i\cdots j-1)]G(s)\right) = \mathbb{E}[\mathrm{Tr}((A_s^{2N})^{j-i})\mathrm{Tr}((A_s^{2N})^{n-(j-i)})].$$

A similar calculation shows that

$$\mathrm{Tr}\left([(1\,\cdots\,n)]G(s)[(i\,j)][P_\ell\otimes P_\ell]_{i,j}\right) = \mathbb{E}[\mathrm{Tr}((A_s^{2N})^{j-i}P_\ell)\mathrm{Tr}((A_s^{2N})^{n-(j-i)}P_\ell)].$$

Thus, we have

$$F'(s)$$
$$= \frac{1}{4N^2}\sum_{1\le i<j\le n}\left\{\mathbb{E}[\mathrm{Tr}((A_s^{2N})^{j-i})\mathrm{Tr}((A_s^{2N})^{n-(j-i)})] - 2\sum_{\ell=1}^{2}\mathbb{E}[\mathrm{Tr}((A_s^{2N})^{j-i}P_\ell)\mathrm{Tr}((A_s^{2N})^{n-(j-i)}P_\ell)]\right\}.$$

The terms inside the overall summation depend on (i,j) only through $j-i$; in other words, the overall sum has the form

$$\sum_{1\le i<j\le n} h_{j-i,n-(j-i)}$$

for a function $h\colon \{1,\ldots,n-1\}^2\to\mathbb{C}$ which is symmetric in its two variables. For such a sum in general we have

$$S \equiv \sum_{1\le i<j\le n} h_{j-i,n-(j-i)} = \sum_{p=1}^{n-1}\sum_{\substack{1\le i<j\le n\\ j-i=p}} h_{p,n-p} = \sum_{p=1}^{n-1}(n-p)h_{p,n-p}$$

since the number of (i,j) with $1\le i<j\le n$ and $j-i=p$ is $(n-p)$. Now using the symmetry and reindexing by $q=n-p$ we have

$$2S = \sum_{p=1}^{n-1}(n-p)h_{p,n-p} + \sum_{q=1}^{n-1} q h_{n-q,q} = \sum_{p=1}^{n-1}(n-p)h_{p,n-p} + \sum_{p=1}^{n-1} p h_{p,n-p} = n\sum_{p=1}^{n-1} h_{p,n-p}.$$

Applying this with the above summations yields the result. □

Having now expressed the desired quantities in terms of $H_{p,q}(s)$, we proceed to encapsulate the many cancellations by re-expressing $H_{p,q}(s)$ as a linear combination of covariances.

Lemma 3.10. *For $s\ge 0$ and $p,q\in\mathbb{N}$, $H_{p,q}(s)$ is given by the linear combination of covariances in* (3.21)*:*

$$N^2H_{p,q}(s) = \frac{1}{4}\mathrm{Cov}[\mathrm{Tr}((A_s^{2N})^p),\mathrm{Tr}((A_s^{2N})^q)] - \mathrm{Cov}[\mathrm{Tr}((A_s^{2N})^pP_1),\mathrm{Tr}((A_s^{2N})^qP_1)].$$

Proof. From (3.19), $N^2 H_{p,q}(s)$ is a difference of two terms. The first is

$$\begin{aligned}\frac{1}{4}&\mathbb{E}[\mathrm{Tr}((A_s^{2N})^p)\mathrm{Tr}((A_s^{2N})^q)] \\ &= \frac{1}{4}\mathrm{Cov}[\mathrm{Tr}((A_s^{2N})^p), \mathrm{Tr}((A_s^{2N})^q)] + \frac{1}{4}\mathbb{E}[\mathrm{Tr}((A_s^{2N})^p)]\mathbb{E}[\mathrm{Tr}((A_s^{2N})^q)].\end{aligned} \tag{3.31}$$

The second term is a sum

$$-\frac{1}{2}\sum_{\ell=1}^{2}\mathbb{E}[\mathrm{Tr}((A_s^{2N})^p P_\ell)\mathrm{Tr}((A_s^{2N})^q P_\ell)].$$

Let R be the block rotation matrix of $\mathbb{C}^{2N}$ by $\frac{\pi}{2}$ in each factor of $\mathbb{C}^N$, so that $RP_1R^* = P_2$. Since the distribution of A_s^{2N} is invariant under rotations, it follows that

$$\mathbb{E}[\mathrm{Tr}((A_s^{2N})^p P_\ell)] \quad \text{and} \quad \mathbb{E}[\mathrm{Tr}((A_s^{2N})^p P_\ell)\mathrm{Tr}((A_s^{2N})^q P_\ell)]$$

do not depend on ℓ (as each is a conjugation-invariant polynomial function in A_s^{2N}). In particular, the two terms in the ℓ-sum are equal, and so the second term in $H_{p,q}(s)$ is

$$\begin{aligned}&-\mathbb{E}[\mathrm{Tr}((A_s^{2N})^p P_\ell)\mathrm{Tr}((A_s^{2N})^q P_\ell)] \\ &\qquad = -\mathrm{Cov}[\mathrm{Tr}((A_s^{2N})^p P_1), \mathrm{Tr}((A_s^{2N})^q P_1)] - \mathbb{E}[\mathrm{Tr}((A_s^{2N})^p P_1)]\mathbb{E}[\mathrm{Tr}((A_s^{2N})^q P_1)].\end{aligned} \tag{3.32}$$

Moreover, since $\mathbb{E}[\mathrm{Tr}((A_s^{2N})^p P_1)] = \mathbb{E}[\mathrm{Tr}((A_s^{2N})^p P_2)]$ and $P_1 + P_2 = I$, we have $\mathbb{E}[\mathrm{Tr}((A_s^{2N})^p P_1)] = \frac{1}{2}\mathbb{E}[\mathrm{Tr}((A_s^{2N})^p)]$. Thus, the last term in (3.32) is $-\frac{1}{4}\mathbb{E}[\mathrm{Tr}((A_s^{2N})^p)]\mathbb{E}[\mathrm{Tr}((A_s^{2N})^q)]$. Combining this with (3.31) then yields the result. $\square$

Now, to begin the process of estimating these covariances, we re-express them using the Schur–Weyl representation, in terms of the two-cycle permutation

$$\gamma_{p,q} = (1\ \cdots\ p)(p+1\ \cdots\ p+q) \in S_{p+q}.$$

(For present notational convenience, we de-emphasize the role of n, letting $q = n - p$.)

Lemma 3.11. *Let $\mathcal{C}_{p,q}(s)$ denote the covariance tensor of* (3.23),

$$\mathcal{C}_{p,q}(s) \equiv \mathbb{E}((A_s^{2N})^{\otimes n}) - \mathbb{E}((A_s^{2N})^{\otimes p}) \otimes \mathbb{E}((A_s^{2N})^{\otimes q}).$$

Then the two covariance terms in (3.21) *are given by*

$$\mathrm{Cov}[\mathrm{Tr}((A_s^{2N})^p), \mathrm{Tr}((A_s^{2N})^q)] = \mathrm{Tr}\left(\mathcal{C}_{p,q}(s)[\gamma_{p,q}]\right), \quad \textit{and} \tag{3.33}$$

$$\mathrm{Cov}[\mathrm{Tr}((A_s^{2N})^p P_1), \mathrm{Tr}((A_s^{2N})^q P_1)] = \mathrm{Tr}\left(\mathcal{C}_{p,q}(s)[P_1 \otimes P_1]_{p,p+q}[\gamma_{p,q}]\right). \tag{3.34}$$

Proof. To begin, observe that for any matrix A, (3.12) yields

$$\mathrm{Tr}(A^p)\mathrm{Tr}(A^q) = \mathrm{Tr}\left(A^{\otimes p} \otimes A^{\otimes q}[\gamma_{p,q}]\right).$$

For notational convenience in this proof, denote the full cycle $(1\cdots p) \in S_p$ by γ_p, so that $\mathrm{Tr}(A^p) = \mathrm{Tr}\left(A^{\otimes p}[\gamma_p]\right)$. It follows that

$$\begin{aligned}&\mathrm{Cov}[\mathrm{Tr}((A_s^{2N})^p), \mathrm{Tr}((A_s^{2N})^q)] \\ &\qquad = \mathbb{E}[\mathrm{Tr}((A_s^{2N})^p)\mathrm{Tr}((A_s^{2N})^q)] - \mathbb{E}[\mathrm{Tr}((A_s^{2N})^p)]\mathbb{E}[\mathrm{Tr}((A_s^{2N})^q)] \\ &\qquad = \mathbb{E}\mathrm{Tr}\left((A_s^{2N})^{\otimes p} \otimes (A_s^{2N})^{\otimes q}[\gamma_{p,q}]\right) - \mathbb{E}\mathrm{Tr}\left((A_s^{2N})^{\otimes p}[\gamma_p]\right)\mathbb{E}\mathrm{Tr}\left((A_s^{2N})^{\otimes q}[\gamma_q]\right) \\ &\qquad = \mathrm{Tr}\Big(\mathbb{E}((A_s^{2N})^{\otimes(p+q)})[\gamma_{p,q}]\Big) - \mathrm{Tr}\Big(\mathbb{E}((A_s^{2N})^{\otimes p}) \otimes \mathbb{E}((A_s^{2N})^{\otimes q})[\gamma_{p,q}]\Big) \\ &\qquad = \mathrm{Tr}\left(\Big[\mathbb{E}((A_s^{2N})^{\otimes(p+q)}) - \mathbb{E}((A_s^{2N})^{\otimes p}) \otimes \mathbb{E}((A_s^{2N})^{\otimes q})\Big][\gamma_{p,q}]\right) \\ &\qquad = \mathrm{Tr}\left(\mathcal{C}_{p,q}(s)[\gamma_{p,q}]\right)\end{aligned}$$

thus proving (3.33). At the same time, using the fact that the projection P_1 is diagonal, we have for any matrix $A \in \mathbb{M}_{2N}$

$$\mathrm{Tr}(A^p P_1)\mathrm{Tr}(A^q P_1) = \mathrm{Tr}\left(A^{\otimes p} \otimes A^{\otimes q}[P_1 \otimes P_1]_{p,p+q}[\gamma_{p,q}]\right),$$

where we remind the reader that $[P_\ell \otimes P_\ell]_{i,j}$ references notation (3.17); here $i = p$ and $j = p+q$ (the final tensor factor). The derivation of (3.34) follows from here analogously to the above computations. □

Thus, from (3.21), (3.33), and (3.34), to estimate $H_{p,n-p}(s)$ we must understand the covariance tensor $\mathcal{C}_{p,n-p}(s)$. To that end, we introduce some notation. For $1 \le i < j \le n$, define the matrix $T_{i,j} \in (\mathbb{M}_{2N})^{\otimes n}$ by

$$T_{i,j} = 2[(i\,j)]\,[P_1 \otimes P_1 + P_2 \otimes P_2]_{i,j}\,. \tag{3.35}$$

Additionally, for $1 \le p \le n$, we introduce $\Phi_p, \Psi_p \in (\mathbb{M}_{2N})^{\otimes n}$ as follows:

$$\Phi_p = -\frac{1}{2N}\sum_{\substack{1\le i<j\le p,\text{ or}\\ p<i<j\le n}}[(i\,j)], \quad \Psi_p = -\frac{1}{2N}\sum_{\substack{1\le i<j\le p,\text{ or}\\ p<i<j\le n}} T_{i,j}, \tag{3.36}$$

with the understanding that, when $p = n$, the sum is simply over $1 \le i < j \le n$.

Lemma 3.12. *Let $p \in \{1, \ldots, n\}$, and let $0 \le s \le t$. Then*

$$\mathbb{E}[(U_s^{2N})^{\otimes p}] \otimes \mathbb{E}[(U_s^{2N})^{\otimes(n-p)}] = e^{-\frac{ns}{2}} e^{s\Phi_p}, \tag{3.37}$$

$$\mathbb{E}[(B_s^{2N})^{\otimes p}] \otimes \mathbb{E}[(B_s^{2N})^{\otimes(n-p)}] = e^{-\frac{ns}{2}} e^{s\Psi_p}, \tag{3.38}$$

$$\mathbb{E}[(A_s^{2N})^{\otimes p}] \otimes \mathbb{E}[(A_s^{2N})^{\otimes(n-p)}] = e^{-\frac{nt}{2}} e^{(t-s)\Phi_p} e^{s\Psi_p}, \tag{3.39}$$

where, in the case $p = n$, we interpret the 0-fold tensor product as the identity as usual.

Proof. Returning to the SDEs (3.26) and (3.28), taking expectations we find that

$$\frac{d}{ds}\mathbb{E}[(U_s^{2N})^{\otimes n}] = -\frac{n}{2}\mathbb{E}[(U_s^{2N})^{\otimes n}] - \frac{1}{2N}\mathbb{E}\left\{(U_s^{2N})^{\otimes n} \cdot \sum_{1\le i<j\le n}[(i\,j)]\right\}, \tag{3.40}$$

$$\frac{d}{ds}\mathbb{E}[(B_s^{2N})^{\otimes n}] = -\frac{n}{2}\mathbb{E}[(B_s^{2N})^{\otimes n}] - \frac{1}{2N}\mathbb{E}\left\{(B_s^{2N})^{\otimes n} \cdot \sum_{1\le i<j\le n} T_{i,j}\right\}. \tag{3.41}$$

Since the processes B^{2N} and U^{2N} start at the identity, it is then immediate to verify that the solutions to these ODEs are

$$\mathbb{E}[(U_s^{2N})^{\otimes n}] = e^{-\frac{ns}{2}} \exp\left\{-\frac{s}{2N}\sum_{1\le i<j\le n}[(i\,j)]\right\}, \quad \text{and} \quad \mathbb{E}[(B_s^{2N})^{\otimes n}] = e^{-\frac{ns}{2}} \exp\left\{-\frac{s}{2N}\sum_{1\le i<j\le n} T_{i,j}\right\}.$$

Now, for the tensor product, we decompose

$$\mathbb{E}[(U_s^{2N})^{\otimes p}] \otimes \mathbb{E}[(U_s^{2N})^{\otimes(n-p)}] = \left(\mathbb{E}[(U_s^{2N})^{\otimes p}] \otimes I^{\otimes(n-p)}\right) \cdot \left(I^{\otimes p} \otimes \mathbb{E}[(U_s^{2N})^{\otimes(n-p)}]\right).$$

We can express these expectations as in (3.40), provided we note that $(i\,j)$ now refers (alternatively) to the action of S_p or S_{n-p} on $(\mathbb{M}_{2N})^{\otimes n} = (\mathbb{M}_{2N})^{\otimes p} \otimes (\mathbb{M}_{2N})^{\otimes(n-p)}$, either trivially in the first factor or the second. The result is that

$$\mathbb{E}[(U_s^{2N})^{\otimes p}] \otimes I^{\otimes(n-p)} = e^{-\frac{ps}{2}} \exp\left\{-\frac{s}{2N} \sum_{1\le i<j\le p} [(i\,j)]\right\}$$

$$I^{\otimes p} \otimes \mathbb{E}[(U_s^{2N})^{\otimes(n-p)}] = e^{-\frac{(n-p)s}{2}} \exp\left\{-\frac{s}{2N} \sum_{p<i'<j'\le n} [(i'\,j')]\right\}.$$

Note that all the $(i\,j)$ terms in the first sum commute with all the $(i'\,j')$ terms in the second sum (since $i<j<i'<j'$) and taking the product, we can combine to yield

$$\mathbb{E}[(U_s^{2N})^{\otimes p}] \otimes \mathbb{E}[(U_s^{2N})^{\otimes(n-p)}] = e^{-\frac{ns}{2}} \exp\left\{-\frac{s}{2N} \sum_{\substack{1\le i<j\le p, \text{ or}\\ p<i<j\le n}} [(i\,j)]\right\} = e^{-\frac{ns}{2}} e^{s\Phi_p},$$

verifying (3.37). An entirely analogous analysis proves (3.38).

Finally, using independence as in (3.25) to factor

$$\begin{aligned}&\mathbb{E}[(A_s^{2N})^{\otimes p}] \otimes \mathbb{E}[(A_s^{2N})^{\otimes(n-p)}] \\ &\qquad = \left(\mathbb{E}[(U_{t-s}^{2N})^{\otimes p}] \otimes \mathbb{E}[(U_{t-s}^{2N})^{\otimes(n-p)}]\right) \cdot \left(\mathbb{E}[(B_s^{2N})^{\otimes p}] \otimes \mathbb{E}[(B_s^{2N})^{\otimes(n-p)}]\right)\end{aligned} \tag{3.42}$$

and substituting $s\mapsto t-s$ in (3.37), (3.39) follows from (3.38) and (3.42). □

Corollary 3.13. *For $0\le p\le n$ and $0\le s\le t$,*

$$\mathcal{C}_{p,n-p}(s) = e^{-\frac{nt}{2}} \left[e^{(t-s)\Phi_n} e^{s\Psi_n} - e^{(t-s)\Phi_p} e^{s\Psi_p}\right]. \tag{3.43}$$

Proof. From (3.39), we have

$$\mathbb{E}[(A_s^{2N})^{\otimes p}] \otimes \mathbb{E}[(A_s^{2N})^{\otimes(n-p)}] = e^{-\frac{nt}{2}} e^{(t-s)\Phi_p} e^{s\Psi_p}$$

for $0\le p\le n$. In particular, taking $p=n$ (in which case the second factor is trivial and omitted) we have

$$\mathbb{E}[(A_s^{2N})^{\otimes n}] = e^{-\frac{nt}{2}} e^{(t-s)\Phi_n} e^{s\Psi_n}.$$

Subtracting these two and using the definition (3.23) of $\mathcal{C}_{p,n-p}(s)$ yields the result. □

The next technical lemma uses Duhamel's formula to rewrite the expression in (3.43) in a more complicated, but more computationally useful, way.

Lemma 3.14. *For $0\le s\le t$,*

$$\mathcal{C}_{p,n-p}(s) = \frac{e^{-\frac{nt}{2}}}{2N} \sum_{1\le i\le p<j\le n} \left(\int_0^{t-s} e^{(t-s-u)\Phi_n} [(i\,j)] e^{u\Phi_p} e^{s\Psi_n}\,du + \int_0^s e^{(t-s)\Phi_p} e^{(s-u)\Psi_n} T_{i,j} e^{u\Psi_p}\,du\right).$$

Proof. We begin by using Duhamel's formula (3.24), applied to (3.43) by adding and subtracting the mixed term $e^{(t-s)\Phi_p}e^{s\Psi_n}$:

$$\begin{aligned}
&e^{(t-s)\Phi_n}e^{s\Psi_n} - e^{(t-s)\Phi_p}e^{s\Psi_p} \\
=&\Big[e^{(t-s)\Phi_n} - e^{(t-s)\Phi_p}\Big]e^{s\Psi_n} + e^{(t-s)\Phi_p}\Big[e^{s\Psi_n} - e^{s\Psi_p}\Big] \\
=&\int_0^1 e^{(1-u)(t-s)\Phi_n}(t-s)(\Phi_n-\Phi_p)e^{u(t-s)\Phi_p}\,du\cdot e^{s\Psi_n} + e^{(t-s)\Phi_p}\int_0^1 e^{(1-u)s\Psi_n}s(\Psi_n-\Psi_p)e^{us\Psi_p}\,du \\
=&\int_0^{t-s} e^{(t-s-u)\Phi_n}(\Phi_n-\Phi_p)e^{u\Phi_p}e^{s\Psi_n}\,du + \int_0^s e^{(t-s)\Phi_p}e^{(s-u)\Psi_n}(\Psi_n-\Psi_p)e^{u\Psi_p}\,du
\end{aligned} \tag{3.44}$$

where we have made the substitution $u \mapsto (t-s)u$ in the first integral and $u \mapsto su$ in the second. Now, from the definition (3.36) of Φ_p and Ψ_p, we have

$$\Phi_p - \Phi_n = \frac{1}{2N}\sum_{1\le i\le p<j\le n}[(i\,j)], \quad \text{and} \quad \Psi_p - \Psi_n = \frac{1}{2N}\sum_{1\le i\le p<j\le n} T_{i,j}.$$

Substituting these into (3.44) yields the result. □

We now reinterpret the exponentials in the integrals as expectations of the processes U^{2N} and B^{2N}, using (3.37) and (3.38). The first integrand is

$$e^{-\frac{nt}{2}}e^{(t-s-u)\Phi_n}[(i\,j)]e^{u\Phi_p}e^{s\Psi_n} = \mathbb{E}[(U^{2N}_{t-s-u})^{\otimes n}]\cdot[(i\,j)]\cdot\mathbb{E}[(U^{2N}_u)^{\otimes p}]\otimes\mathbb{E}[(U^{2N}_u)^{\otimes(n-p)}]\cdot\mathbb{E}[(B^{2N}_s)^{\otimes n}].$$

By definition U^{2N} and B^{2N} are independent. Let us introduce two more copies V^{2N}, W^{2N} of U^{2N} so the $\{U^{2N}, V^{2N}, W^{2N}, B^{2N}\}$ are all independent. Then this product may be expressed as the expectation of

$$R^p_{i,j}(u;s,t) \equiv (U^{2N}_{t-s-u})^{\otimes n}\cdot[(i\,j)]\cdot(V^{2N}_u)^{\otimes p}\otimes(W^{2N}_u)^{\otimes(n-p)}\cdot(B^{2N}_s)^{\otimes n}. \tag{3.45}$$

Similarly, if we introduce independent copies C^{2N} and D^{2N} of B^{2N} so that all the the processes U^{2N}, V^{2N}, W^{2N}, B^{2N}, C^{2N}, D^{2N} are independent, the second integrand can be expressed as the expectation of

$$Q^p_{i,j}(u;s,t) \equiv (V^{2N}_{t-s})^{\otimes p}\otimes(W^{2N}_{t-s})^{\otimes(n-p)}\cdot(B^{2N}_{s-u})^{\otimes n}\cdot T_{i,j}\cdot(C^{2N}_u)^{\otimes p}\otimes(D^{2N}_u)^{\otimes(n-p)}. \tag{3.46}$$

To summarize, we have computed the following.

Lemma 3.15. *Let $0\le s\le t$ and $p\in\{1,\dots,n\}$. Then*

$$\mathcal{C}_{p,n-p}(s) = \frac{1}{2N}\sum_{1\le i\le p<j\le n}\left(\int_0^{t-s}\mathbb{E}[R^p_{i,j}(u;t,s)]\,du + \int_0^s \mathbb{E}[Q^p_{i,j}(u;t,s)]\,du\right). \tag{3.47}$$

We will now use this, together with (3.21), (3.33), and (3.34), to estimate $H_{p,n-p}(s)$. We estimate each of the two covariance terms separately, in the following two lemmas.

Lemma 3.16. *For $p\in\{1,\dots,n\}$ and $0\le s\le t$,*

$$\left|\mathrm{Cov}[\mathrm{Tr}((A^{2N}_s)^p),\mathrm{Tr}((A^{2N}_s)^{n-p})]\right| \le p(n-p)(t+3s). \tag{3.48}$$

Proof. From (3.47) together with (3.33), the quantity whose modulus we wish to estimate is

$$\sum_{1\le i\le p<j\le n}\left(\int_0^{t-s}\frac{1}{2N}\mathbb{E}\mathrm{Tr}\left(R^p_{i,j}(u;t,s)\cdot[\gamma_{p,n-p}]\right)\,du + \int_0^s\frac{1}{2N}\mathbb{E}\mathrm{Tr}\left(Q^p_{i,j}(u;t,s)\cdot[\gamma_{p,n-p}]\right)\,du\right).$$

For the first term, we note $(U^{2N}_{t-s-u})^{\otimes n}[\gamma_{p,n-p}] = [\gamma_{p,n-p}](U^{2N}_{t-s-u})^{\otimes n}$ (the Schur-Weyl representation of any permutation in S_n commutes with a matrix of the form $M^{\otimes n}$). It follows from the trace property that

$$\begin{aligned}\mathbb{E}\mathrm{Tr}\left(R^p_{i,j}(u;s,t)\cdot[\gamma_{p,n-p}]\right) &= \mathbb{E}\mathrm{Tr}\left((V^{2N}_u)^{\otimes p}\otimes(W^{2N}_u)^{\otimes(n-p)}\cdot(B^{2N}_s)^{\otimes n}\cdot[\gamma_{p,n-p}]\cdot(U^{2N}_{t-s-u})^{\otimes n}\cdot[(i\,j)]\right)\\ &= \mathbb{E}\mathrm{Tr}\left((V^{2N}_u)^{\otimes p}\otimes(W^{2N}_u)^{\otimes(n-p)}\cdot(B^{2N}_s)^{\otimes n}\cdot(U^{2N}_{t-s-u})^{\otimes n}\cdot[\gamma_{p,n-p}]\cdot[(i\,j)]\right).\end{aligned}$$

Since $i \le p < j$, the permutation $\gamma_{p,n-p}(i\,j)$ is a single cycle. Thus, by (3.12), the $\otimes n$-fold trace reduces to a trace of some p,n,i,j-dependent word in V^{2N}_u, W^{2N}_u, B^{2N}_s, and U^{2N}_{t-s-u}. This word is a random element of $\mathbb{U}_{2N}$, and hence

$$\frac{1}{2N}\mathbb{E}\mathrm{Tr}\left(R^p_{i,j}(u;s,t)\cdot[\gamma_{p,n-p}]\right) = \frac{1}{2N}\mathbb{E}\mathrm{Tr}\Big(\text{a random matrix in } \mathbb{U}_{2N}\Big) \text{ which} \therefore \text{has modulus } \le 1.$$

Hence, the first integral is

$$\left|\int_0^{t-s}\frac{1}{2N}\mathbb{E}\mathrm{Tr}\left(R^p_{i,j}(u;t,s)\cdot[\gamma_{p,n-p}]\right)\,du\right| \le (t-s). \tag{3.49}$$

For the second term, the fact that $T_{i,j} = 2[(i\,j)][P_1\otimes P_1 + P_2\otimes P_2]_{i,j}$ only acts non-trivially in the i,j factors, and $i \le p < j$, shows that (as above) $T_{i,j}$ commutes with $(C^{2N}_u)^{\otimes p}\otimes(D^{2N}_u)^{\otimes(n-p)}$. Hence, we can express the second integrand as

$$\begin{aligned}&\mathbb{E}\mathrm{Tr}\left(Q^p_{i,j}(u;t,s)\cdot[\gamma_{p,n-p}]\right)\\ =&\mathbb{E}\mathrm{Tr}\left((V^{2N}_{t-s})^{\otimes p}\otimes(W^{2N}_{t-s})^{\otimes(n-p)}\cdot(B^{2N}_{s-u})^{\otimes n}\cdot T_{i,j}\cdot(C^{2N}_u)^{\otimes p}\otimes(D^{2N}_u)^{\otimes(n-p)}\cdot[\gamma_{p,n-p}]\right)\\ =&\mathbb{E}\mathrm{Tr}\left((V^{2N}_{t-s})^{\otimes p}\otimes(W^{2N}_{t-s})^{\otimes(n-p)}\cdot(B^{2N}_{s-u})^{\otimes n}\cdot T_{i,j}\cdot[\gamma_{p,n-p}]\cdot(C^{2N}_u)^{\otimes p}\otimes(D^{2N}_u)^{\otimes(n-p)}\right)\\ =&2\sum_{\ell=1}^{2}\mathbb{E}\mathrm{Tr}\left((C^{2N}_u)^{\otimes p}\otimes(D^{2N}_u)^{\otimes(n-p)}\cdot(V^{2N}_{t-s})^{\otimes p}\otimes(W^{2N}_{t-s})^{\otimes(n-p)}\cdot(B^{2N}_{s-u})^{\otimes n}\cdot[P_\ell\otimes P_\ell]_{i,j}[(i\,j)]\cdot[\gamma_{p,n-p}]\right)\end{aligned} \tag{3.50}$$

where we have used the fact that $[\gamma_{p,n-p}]$ commutes with any matrix of the form $C^{\otimes p}\otimes D^{\otimes(n-p)}$ in the second equality, and then the trace property in the third equality. As above, each of these terms reduces to a trace of a word, this time of the form

$$2\sum_{\ell=1}^{2}\mathbb{E}\mathrm{Tr}(\mathcal{U}P_\ell\mathcal{V}P_\ell)$$

where $\mathcal{U}$ and $\mathcal{V}$ are random matrices in $\mathbb{U}_{2N}$ (depending on p,n,i,j). Since $\|P_\ell\| \le 1$, the modulus of each term is $\le 2N$, giving an overall factor of $\le 8N$. Combining with the $\frac{1}{2N}$ in the integral, this gives

$$\left|\int_0^{s}\frac{1}{2N}\mathbb{E}\mathrm{Tr}\left(Q^p_{i,j}(u;t,s)\cdot[\gamma_{p,n-p}]\right)\,du\right| \le 4s. \tag{3.51}$$

Hence, from (3.49) and (3.51), the modulus of the desired covariance is bounded by

$$\sum_{1\le i\le p<j\le n}[(t-s)+2s] = p(n-p)(t+3s),$$

yielding (3.48). $\square$

Lemma 3.17. *For $p \in \{1, \ldots, n\}$ and $0 \le s \le t$,*

$$\left|\mathrm{Cov}[\mathrm{Tr}((A_s^{2N})^p P_1), \mathrm{Tr}((A_s^{2N})^{n-p} P_1)]\right| \le p(n-p)\,(t+3s)\,. \tag{3.52}$$

Proof. From (3.47) together with (3.34), the quantity whose modulus we wish to estimate is

$$\sum_{1\le i\le p<j\le n} \left(\int_0^{t-s} \frac{1}{2N}\mathbb{E}\mathrm{Tr}\left(R^p_{i,j}(u;t,s)\cdot[P_1\otimes P_1]_{p,n}\cdot[\gamma_{p,n-p}]\right)\,du \right. \\ \left. + \int_0^s \frac{1}{2N}\mathbb{E}\mathrm{Tr}\left(Q^p_{i,j}(u;t,s)\cdot[P_1\otimes P_1]_{p,n}\cdot[\gamma_{p,n-p}]\right)\,du\right). \tag{3.53}$$

For the first term, we expand the integrand, commuting $[(i\,j)]$ past $(U^{2N}_{t-s-u})^{\otimes n}$ as in the proof of Lemma 3.16, to give

$$\frac{1}{2N}\mathbb{E}\mathrm{Tr}\left((U^{2N}_{t-s-u})^{\otimes n}\cdot(V^{2N}_u)^{\otimes p}\otimes(W^{2N}_u)^{\otimes(n-p)}\cdot(B^{2N}_s)^{\otimes n}\cdot[P_1\otimes P_1]_{p,n}\cdot[\gamma_{p,n-p}]\cdot[(i\,j)]\right).$$

As above, since $\gamma_{p,n-p}\cdot(i,j)$ is a single cycle, this trace reduces to a trace over $\mathbb{C}^{2N}$, of the form

$$\frac{1}{2N}\mathbb{E}\mathrm{Tr}[\mathcal{U}'P_1\mathcal{V}'P_1]$$

where $\mathcal{U}'$ and $\mathcal{V}'$ are random unitary matrices in $\mathbb{U}_{2N}$ composed of certain i,j,p,n-dependent words in U^{2N}_{t-s-u}, V^{2N}_u, W^{2N}_u, and B^{2N}_s. As $\|P_1\|\le 1$, it follows that this normalized trace is ≤ 1, and so the first integral in (3.53) is $\le (t-s)$ in modulus, as in (3.49).

Similarly, we expand the second term as in (3.50), which gives the sum over $\ell\in\{1,2\}$ of the expected normalized trace of

$$(V^{2N}_{t-s})^{\otimes p}\otimes(W^{2N}_{t-s})^{\otimes(n-p)}\cdot(B^{2N}_{s-u})^{\otimes n}\cdot[\gamma_{p,n-p}]\cdot[(i,j)]\cdot[P_\ell\otimes P_\ell]_{i,j}\cdot(C^{2N}_u)^{\otimes p}\otimes(D^{2N}_u)^{\otimes(n-p)}\cdot[P_1\otimes P_1]_{p,n}.$$

As in the above cases, since $\gamma_{p,n-p}(i,j)$ is a single cycle, this is equal to a single trace $\mathrm{Tr}(A_1\cdots A_n)$ where $A_1,\ldots,A_n$ belong to the set $\{V^{2N}_{t-s}, W^{2N}_{t-s}, B^{2N}_{s-u}, C^{2N}_u, D^{2N}_u, P_\ell, P_1\}$. As each of these is either a random unitary matrix or a projection, it follows that the expected normalized trace has modulus ≤ 1, and so the $\frac{1}{N}$-weighted sum of 2 terms, each of modulus $\le 2N$, gives a contribution no bigger than 4. The remainder of the proof is exactly as the end of the proof of Lemma 3.16. ☐

Finally, we are ready to conclude this section.

Proof of Proposition 3.6. From (3.20), we have

$$|\nu^N_t(n)-\nu^{2N}_t(n)| \le \frac{n}{2}\sum_{p=1}^{n-1}\int_0^t |H_{p,n-p}(s)|\,ds.$$

Lemma 3.10 then gives

$$|H_{p,q}(s)| \le \frac{1}{4N^2}\left|\mathrm{Cov}[\mathrm{Tr}((A^{2N}_s)^p), \mathrm{Tr}((A^{2N}_s)^q)]\right| + \frac{1}{N^2}\left|\mathrm{Cov}[\mathrm{Tr}((A^{2N}_s)^p P_1), \mathrm{Tr}((A^{2N}_s)^q P_1)]\right|.$$

Combining this with (3.48) and (3.52) therefore yields

$$|H_{p,q}(s)| \le \frac{p(n-p)}{N^2}\cdot\frac{5}{4}(t+3s).$$

Integration then gives

$$|\nu_t^N(n) - \nu_t^{2N}(n)| \le \frac{n}{2}\sum_{p=1}^{n-1}\frac{p(n-p)}{N^2}\cdot\frac{25}{8}t^2 = \frac{25t^2n}{16N^2}\sum_{p=1}^{n-1}p(n-p).$$

The sum over p has the exact value $\frac{1}{6}(n^3-n) \le \frac{n^3}{6}$. The blunt estimate $\frac{25}{96} < \frac{3}{4}$ yields the result. $\square$

4 Strong Convergence

In this section, we prove Theorem 1.4. We begin by showing (in Section 4.1) that the eigenvalues of the unitary Brownian motion at a fixed time converge to their "classical locations", and we use this to prove that the unitary Brownian motion can be uniformly approximated by a function of a Gaussian unitary ensemble (for time $t < 4$). We then use this, together with Male's Theorem 2.3, to prove Theorem 1.4.

4.1 Marginals of Unitary Brownian Motion and Approximation by GUE^N

We begin with the following general result on convergence of empirical measures. As usual, for a probability measure μ on $\mathbb{R}$, the cumulative distribution function F_μ is the nondecreasing function $F_\mu(x) = \mu((-\infty, x])$ and its right-inverse is the left-continuous nondecreasing function

$$F_\mu^{-1}(y) = \inf\{x \in \mathbb{R} : y \le F_\mu(x)\},$$

with $F_\mu^{-1}(0) = -\infty$. If μ has a density ρ, we may abuse notation and write $F_\mu = F_\rho$.

Proposition 4.1. *For each $N \in \mathbb{N}$, let $(x_k^N)_{k=1}^N$ be points in $\mathbb{R}$ with $x_1^N \le \cdots \le x_N^N$. Let $\mu^N = \frac{1}{N}\sum_{k=1}^N \delta_{x_k^N}$ be the associated empirical measures. Suppose the following hold true.*

(1) There is a compact interval $[a_-, a_+]$ and a continuous probability density ρ with $\operatorname{supp}\rho = [a_-, a_+]$ so that, with $\mu(dx) = \rho(x)\,dx$, we have $\mu^N \rightharpoonup \mu$ weakly as $N \to \infty$.

(2) $x_1^N \to a_-$ and $x_N^N \to a_+$ as $N \to \infty$.

For $r \in [0,1]$, define $x^*(r) = F_\mu^{-1}(r)$ if $r \in (0,1)$, and $x^*(0) = a_-$, $x^*(1) = a_+$. Then, as $N \to \infty$,

$$\max_{1\le k\le N}\left|x_k^N - x^*(\tfrac{k}{N})\right| \to 0.$$

Proof. For any N, the right-inverse of F_{μ_N} satisfies

$$F_{\mu_N}^{-1}(r) = x_k^N \quad\text{and}\quad F_{\mu_N}(F_{\mu_N}^{-1}(r)) = \frac{k}{N}, \tag{4.1}$$

for any $k \in \{1,\ldots,N\}$ and $r \in (\frac{k-1}{N}, \frac{k}{N}]$. Since μ^N converges weakly towards μ and F_μ is continuous, the sequence F_{μ^N} converges pointwise towards F_μ. What is more, using that μ is compactly supported, a variant of Dini's theorem (cf. [39, Problem 127 Chapter II]) implies further that $F_{\mu^N} \to F$ uniformly on $\mathbb{R}$. Let us consider $\Psi : [0,1] \to [a_-, a_+]$, the continuous inverse of the one-to-one map $F_\mu : [a_-, a_+] \to [0,1]$. It satisfies $\Psi(r) = F_\mu^{-1}(r)$, for any $r \in (0,1]$ and $\Psi(F_\mu(x)) = \max(a_-, \min(a_+, x))$, for any $x \in \mathbb{R}$. Combining this latter equality with (4.1), and using the uniform continuity of Ψ and the uniform convergence of F_{μ_N} towards F_μ, implies that $\max(a_-, \min(a_+, F_{\mu_N}^{-1}))$ converges uniformly towards Ψ on $(0,1]$. The first equality of (4.1) together with the assumption (2) yield then the claim.

$\square$

For example, if $(x_k^N)_{k=1}^N$ are the ordered eigenvalues of a GUE^N, then Wigner's law (and the corresponding spectral edge theorem) show that the empirical spectral distribution satisfies the conditions of Proposition 4.1 almost surely, where the limit measure is the semicircle law $\mu = \sigma_1 \equiv \frac{1}{2\pi}\sqrt{(4-x^2)_+}\,dx$. In particular, when $\frac{k(N)}{N} \to r$, we have $x_{k(N)}^N \to F_{\sigma_1}^{-1}(r)$. These values are sometimes called the *classical locations* of the eigenvalues. In the case of a GUE^N, much more is known; for example, [23] showed that the eigenvalues have variance of $O(\frac{\log N}{N^2})$ in the bulk and $O(N^{-4/3})$ at the edge, and further standardizing them, their limit distribution is Gaussian in the bulk and Tracy–Widom at the edge. For our purposes, the macroscopic statement of Proposition 4.1 will suffice.

Now, fix $t \in [0,4)$. From Theorem 2.5, the law ν_t of the free unitary Brownian motion u_t has an analytic density ϱ_t supported on a closed arc strictly contained in $\mathbb{U}_1$, and has the form $\varrho_t(e^{ix}) = \rho_t(x)$ for some strictly positive continuous probability density function $\rho_t\colon(-\pi,\pi)\to\mathbb{R}$ which is symmetric about 0 and supported in a symmetric interval $[-a(t),a(t)]$ where $a(t) = \frac{1}{2}\sqrt{t(4-t)} + \arccos(1-t/2)$; cf. (2.7). For $0<r<1$, define that *classical locations* $\upsilon^*(t,r)$ of the eigenvalues of unitary Brownian motion as follows:

$$\upsilon^*(t,r) = \exp\left(iF_{\rho_t}^{-1}(r)\right),$$

and also set $\upsilon^*(t,0) = e^{-ia(t)}$ and $\upsilon^*(t,1) = e^{ia(t)}$.

Corollary 4.2. *Let $0 \le t < 4$, and let V_t^N be a random unitary matrix distributed according to the heat kernel on $\mathbb{U}_N$ at time t (i.e. equal in distribution to the t-marginal of the unitary Brownian motion U_t^N). Enumerate the eigenvalues of V_t^N as $\upsilon_1^N(t),\ldots,\upsilon_N^N(t)$, in increasing order of complex argument in $(-\pi,\pi)$. Then,*

$$\lim_{N\to\infty} |\upsilon_k^N(t) - \upsilon^*(t,\tfrac{k}{N})| = 0 \ \ a.s.$$

Proof. Let $x_k^N(t) = -i\log \upsilon_k^N(t)$, where we use the branch of the logarithm cut along the negative real axis. Note: by Theorem 1.1, for sufficiently large N, $\upsilon_k^N(t)$ are outside a t-dependent neighborhood of -1, and so the log function is continuous. The empirical law of $\{\upsilon_k^N(t)\colon 1\le k\le N<\infty\}$ converges weakly a.s. to ν_t (cf. (1.2)), and so by continuity, the empirical measure of $\{x_k^N(t)\colon 1\le k\le N<\infty\}$ converges a.s. to the measure with density ρ_t. Moreover, Theorem 1.1 shows that $\upsilon_1^N(t)\to e^{-ia(t)}$ and $\upsilon_N^N(t)\to e^{ia(t)}$ a.s., and so $x_1^N(t)\to -a(t)$ while $x_N^N(t)\to a(t)$ a.s. Hence, by Proposition 4.1, $\max_{1\le k\le N}|x_k^N(t) - F_{\rho_t}^{-1}(\frac{k}{N})| \to 0$. Taking $\exp(i\cdot)$ of these statements yields the corollary. □

Now, let us combine this result with the comparable one for the GUE^N. Let $g_t\colon\mathbb{R}\to\mathbb{R}$ be given by $g_t = F_{\rho_t}^{-1}\circ F_{\sigma_1}$; this is an increasing, continuous map that pushes σ_1 forward to ρ_t. Define $f_t\colon\mathbb{R}\to\mathbb{U}_1$ by

$$f_t = \exp(ig_t), \qquad \therefore\ \nu_t = (f_t)_*(\sigma_1). \tag{4.2}$$

The main result of this section is that, rather than just pushing the semicircle law forward to the law of free unitary Brownian motion, g_t in fact pushes a GUE^N forward, asymptotically, to V_t^N (for fixed $t\in[0,4)$).

Proposition 4.3. *Let $0\le t<4$, and let V_t^N be a random unitary matrix distributed according to the heat kernel on $\mathbb{U}_N$ at time t (i.e. equal in distribution to the t-marginal of the unitary Brownian motion U_t^N). There exists a self-adjoint random matrix X^N with the following properties:*

(1) X^N is a GUE^N.

(2) The eigenvalues of X^N are independent from V_t^N, and $\{V_t^N, X^N\}$ have the same eigenvectors.

(3) $\|f_t(X^N) - V_t^N\|_{\mathbb{M}_N} \to 0$ a.s. as $N\to\infty$.

Proof. Let $(\upsilon_k^N(t))_{k=1}^N$ denote the eigenvalues of V_t^N in order of increasing argument in $(-\pi,\pi)$, as in Corollary 4.2. It is almost surely true that $\arg(\upsilon_1^N(t)) < \cdots < \arg(\upsilon_N^N(t))$, and so we work in this event only. Let $\mathscr{E}_k^N$ denote the eigenspace of the eigenvalue $\upsilon_k^N(t)$. This space has complex dimension 1 a.s., and so we may select a unit length vector E_k^N from this space, with phase chosen uniformly at random in $\mathbb{U}_1$, independently for each of $E_1^N,\ldots,E_N^N$. Then, by orthogonality of distinct eigenspaces, the random matrix $E^N = [\, E_1^N \cdots E_N^N \,]$ is in $\mathbb{U}_N$; what's more, since the distribution of V_t^N is invariant under conjugation by unitaries, we may further assume that E^N is Haar distributed on $\mathbb{U}_N$. Now, for each N, fix a random vector $\boldsymbol{\mu}^N = (\mu_1^N,\ldots,\mu_N^N)$ independent from V_t^N with joint density $f_{\boldsymbol{\mu}^N}(x_1,\ldots,x_N)$ equal to the known joint density of eigenvalues of a GUE^N, i.e. proportional to

$$f_{\boldsymbol{\mu}^N}(x_1,\ldots,x_n) \sim \prod_{j=1}^N e^{-\frac{N}{2}x_j^2} \prod_{1\le j<k\le N} |x_j - x_k|^2.$$

Then we define

$$X^N \equiv E^N \begin{bmatrix} \mu_1^N & 0 & \cdots & 0 \\ 0 & \mu_2^N & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mu_N^N \end{bmatrix} (E^N)^*.$$

It is well known (cf. [1, 37]) that the distribution of X^N is the GUE^N, verifying item (1). Item (2) holds by construction of X^N. It remains to see that (3) holds true. Note, since operator norm is invariant under unitary conjugation, we simply have

$$\|f_t(X^N) - V_t^N\|_{\mathbb{M}_N} = \max_{1\le k\le N} |f_t(\mu_k^N) - \upsilon_k^N|. \tag{4.3}$$

But for any $k\in\{1,\ldots,N\}$, $\upsilon^*(t,\frac{k}{N}) = f_t(\mu_k^N)$ and Corollary 4.2 yields the result. □

4.2 Strong Convergence of the Process $(U_t^N)_{t\ge 0}$

Since the Gaussian unitary ensemble is selfadjoint, we may extend Male's Theorem 2.3 to continuous functions in independent GUE^Ns.

Lemma 4.4. *Let $\mathbf{A}^N = (A_1^N,\ldots,A_n^N)$ be a collection of random matrix ensembles that converges strongly to some $\mathbf{a} = (a_1,\ldots,a_n)$ in a W^*-probability space $(\mathscr{A},\tau)$. Let $\mathbf{X}^N = (X_1^N,\ldots,X_k^N)$ be independent Gaussian unitary ensembles independent from $\mathbf{A}^N$, and let $\mathbf{x} = (x_1,\ldots,x_k)$ be freely independent semicircular random variables in $\mathscr{A}$ all free from $\mathbf{a}$. Let $\mathbf{f} = (f_1,\ldots,f_k)\colon \mathbb{R}\to\mathbb{C}^k$ be continuous functions, and let $\mathbf{f}(\mathbf{X}^N) = (f_1(X_1^N),\ldots,f_k(X_k^N))$ and $\mathbf{f}(\mathbf{x}) = (f_1(x_1),\ldots,f_k(x_k))$. Then $(\mathbf{A}^N,\mathbf{f}(\mathbf{X}^N))$ converges strongly to $(\mathbf{a},\mathbf{f}(\mathbf{x}))$.*

Proof. We begin with the case $k=1$. If p is any polynomial, by Theorem 2.3, $(\mathbf{A}^N, p(X_1^N))$ converges strongly to $(\mathbf{a}, p(x_1))$ by definition. Now, let $\epsilon>0$, and fix a noncommutative polynomial P in $n+1$ indeterminates. Then $P(\mathbf{a},y)$ is a finite sum of monomials, each of the form

$$Q_0(\mathbf{a})yQ_1(\mathbf{a})y\cdots Q_{d-1}(\mathbf{a})yQ_d(\mathbf{a})$$

for some noncommutative polynomials $Q_0,\ldots,Q_d$ and nonnegative integers d. Let d_P be the "degree" of P: the maximum number of $Q_k(\mathbf{a})$ terms that appears in any monomial in the above expansion of $P(\mathbf{a},y)$. Let $M = 1+$ the sum of all the products $\|Q_0(\mathbf{a})\|\cdots\|Q_d(\mathbf{a})\|$ over all monomial terms appearing in P. Given a fixed $\kappa>2$, by the Weierstrass approximation theorem, for any $\epsilon>0$, there is a polynomial p in one indeterminate so that

$$\|p - f_1\|_{L^\infty[-\kappa,\kappa]} < \frac{\epsilon}{8d_P M(1+\|f_1\|_{L^\infty[-\kappa,\kappa]})^{d_P}}. \tag{4.4}$$

It follows that, for small enough $\epsilon > 0$, we also have $\|p\|_{L^\infty[-\kappa,\kappa]} \le 1 + \|f_1\|_{L^\infty[-\kappa,\kappa]}$. Now we break up the difference in the usual manner,

$$\begin{aligned}\big|\|P(\mathbf{A}^N, f_1(X_1^N))\| - \|P(\mathbf{a}, f_1(x_1))\|\big| \le{}& \big|\|P(\mathbf{A}^N, f_1(X_1^N))\| - \|P(\mathbf{A}^N, p(X_1^N))\|\big| \\ &+ \big|\|P(\mathbf{A}^N, p(X_1^N))\| - \|P(\mathbf{a}, p(x_1))\|\big| \\ &+ |\|P(\mathbf{a}, p(x_1))\| - \|P(\mathbf{a}, f(x_1))\||\,. \end{aligned} \tag{4.5}$$

By the known strong convergence of $(\mathbf{A}^N, p(X_1^N))$ to $(\mathbf{a}, p(x_1))$, the middle term is $< \frac{\epsilon}{4}$ for all sufficiently large N. For the first and third terms, we use the reverse triangle inequality; in the third term this gives

$$|\|P(\mathbf{a}, p(x_1))\| - \|P(\mathbf{a}, f(x_1))\|| \le \|P(\mathbf{a}, p(x_1)) - P(\mathbf{a}, f(x_1))\|.$$

Let $y = p(x_1)$ and $z = f_1(x_1)$. We may estimate the norm of the difference using the triangle inequality summing over all monomial terms; then we have a sum of terms of the form

$$\|Q_0(\mathbf{a})yQ_1(\mathbf{a})y\cdots Q_{d-1}(\mathbf{a})yQ_d(\mathbf{a}) - Q_0(\mathbf{a})zQ_1(\mathbf{a})z\cdots Q_{d-1}(\mathbf{a})zQ_d(\mathbf{a})\|. \tag{4.6}$$

By introducing intermediate mixed terms of the form $Q_0(\mathbf{a})y\cdots Q_{k-1}(\mathbf{a})yQ_k(\mathbf{a})z\cdots Q_{d-1}(\mathbf{a})zQ_d(\mathbf{a})$ to give a telescoping sum, we can estimate the term in (4.6) by

$$\|Q_0(\mathbf{a})\|\cdots\|Q_d(\mathbf{a})\|\sum_{k=1}^{d}\|y\|^{k-1}\|z\|^{d-k}\|y-z\|. \tag{4.7}$$

Since $\|y\| = \|p(x_1)\| = \|p\|_{L^\infty[-\kappa,\kappa]} \le 1 + \|f_1\|_{L^\infty[-\kappa,\kappa]}$ and $\|z\| = \|f_1(x_1)\| \le 1 + \|f_1\|_{L^\infty[-\kappa,\kappa]}$, each term in the previous sum is bounded by

$$\|y\|^{k-1}\|z\|^{d-k} \le (1 + \|f_1\|_{L^\infty[-\kappa,\kappa]})^{d-1} \le (1 + \|f_1\|_{L^\infty[-\kappa,\kappa]})^{d_P}.$$

Since $\|y - z\| = \|p - f_1\|_{L^\infty[-\kappa,\kappa]}$, combining this with (4.4) shows that the third term in (4.5) is $< \frac{\epsilon}{4}$ for all large N.

The first term in (4.5) is handled in an analogous fashion, with the caveat that the prefactor in (4.7) is replaced by $\|Q_0(\mathbf{A}^N)\|\cdots\|Q_d(\mathbf{A}^N)\|$. Here we use the fact that X_1^N converges strongly towards x_1 to ensure that almost surely, for N large enough,

$$\mathrm{spec}(X_1^N) \subset [-\kappa, \kappa],$$

together with the assumption of strong convergence of $\mathbf{A}^N \to \mathbf{a}$ to show that, for all sufficiently large N,

$$\|Q_0(\mathbf{A}^N)\|\cdots\|Q_d(\mathbf{A}^N)\| \le \max\{1, 2\cdot\|Q_0(\mathbf{a})\|\cdots\|Q_d(\mathbf{a})\|\}.$$

Then we see that the first term in (4.5) is $< \frac{\epsilon}{2}$ for all large N, and so we have bounded the sum $< \epsilon$ for all large N, concluding the proof of the lemma in the case $k = 1$.

Now suppose we have verified the conclusion of the lemma for a given k. We proceed by induction. Taking $(\mathbf{A}^N, f_1(X_1^N), \ldots, f_k(X_k^N))$ as our new input vector, since $f_{k+1}(X_{k+1}^N)$ is independent from all previous terms, the induction hypothesis and the preceding argument in the case $k = 1$ give strong convergence of the augmented vector $(\mathbf{A}^N, f_1(X_1^N), \ldots, f_k(X_k^N), f_{k+1}(X_{k+1}^N))$ as well. Hence, the proof is complete by induction. □

This finally brings us to the proof of Theorem 1.4.

Proof of Theorem 1.4. As above, let $\mathbf{A}^N = (A_1^N, \ldots, A_n^N)$ and let $\mathbf{a} = (a_1, \ldots, a_n)$ be the strong limit. By reindexing the order of the variables in the noncommutative polynomial P appearing in the definition of strong convergence, it suffices to prove the theorem in the case of time-ordered entries: $U_{t_1}^N, \ldots, U_{t_k}^N$ with $t_1 \le t_2 \le \cdots \le t_k$. What's more, we may assume without loss of generality that the time increments $s_1 = t_1, s_2 = t_2 - t_1, \ldots, s_k = t_k - t_{k-1}$ are all in $[0, 4)$. Indeed, if we know the theorem holds in this case, then for a list of ordered times with some gaps 4 or larger, we may introduce intermediate times until all gaps are < 4; then the restricted theorem implies strong convergence for this longer list of marginals, which trivially implies strong convergence for the original list.

Now, set $V_{s_1}^N = U_{t_1}^N$, and $V_{s_j}^N = (U_{t_{j-1}}^N)^* U_{t_j}^N$ for $2 \le j \le k$. As discussed in Section 2.1, these increments of the process are independent, and $V_{s_j}^N$ has the same distribution as $U_{s_j}^N$. Hence, by Proposition 4.3, there are k independent GUE^Ns $X_1^N, \ldots, X_k^N$, and continuous functions $f_{s_j} \colon \mathbb{R} \to \mathbb{C}$, so that $\|f_{s_j}(X_j^N) - V_{s_j}^N\|_{\mathbb{M}_N} \to 0$ as $N \to \infty$. Since the $V_{s_j}^N$ are all independent from $\mathbf{A}^N$, so are the X_j^N. Hence, by Lemma 4.4, taking $x_1, \ldots, x_k$ freely independent semicircular random variables all free from $\mathbf{a}$, it follows that

$$(\mathbf{A}^N, f_{s_1}(X_1^N), \ldots, f_{s_k}(X_k^N)) \text{ converges strongly to } (\mathbf{a}, f_{s_1}(x_1), \ldots, f_{s_k}(x_k)).$$

By the definition of the mapping f_s (cf. (4.2)), $f_{s_j}(x_j)$ has distribution ν_{s_j}, and as all variables in sight are free, $(\mathbf{a}, f_{s_1}(x_1), \ldots, f_{s_k}(x_k))$ has the same distribution as $(\mathbf{a}, v_{s_1}, \ldots, v_{s_k})$ where $(v_s)_{s \ge 0}$ is a free unitary Brownian motion, freely independent from $\mathbf{a}$.

It now follows, since $\|f_{s_j}(X_j^N) - V_{s_j}^N\|_{\mathbb{M}_N} \to 0$, that

$$(\mathbf{A}^N, V_{s_1}^N, \ldots, V_{s_k}^N) \text{ converges strongly to } (\mathbf{a}, v_{s_1}, \ldots, v_{s_k}).$$

(The proof is very similar to the proof of Lemma 4.4.) Finally, we can recover the original variables $U_{t_j}^N = V_{s_1}^N V_{s_2}^N \cdots V_{s_j}^N$. Therefore

$$(\mathbf{A}^N, U_{t_1}^N, \ldots, U_{t_k}^N) \text{ converges strongly to } (\mathbf{a}, v_{s_1}, v_{s_1} v_{s_2} \ldots, v_{s_1} v_{s_2} \cdots v_{s_k}).$$

The discussion at the end of Section 2.3 shows that $(v_{s_1}, v_{s_1} v_{s_2}, \ldots, v_{s_1} v_{s_2} \cdots v_{s_k})$ has the same distribution as $(u_{t_1}, u_{t_2}, \ldots, u_{t_k})$ where $(u_t)_{t \ge 0}$ is a free unitary Brownian motion in the W^*-algebra generated by $(v_s)_{s \ge 0}$, and is therefore freely independent from $\mathbf{a}$. This concludes the proof. $\square$

5 Application to the Jacobi Process

In this final section we combine our main Theorem 1.4 with some of the results of the first and third authors' earlier paper [12], to show that the Jacobi process (cf. (5.1) and (5.2)) has spectral edges that evolve with finite propagation speed.

There are three classical Hermitian Gaussian ensembles that have been well studied. The first is the Gaussian Unitary Ensemble described in detail above, whose analysis was initiated by Wigner [50] and began random matrix theory. The second is the *Wishart Ensemble*, also known (through its applications in statistics) as a *sample covariance matrix*. Let $a \ge 1$, and let $X = X^N$ be an $N \times \lceil aN \rceil$ matrix all of whose entries are independent normal random variables of variance $\frac{1}{N}$; then $W = XX^*$ is a Wishart ensemble with parameter a. As $N \to \infty$, its empirical spectral distribution converges almost surely to a law known as the *Marchenko-Pastur distribution*; this was proved in [36]. As with the Gaussian Unitary Ensemble, it also has a spectral edge, and the largest eigenvalue when properly renormalized has the Tracy-Widom law.

The third Hermitian Gaussian ensemble is the *Jacobi Ensemble*. Let W_a and W_b' be independent Wishart ensembles of parameters $a, b \geq 1$. Then it is known that $W_a + W_b'$ is a Wishart ensemble of parameter $a + b$, and is a.s. invertible (cf. [11, Lemma 2.1]). The associated Jacobi Ensemble is

$$J = J_{a,b} = (W_a + W_b')^{-\frac{1}{2}} W_a (W_a + W_b')^{-\frac{1}{2}}. \tag{5.1}$$

Such matrices have been studied in the statistics literature for over thirty years; they play a key role in MANOVA (multivariate analysis of variance) and are sometimes simply called MANOVA matrices. The joint law of eigenvalues is explicitly known, but the large-N limit is notoriously harder than the Gaussian Unitary and Wishart Ensembles. In [11], the present first author made the following discovery which led to a new approach to the asymptotics of the ensemble: its joint law can be described by a product of randomly rotated projections, as follows. (For the sake of making the statement simpler, we assume a, b are such that aN and bN are integers.)

Theorem 5.1 ([11], Theorem 2.2)**.** *Let $J_{a,b} = J_{a,b}^N$ be an $N \times N$ Jacobi ensemble with parameters $a, b \geq 1$. Let $P, Q \in \mathbb{M}_{(a+b)N}$ be (deterministic) orthogonal projections with $\mathrm{rank}(P) = bN$ and $\mathrm{rank}(Q) = N$. Let $U \in \mathbb{U}_{(a+b)N}$ be a random unitary matrix sampled from the Haar measure. Then QU^*PUQ, viewed as a random matrix in $\mathbb{M}_N$ via the unitary isomorphism $\mathbb{M}_N \cong Q\mathbb{M}_{(a+b)N}Q$, has the same distribution as $J_{a,b}$.*

Given two closed subspaces $\mathbb{V}, \mathbb{W}$ of a Hilbert space $\mathbb{H}$, if $P\colon \mathbb{H} \to \mathbb{V}$ and $Q\colon \mathbb{H} \to \mathbb{W}$ are the orthogonal projections, then the operator QPQ is known as the *operator valued angle* between the two subspaces. (Indeed, in the finite-dimensional setting, the eigenvalues of QPQ are trigonometric polynomials in the principal angles between the subspaces $\mathbb{V}$ and $\mathbb{W}$.) Thus, the law of the Jacobi ensemble records all the remaining information about the angles between two uniformly randomly rotated subspaces of fixed ranks. These observations were used to make significant progress in understanding the Jacobi Ensemble in statistical applications (cf. [29]), and to generalize many of these results to the full expected universality class (beyond Gaussian entries) in the limit (cf. [21]).

In terms of the large-N limit: letting $\alpha = \frac{b}{a+b}$ and $\beta = \frac{1}{a+b}$, we have $\mathrm{tr}P = \alpha$ and $\mathrm{tr}Q = \beta$ fixed as N grows, and therefore there are limit projections p, q of these same traces. The chosen Haar distributed unitary matrices converge in noncommutative distribution to a Haar unitary operator u freely independent from p, q, and so the empirical spectral distribution of $J_{a,b}$ converges to the law of qu^*puq, which was explicitly computed in [49] as an elementary example of free multiplicative convolution:

$$\mathrm{Law}_{qu^*puq} = (1 - \min\{\alpha, \beta\})\delta_0 + \max\{\alpha + \beta - 1, 0\}\delta_1 + \frac{\sqrt{(r_+ - x)(x - r_-)}}{2\pi x(1-x)} \mathbb{1}_{[r_-, r_+]}\, dx,$$

where $r_\pm = \alpha + \beta - 2\alpha\beta \pm 2\sqrt{\alpha\beta(1-\alpha)(1-\beta)}$. Furthermore, it was shown in [29] that the Jacobi Ensemble has a spectral edge, the rate of convergence of the largest eigenvalue is $N^{-2/3}$ (as with the Gaussian Unitary and Wishart Ensembles), and the rescaled limit distribution of the largest eigenvalue is the Tracy–Widom law of [44].

Simultaneously to these developments, Voiculescu [48] introduced *free liberation*. Given two subalgebras A, B of a W^*-probability space $(\mathscr{A}, \tau)$ and a Haar unitary operator $u \in \mathscr{A}$ that is freely independent from A, B, the rotated subalgebra u^*Au is freely independent from B. If $(u_t)_{t\geq 0}$ is a free unitary Brownian motion freely independent from A, B, it is not generally true that $u_t^*Au_t$ is free from B for any finite t (in particular when $t = 0$ we just have A, B), but since the (strong operator) limit as $t \to \infty$ of u_t is a Haar unitary, this process "liberates" A and B. This concept was used to define several important regularized versions of measures associated to free entropy and free information theory, and to this days plays an important role in free probability theory. The special case that A, B

are algebras generated by two projections has been extensively studied [15, 16, 17, 18, 19, 25, 26, 27], as the best special case where one can hope to compute all quantities fairly explicitly.

In the first and third authors' paper [12, Section 3.2], the following was proved.

Theorem 5.2 ([12], Lemmas 3.2–3.6)**.** *Let p, q be orthogonal projections with traces α, β, and let $(u_t)_{t\geq 0}$ be a free unitary Brownian motion freely independent from p, q. Let $\mu_t = \mathrm{Law}_{qu_t^* pu_t q}$. Then*

$$\mu_t = (1 - \min\{\alpha, \beta\})\delta_0 + \max\{\alpha + \beta - 1, 0\}\delta_1 + \widetilde{\mu}_t$$

where $\widetilde{\mu}_t$ is a positive measure (of mass $\min\{\alpha, \beta\} - \max\{\alpha + \beta - 1, 0\}$). Let I_1, I_2 be two disjoint open subintervals of $(0, 1)$. If $\operatorname{supp} \widetilde{\mu}_{t_0} \subset I_1 \sqcup I_2$ for some $t_0 \geq 0$, then $\operatorname{supp} \widetilde{\mu}_t \subset I_1 \sqcup I_2$ for $|t - t_0|$ sufficiently small; moreover, $\widetilde{\mu}_t(I_1)$ and $\widetilde{\mu}_t(I_2)$ do not vary with t close to t_0.

If $\widetilde{\mu}_t$ has a continuous density on $(0, 1)$ for $t > 0$, and $x_{t_0} \in (0, 1)$ is a boundary point of $\operatorname{supp} \widetilde{\mu}_{t_0}$, then for $|t - t_0|$ sufficiently small there is a C^1 function $t \mapsto x(t)$ with $x(t_0) = x_{t_0}$ so that $x(t)$ is a boundary point of $\operatorname{supp} \widetilde{\mu}_{t_0}$.

Finally, in the special case $\alpha = \beta = \frac{1}{2}$, for all $t > 0$, $\widetilde{\mu}_t$ possesses a continuous density which is analytic on the interior of its support.

Remark 5.3. (1) It is expected that the final statement, regarding the existence of a continuous density, holds true for all $\alpha, \beta \in (0, 1)$; at present time, this is only known for $\alpha = \beta = \frac{1}{2}$. Nevertheless, the "islands stay separated" result holds in general.

(2) Our method of proof of the regularity of $\widetilde{\mu}_t$ involved a combination of free probability, complex analysis, and PDE techniques. In [28], Izumi and Ueda partly extended this framework beyond the $\alpha = \beta = \frac{1}{2}$ case; but they were still not able to prove continuity of the measure. They did, however, give a much simpler proof of the result in the case $\alpha = \beta = \frac{1}{2}$: here, $\widetilde{\mu}_t$ can be described as the so-called Szegő transform (from the unit circle to the unit interval) of the law of $v_0 u_t$, where v_0 is determined by the law of qpq. Via this description, the regularity result is an immediate consequence of Theorem 2.5 above.

(3) Let us note that $\alpha = \beta = \frac{1}{2}$ corresponds to $a = b = 1$, meaning the "square" Jacobi ensemble. This is, of course, the case that is least interesting to statisticians: in MANOVA problems the data sets are typically time series, where there are many more samples than detection sites, meaning that $a, b \ggg 1$. In fact, it is generally questioned whether the Jacobi Ensemble is a realistic regime for real world applications, rather than building the Wishart Ensembles out of $N \times M$ Gaussian matrices where $\frac{M}{N} \to \infty$.

Thus, it is natural to consider the corresponding finite-t deformation of the Jacobi Ensemble. The *matrix Jacobi process* J_t^N associated to the projections $P^N, Q^N \in \mathbb{M}_N$, is given by

$$J_t^N = Q^N (U_t^N)^* P^N U_t^N Q^N \tag{5.2}$$

where $(U_t^N)_{t\geq 0}$ is a Brownian motion in $\mathbb{U}_N$. (Typically P^N, Q^N are deterministic; they may also be chosen randomly, in which case U_t^N must be chosen independent from them.) This is a diffusion process in $\mathbb{M}_N^{[0,1]}$: it lives a.s. in the space of matrices $M \in \mathbb{M}_N$ with $0 \leq M \leq 1$ (i.e. M is self-adjoint and has eigenvalues in $[0, 1]$). Note that the initial value is $J_0^N = Q^N P^N Q^N$, the operator-valued angle between the subspaces in the images of P^N, Q^N. In particular, the Jacobi process records (through its eigenvalues) the evolution of the principal angles between two subspaces as they are continuously rotated by a unitary Brownian motion.

In the case $N = 1$, the process (5.2) precisely corresponds to what is classically known as the Jacobi process: the Markov process on $[0, 1]$ with generator $\mathscr{L} = x(x - 1)\frac{\partial^2}{\partial x^2} - (cx + d)\frac{\partial}{\partial x}$, where

$c = 2\min\{\alpha, \beta\} - 1$, $d = |\alpha - \beta|$. This is where the name comes from, as the orthogonal polynomials associated to this Markov process are the Jacobi polynomials, cf. [19].

Remark 5.4. Comparing to Theorem 5.1, we have now compressed the projections and the Brownian motion into $\mathbb{M}_N$ from the start. We could instead formulate the process as in that theorem by choosing projections and Brownian motion in a larger space, which would have the effect of using a "corner" of a higher-dimensional Brownian motion instead of U_t^N. While this makes a difference for the finite-dimensional distribution, it does not affect the large-N behavior.

This brings us to our main application. First note that, from our main Theorem 1.4, the Jacobi process converges strongly.

Corollary 5.5. *Let P^N, Q^N be deterministic orthogonal projections in $\mathbb{M}_N$, and suppose $\{P^N, Q^N\}$ converges strongly to $\{p, q\}$. Let $(u_t)_{t\geq 0}$ be a free unitary Brownian motion freely independent from p, q. Then for each $t \geq 0$ the Jacobi process marginal J_t^N converges strongly to $j_t = qu_t^* pu_t q$. What's more, if $f \in C[0,1]$ is any continuous test function, then $\|f(J_t^N)\| \to \|f(j_t)\|$ a.s. as $N \to \infty$.*

Proof. The strong convergence statement is an immediate corollary to Theorem 1.4, with $A_1^N = P^N$, $A_2^N = Q^N$, and $n = 2, k = 1$. The extension to continuous test functions beyond polynomials is then an elementary Weierstrass approximation argument. □

Example 5.6. For fixed $k \in \mathbb{N}$, select two orthogonal projections $P, Q \in \mathbb{M}_k$. Then define $P^N, Q^N \in \mathbb{M}_{kN}$ by $P^N = P \otimes I^N$ and $Q^N = Q \otimes I^N$. (Here we are identifying $\mathbb{M}_k \otimes \mathbb{M}_N \cong \mathbb{M}_{kN}$ via the Kronecker product.) If F is a noncommutative polynomial in two indeterminates, then

$$F(P^N, Q^N) = F(P, Q) \otimes I^N$$

and it follows immediately that $\{P^N, Q^N\}$ converges strongly to $\{P, Q\}$ (i.e. the W^*-probability space can be taken to be $(\mathbb{M}_k, \mathrm{tr})$). Expanding this space to include a free unitary Brownian motion freely independent from $\{P, Q\}$ and setting $j_t = Qu_t^* P u_t Q$, Corollary 5.5 yields that the Jacobi process J_t^{kN} with initial value $Q^N P^N Q^N$ converges strongly to j_t.

Figure 2 illustrates the eigenvalues of J_t^{kN} with $k = 4$, $N = 100$, and initial projections given by

$$P = \begin{bmatrix} 0.2 & 0.4 \\ 0.4 & 0.8 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0.8 & 0.4 \\ 0.4 & 0.2 \end{bmatrix} \otimes \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \qquad Q = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \otimes I_2$$

which have been selected so that the initial operator-valued angle QPQ has non-trivial eigenvalues 0.2 and 0.8; this therefore holds as well for $Q^N P^N Q^N$ for all N. This implies that the subspaces $P^N(\mathbb{C}^{kN})$ and $Q^N(\mathbb{C}^{kN})$ have precisely two distinct principal angles.

As is plainly visible in Figure 2, for small time, the eigenvalues (which are fixed trigonometric polynomials in the principal angles) stay close to their initial values. That is: despite the fact that the diffusion's measure is fully supported on $\mathbb{M}_N^{[0,1]}$ for every $t, N > 0$, the eigenvalues move with finite speed for all large N. That is our final theorem.

Theorem 5.7. *For each $N \geq 1$, let $(U_t^N)_{t\geq 0}$ be a Brownian motion on $\mathbb{U}_N$, let $\mathbb{V}^N$ and $\mathbb{W}^N$ be subspaces of $\mathbb{C}^N$, and suppose that the orthogonal projections onto these subspaces converge jointly strongly as $N \to \infty$. Suppose there is a fixed finite set $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_k\}$ of angles so that all principal angles between $\mathbb{V}^N$ and $\mathbb{W}^N$ are in $\boldsymbol{\theta}$ for all N. Fix any open neighborhood $\mathscr{O}$ of $\boldsymbol{\theta}$. Then there exists some $T > 0$ so that, for all $t \in [0, T]$, with probability* 1, all *principal angles between $U_t^N(\mathbb{V}^N)$ and $\mathbb{W}^N$ are in $\mathscr{O}$ for all sufficiently large N.*

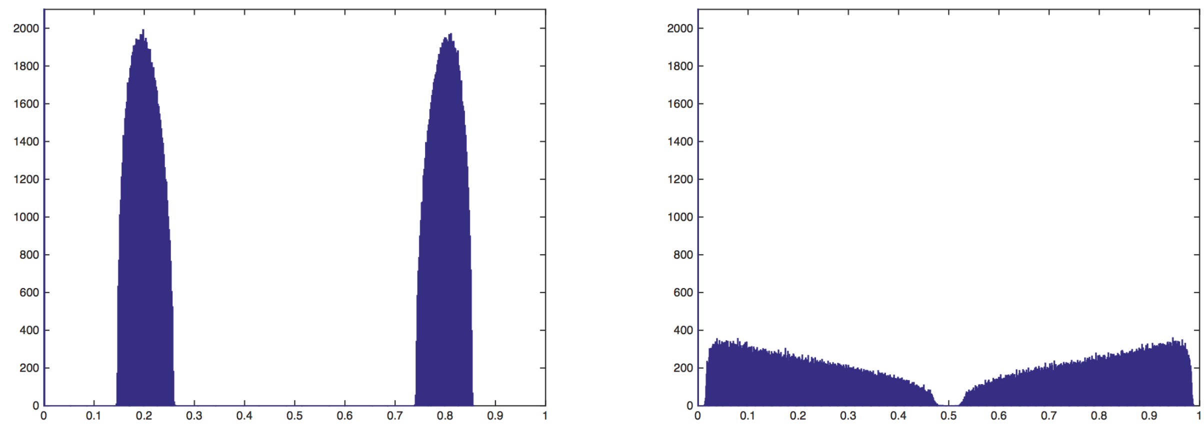

Figure 2: The spectral distribution of the Jacobi process J_t^{kN} of Example 5.6 with $k = 4$, $N = 100$ and times $t = 0.01$ (on the left) and $t = 0.25$ (on the right). The histograms were made with 1000 trials each, yielding 4×10^5 eigenvalues sorted into 1000 bins.

Proof. Let P^N and Q^N be the projections onto $\mathbb{V}^N$ and $\mathbb{W}^N$. Then there is a fixed list $\boldsymbol{\lambda} = \{\lambda_1, \ldots, \lambda_k\}$ in $[0,1]$ so that all eigenvalues of $Q^N P^N Q^N$ are in $\boldsymbol{\lambda}$. (The eigenvalues λ_j are certain fixed trigonometric polynomials in $\boldsymbol{\theta}$). Let J_t^N be the Jacobi process associated to P^N, Q^N, and let j_t be the associated large-N limit. By Corollary 5.5, for any $t \geq 0$ and any $f \in C[0,1]$, $\|f(J_t^N)\| \to \|f(j_t)\|$ a.s. as $N \to \infty$.

Applying this at time $t = 0$, let $\lambda_i, \lambda_j \in \boldsymbol{\lambda}$ with $\lambda_i < \lambda_j$ such that no elements of $\boldsymbol{\lambda}$ are in the interval (λ_i, λ_j). Now let f be a continuous bump function supported in (λ_i, λ_j). Then $f(J_0^N) = 0$, and it therefore follows that $\|f(j_0)\| = 0$. As this holds for all bump function supported in (λ_i, λ_j), it follows that $\mathrm{spec}(j_0)$ does not intersect (λ_i, λ_j). Thus j_0 has pure point spectrum precisely equal to $\boldsymbol{\lambda}$.

Now, fix any $\epsilon > 0$; by (induction on) Theorem 5.2, for sufficiently small $t > 0$, $\mathrm{spec}(j_t)$ is contained in $\boldsymbol{\lambda}_\epsilon$ (the union of ϵ-balls centered at the points of $\boldsymbol{\lambda}$). Now, suppose (for a contradiction) that, for some N_0, $J_t^{N_0}$ possesses an eigenvalue $\lambda \in (0,1) \setminus \boldsymbol{\lambda}_\epsilon$. Let g be a bump function supported in $(0,1) \setminus \boldsymbol{\lambda}_\epsilon$ that is equal to 1 on a neighborhood of λ; then $\|g(J_t^{N_0})\| \geq 1$. But, by Corollary 5.5, we know $\|g(J_t^N)\| \to \|g(j_t)\| = 0$ a.s. as $N \to \infty$. Thus, for all sufficiently large N, $\|g(J_t^N)\| < 1$, which implies that λ is not an eigenvalue. As this holds for any point in $(0,1) \setminus \boldsymbol{\lambda}_\epsilon$, it follows that $\mathrm{spec}(J_t^N)$ is almost surely contained in $\boldsymbol{\lambda}_\epsilon$ for all sufficiently large N.

The result now follows from the fact that the principal angles between $U_t^N(\mathbb{V})$ and $\mathbb{W}$ are fixed continuous functions (trigonometric polynomials) in the eigenvalues of J_t^N. $\square$

Acknowledgments

This paper was initiated during a research visit of BC to TK in late 2012; substantial subsequent progress was made during a one-month visit of AD to BC at AIMR, Sendai during the fall of 2013. We would like to express our great thanks to UC San Diego and AIMR for their warm hospitality, fruitful work environments, and financial support that aided in the completion of this work. We would like to thank Ioana Dumitriu for helping to design the simulations that produced the figures. Finally, we express our gratitude to the anonymous referees whose careful reading and detailed comments significantly improved the presentation of this manuscript (particularly in Section 3.2).

References

[1] ANDERSON, G. W., GUIONNET, A., AND ZEITOUNI, O. *An introduction to random matrices*, vol. 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010.

[2] BAI, Z. D. Methodologies in spectral analysis of large-dimensional random matrices, a review. *Statist. Sinica 9*, 3 (1999), 611–677. With comments by G. J. Rodgers and Jack W. Silverstein; and a rejoinder by the author.

[3] BAI, Z. D., AND YIN, Y. Q. Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a Wigner matrix. *Ann. Probab. 16*, 4 (1988), 1729–1741.

[4] 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.

[5] 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.

[6] 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.

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

[8] CÉBRON, G. Free convolution operators and free Hall transform. *Journal of Functional Analysis 265*, 11 (2013), 2645–2708.

[9] CÉBRON, G., AND KEMP, T. Fluctuations of Brownian motions on $\mathbb{GL}_N$. *arXiv:1409.5624* (2014).

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

[11] COLLINS, B. Product of random projections, Jacobi ensembles and universality problems arising from free probability. *Probab. Theory Related Fields 133*, 3 (2005), 315–344.

[12] COLLINS, B., AND KEMP, T. Liberation of projections. *J. Funct. Anal. 266*, 4 (2014), 1988–2052.

[13] COLLINS, B., AND MALE, C. The strong asymptotic freeness of Haar and deterministic matrices. *Ann. Sci. Éc. Norm. Supér. (4) 47*, 1 (2014), 147–163.

[14] DAHLQVIST, A. Integration formula for Brownian motion on classical compact Lie groups. *To appear in Ann. Inst. H. Poincaré Probab. Statist.* (2016).

[15] DEMNI, N. Free Jacobi process. *J. Theoret. Probab. 21*, 1 (2008), 118–143.

[16] DEMNI, N. Free Jacobi process associated with one projection: local inverse of the flow. *arXiv:1502.00013* (2015).

[17] DEMNI, N., HAMDI, T., AND HMIDI, T. Spectral distribution of the free Jacobi process. *Indiana Univ. Math. J. 61*, 3 (2012), 1351–1368.

[18] DEMNI, N., AND HMIDI, T. Spectral distribution of the free Jacobi process associated with one projection. *Colloq. Math. 137*, 2 (2014), 271–296.

[19] DEMNI, N., AND ZANI, M. Large deviations for statistics of the Jacobi process. *Stochastic Process. Appl. 119*, 2 (2009), 518–533.

[20] DRIVER, B. K., HALL, B. C., AND KEMP, T. The large-N limit of the Segal-Bargmann transform on $\mathbb{U}_N$. *J. Funct. Anal. 265*, 11 (2013), 2585–2644.

[21] ERDŐS, L., AND FARRELL, B. Local eigenvalue density for general MANOVA matrices. *J. Stat. Phys. 152*, 6 (2013), 1003–1032.

[22] EVANS, L. C. *Partial differential equations*, second ed., vol. 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2010.

[23] GUSTAVSSON, J. Gaussian fluctuations of eigenvalues in the GUE. *Ann. Inst. H. Poincaré Probab. Statist. 41*, 2 (2005), 151–178.

[24] HAAGERUP, U., AND THORBJØRNSEN, S. A new application of random matrices: $\mathrm{Ext}(C^*_{\mathrm{red}}(F_2))$ is not a group. *Ann. of Math. (2) 162*, 2 (2005), 711–775.

[25] HIAI, F., MIYAMOTO, T., AND UEDA, Y. Orbital approach to microstate free entropy. *Internat. J. Math. 20*, 2 (2009), 227–273.

[26] HIAI, F., AND UEDA, Y. Notes on microstate free entropy of projections. *Publ. Res. Inst. Math. Sci. 44*, 1 (2008), 49–89.

[27] HIAI, F., AND UEDA, Y. A log-Sobolev type inequality for free entropy of two projections. *Ann. Inst. Henri Poincaré Probab. Stat. 45*, 1 (2009), 239–249.

[28] IZUMI, M., AND UEDA, Y. Remarks on free mutual information and orbital free entropy. *Nagoya Math. J. 220* (2015), 45–66.

[29] JOHNSTONE, I. M. Multivariate analysis and Jacobi ensembles: largest eigenvalue, Tracy-Widom limits and rates of convergence. *Ann. Statist. 36*, 6 (2008), 2638–2716.

[30] KEMP, T. Heat Kernel Empirical Laws on $\mathbb{U}_N$ and $\mathbb{GL}_N$. *To appear in J. Theoret. Probab.* (2015).

[31] KEMP, T. The Large-N Limits of Brownian Motions on $\mathbb{GL}_N$. *To appear in Int. Math. Res. Not.* (2015).

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

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

[34] LÉVY, T., AND MAÏDA, M. Central limit theorem for the heat kernel measure on the unitary group. *Journal of Functional Analysis 259*, 12 (Dec. 2010), 3163–3204.

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

[36] MARCHENKO, V. A., AND PASTUR, L. A. Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb. (N.S.) 72*, 114 (1967), 507–536.

[37] MEHTA, M. L. *Random matrices*, third ed., vol. 142 of *Pure and Applied Mathematics (Amsterdam)*. Elsevier/Academic Press, Amsterdam, 2004.

[38] NICA, A., AND SPEICHER, R. *Lectures on the combinatorics of free probability*, vol. 335 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 2006.

[39] PÓLYA, G., AND SZEGŐ, G. *Problems and theorems in analysis. I*. Classics in Mathematics. Springer-Verlag, Berlin, 1998. Series, integral calculus, theory of functions, Translated from the German by Dorothee Aeppli, Reprint of the 1978 English translation.

[40] RAINS, E. M. Combinatorial properties of Brownian motion on the compact classical groups. *J. Theoret. Probab. 10*, 3 (1997), 659–679.

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

[42] SALOFF-COSTE, L. *Aspects of Sobolev-type inequalities*, vol. 289 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 2002.

[43] SENGUPTA, A. *Traces in two-dimensional QCD: the large N-limit*. Aspects of Mathematics, E38. Friedr. Vieweg & Sohn, Wiesbaden, 2008.

[44] TRACY, C. A., AND WIDOM, H. Level-spacing distributions and the Airy kernel. *Comm. Math. Phys. 159*, 1 (1994), 151–174.

[45] 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.

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

[47] VOICULESCU, D. A strengthened asymptotic freeness result for random matrices with applications to free entropy. *Internat. Math. Res. Notices*, 1 (1998), 41–63.

[48] VOICULESCU, D. The analogues of entropy and of Fisher's information measure in free probability theory. VI. Liberation and mutual free information. *Adv. Math. 146*, 2 (1999), 101–166.

[49] VOICULESCU, D. V., DYKEMA, K. J., AND NICA, A. *Free random variables*, vol. 1 of *CRM Monograph Series*. American Mathematical Society, Providence, RI, 1992. A noncommutative probability approach to free products with applications to random matrices, operator algebras and harmonic analysis on free groups.

[50] WIGNER, E. P. On the distribution of the roots of certain symmetric matrices. *Ann. of Math. (2) 67* (1958), 325–327.