Annales Henri Lebesgue
1 (2018) 247-265

ANNALES
HENRI LEBESGUE

ELIZABETH MECKES
TAI MELCHER

CONVERGENCE OF THE EMPIRICAL SPECTRAL MEASURE OF UNITARY BROWNIAN MOTION

CONVERGENCE DE LA MESURE SPECTRALE EMPIRIQUE D'UN MOUVEMENT BROWNIEN UNITAIRE

Abstract. — Let $\{U_t^N\}_{t\geqslant 0}$ be a standard Brownian motion on $\mathbb{U}(N)$. For fixed $N\in\mathbb{N}$ and $t>0$, we give explicit almost-sure bounds on the L_1-Wasserstein distance between the empirical spectral measure of U_t^N and the large-N limiting measure. The bounds obtained are tight enough that we are able to use them to study the evolution of the eigenvalue process in time, bounding the rate of convergence of paths of the measures on compact time intervals. The proofs use tools developed by the first author to obtain rates of convergence of the empirical spectral measures in classical random matrix ensembles, as well as recent estimates for the rates of convergence of moments of the ensemble-averaged spectral distribution.

Résumé. — Soit $\{U_t^N\}_{t\geqslant 0}$ un mouvement brownien standard sur $\mathbb{U}(N)$. Étant donnés $N\in\mathbb{N}$ et $t>0$, nous donnons des bornes presque sûres explicites sur la distance de Wasserstein L^1 entre la mesure spectrale empirique de U_t^N et la mesure limite en N. Nos bornes sont assez précises pour permettre l'étude de l'évolution du processus des valeurs propres, en bornant la vitesse de convergence de chemins de mesures sur des intervalles de temps compacts. Les

Keywords: Unitary Brownian motion, empirical spectral measure, heat kernel measure, concentration.
2010 *Mathematics Subject Classification:* 60B20, 58J65.
DOI: `https://doi.org/10.5802/ahl.7`
(*) E. Meckes' research supported in part by NSF DMS 1612589. T. Melcher's research supported in part by NSF DMS 1255574.

preuves reposent sur des outils développés par le premier auteur pour obtenir des vitesses de convergence sur la mesure spectrale empirique dans des ensembles de matrices aléatoires classiques, ainsi que des estimées récentes sur la vitesse de convergence des moments pour la distribution spectrale moyennée sur l'ensemble.

1. Introduction

This paper studies the convergence of the empirical spectral measure of Brownian motion on the unitary group $\mathbb{U}(N)$ to its large N limit. Brownian motion on large unitary groups has generated significant interest in recent years, due in part to its relationships with two-dimensional Yang–Mills theory and with the object from free probability theory called free unitary Brownian motion. As is typical in random matrix theory, a particular focus is on the asymptotic behavior (as N tends to infinity) of the eigenvalues. Such results are usually formulated as limiting results for the (random) spectral measure of unitary Brownian motion at a fixed time t, as N tends to infinity; see for example [Rai97, Xu97, Bia97a, Bia97b, Lév08, LM10, DHK13, Kem15, CDK18] and the references therein.

A significant focus in random matrix theory in recent years has been in obtaining *non-asymptotic* results; that is, quantitative results describing the behavior of random matrices of fixed (large) size; see, for example, [RV10]. In this context, many tools have been developed to study the spectral distributions of random matrices in fixed high dimensions. Among them is an approach developed by the first author with M. Meckes (see [MM17] for a survey) which allows for quantitative estimates on rates of convergence of the empirical spectral measure in a wide assortment of random matrix ensembles. This approach is based on concentration of measure and bounds for suprema of stochastic processes, in combination with more classical tools from matrix analysis, approximation theory, and Fourier analysis. In the present paper, we combine some of these techniques with recent estimates on the rates of convergence of the moments for the empirical spectral distribution of unitary Brownian motion [CDK18] to prove asymptotically almost sure rates of convergence. The concentration inequalities we obtain are sharp enough to be used to study the evolution of the eigenvalue process in time; we are able to control the rate of convergence not just at a fixed finite time, but to compare paths of the spectral measures over compact time intervals to the deterministic path described by the evolution of the large N limiting measures.

Statement of results

Let $\mathbb{U}(N)$ denote the unitary group and $\mathfrak{u}(N)$ its Lie algebra of skew-Hermitian matrices equipped with the scaled (real) inner product $\langle U, V\rangle_N := N\Re\operatorname{tr}(UV^*)$. This is the unique scaling that gives meaningful limiting behavior as $N \to \infty$; see for example Remark 3.4 of [DHK13]. The inner product on $\mathfrak{u}(N)$ induces a left-invariant Riemannian metric on $\mathbb{U}(N)$, and we may define Brownian motion on $\mathbb{U}(N)$ as the Markov diffusion $\{U_t^N\}_{t\geqslant 0}$ issued from the identity with generator $\frac{1}{2}\Delta_N$, that is, one half the left-invariant Laplacian on $\mathbb{U}(N)$ with respect to this metric. One may equivalently describe U_t^N as the solution to the Itô stochastic differential equation

$$\mathrm{d}U_t^N = U_t^N \mathrm{d}W_t^N - \frac{1}{2}U_t^N \mathrm{d}t$$

with $U_0^N = I_N$, where W_t is a standard Brownian motion on $\mathfrak{u}(N)$ (for example, take $\{\xi_k\}_{k=0}^{N^2-1}$ an orthonormal basis of $\mathfrak{u}(N)$ with respect to the given inner product and $W_t^N = \sum_{j=0}^{N^2-1} b_t^j \xi_j$, where the b_t^j are independent standard Brownian motions on $\mathbb{R}$). This realization of unitary Brownian motion is computationally more useful and is mainly what will be used in the sequel. It should be noted that another standard description of the unitary Brownian motion is via a stochastic differential equation with respect to a Hermitian Brownian motion, which results in a difference of a factor of i in the diffusion coefficient. For $t > 0$, let $\rho_t^N = \mathrm{Law}(U_t^N)$ denote the end point distribution of Brownian motion; ρ_t^N is called the heat kernel measure on $\mathbb{U}(N)$.

Our primary object of interest is the empirical spectral measure of unitary Brownian motion. A matrix $U \in \mathbb{U}(N)$ has N complex eigenvalues of modulus one which we denote by $e^{i\theta_1}, \dots, e^{i\theta_N}$ (repeated according to multiplicity), and the spectral measure of U is defined to be the probability measure on the unit circle $\mathbb{S}^1$ given by

$$\mu_U := \frac{1}{N}\sum_{j=1}^{N} \delta_{e^{i\theta_j}}.$$

In particular, for $f \in C(\mathbb{S}^1)$

$$\int_{\mathbb{S}^1} f \mathrm{d}\mu_U = \frac{1}{N}\sum_{j=1}^{N} f(e^{i\theta_j}).$$

For each fixed $t > 0$, U_t^N is a random unitary matrix, and we denote its empirical spectral measure by $\mu_t^N := \mu_{U_t^N}$. In [Bia97a], Biane showed that the random probability measure μ_t^N converges weakly almost surely to a deterministic probability measure, which we denote by ν_t: that is, for all $f \in C(\mathbb{S}^1)$,

$$\lim_{N\to\infty} \int_{\mathbb{S}^1} f \mathrm{d}\mu_t^N = \int_{\mathbb{S}^1} f \,\mathrm{d}\nu_t \text{ a.s.}$$

The measure ν_t represents in some sense the spectral distribution of a "free unitary Brownian motion". For $t > 0$, ν_t possesses a continuous density that is symmetric about $1 \in \mathbb{S}^1$. When $0 < t < 4$, ν_t is supported on an arc strictly contained in the circle; for $t \geqslant 4$, $\mathrm{supp}(\nu_t) = \mathbb{S}^1$. The paper [CDK18] presents a nice brief summary of these and other properties of ν_t and the construction of free unitary Brownian motion.

In the present paper, we give estimates on the L_1-Wasserstein distance between the empirical spectral distribution μ_t^N and its limiting spectral measure ν_t. For Borel probability measures μ and ν on a Polish space $(\mathcal{X}, d)$, the L_1-Wasserstein distance between μ and ν is defined by

$$W_1(\mu,\nu) := \inf\left\{\int d(x,y)\,\mathrm{d}\pi(x,y) : \pi \text{ is a coupling of } \mu \text{ and } \nu\right\}.$$

The following equivalent dual representation of W_1 is known as the Kantorovich–Rubinstein formula (see, e.g., Particular Case 5.16 in [Vil09]):

$$W_1(\mu,\nu) = \sup\left\{\int f\,\mathrm{d}\mu - \int f\,\mathrm{d}\nu : |f|_L \leqslant 1\right\},$$

where $|f|_L$ denotes the Lipschitz constant of f. For X_1, X_2 random elements of $\mathcal{X}$, $W_1(X_1, X_2)$ should be interpreted as the W_1-distance between the distributions of X_1 and X_2.

The main results of this paper are the following.

Theorem 1.1. — *Let $\{U_t^N\}_{t\geqslant 0}$ be a Brownian motion on $\mathbb{U}(N)$. For $t>0$, let μ_t^N denote the empirical spectral measure U_t as above, and let $\overline{\mu}_t^N$ denote the ensemble-averaged spectral measure of U_t^N defined by*

$$\int_{\mathbb{S}^1} f\mathrm{d}\overline{\mu}_t^N := \mathbb{E}\int_{\mathbb{S}^1} f\mathrm{d}\mu_t^N.$$

Then there is a constant $C\in(0,\infty)$ such that with probability one, for all $N\in\mathbb{N}$ sufficiently large and $t>0$,

$$W_1(\mu_t^N, \overline{\mu}_t^N) \leqslant C\left(\frac{t}{N^2}\right)^{1/3}.$$

and, for all $N\in\mathbb{N}$ sufficiently large and $t\geqslant 8(\log N)^2$,

$$W_1(\mu_t^N, \overline{\mu}_t^N) \leqslant \frac{C}{N^{2/3}}.$$

Theorem 1.2. — *Let ν_t be the limiting spectral measure for unitary Brownian motion described above. There is a constant $C\in(0,\infty)$ such that for all $N\in\mathbb{N}$ and $t>0$*

$$W_1(\overline{\mu}_t^N, \nu_t) \leqslant C\min\left\{\frac{t^{2/5}\log N}{N^{2/5}}, e^{-\frac{t(1+o(1))}{8\log(N)}} + \frac{1}{N}\right\}.$$

One may infer from these bounds direct (a.s.) estimates on the rate of convergence of the empirical spectral distribution to its limiting distribution for all sufficiently large N. To the authors' knowledge, these results constitute the first known rates of convergence for μ_t^N itself; previously the only known convergence rates were for moments of the ensemble-averaged spectral measure $\overline{\mu}_t^N$ [CDK18].

A key advantage of such rates is that they may be applied to obtain almost sure convergence of paths of spectral measures. The following theorem gives uniform bounds on the Wasserstein distance between the empirical spectral measures and the deterministic limiting measures on compact time intervals.

Theorem 1.3. — *Let $T\geqslant 0$. There are constants c, C such that for all $x\geqslant c\frac{T^{2/5}\log(N)}{N^{2/5}}$,*

$$\mathbb{P}\left(\sup_{0\leqslant t\leqslant T} W_1(\mu_t^N,\nu_t) > x\right) \leqslant C\left(\frac{T}{x^2}+1\right)e^{-\frac{N^2x^2}{T}}.$$

In particular, with probability one for N sufficiently large

$$\sup_{0\leqslant t\leqslant T} W_1(\mu_t^N,\nu_t) \leqslant c\frac{T^{2/5}\log(N)}{N^{2/5}}.$$

As a technical tool, we also determine rates for the convergence in time of Biane's measure to the uniform distribution on $\mathbb{S}^1$.

Proposition 1.4. — *Let ν_t denote the limiting spectral measure and ν the uniform measure on $\mathbb{S}^1$. Then there is a constant $C \in (0,\infty)$ so that for all $t \geqslant 1$*

$$W_1(\nu_t, \nu) \leqslant C t^{3/2} e^{-t/4}.$$

The organization of the paper is as follows. In Section 2, we establish improved concentration estimates for heat kernel measure on $\mathbb{U}(N)$ via a coupling of Brownian motions on $\mathbb{S}^1$ and $\mathbb{SU}(N)$. These estimates are then used in Section 3 to prove Theorem 1.1. In Section 4 we use Fourier and classical approximation methods, as well as the previously mentioned coupling argument, to give bounds on the rate of convergence of the ensemble-averaged spectral measure to the limiting measure ν_t as in Theorem 1.2. In this section, we also give the proof of Proposition 1.4 using similar methods. Finally, in Section 5, we prove a tail bound on the metric radius of the unitary Brownian motion and a continuity result for the family of measures $\{\nu_t\}_{t>0}$, which are then both used to give the proof of Theorem 1.3.

2. A concentration inequality for heat kernel measure

In this section, we will consider metric probability spaces (X, d, ρ) (where ρ is a Borel probability measure) with the following concentration property: that there exists $C > 0$ such that, for all $r > 0$ and $F : X \to \mathbb{R}$ Lipschitz with Lipschitz constant L and $\mathbb{E}|F| < \infty$,

$$\rho\left(|F - \mathbb{E}F| \geqslant r\right) \leqslant 2e^{-r^2/L^2C}. \tag{2.1}$$

Concentration estimates of this type are standard for heat kernel measure on a Riemannian manifold with curvature bounded below. We recall here the necessary results. Let (M, g) be a complete Riemannian manifold, and let Δ denote the Laplace–Beltrami operator acting on $C^\infty(M)$. We write $P_t = e^{t\Delta/2}$ to denote the heat semigroup; that is, for $t > 0$ and any sufficiently nice function $f : M \to \mathbb{R}$,

$$P_t f(x) = \mathbb{E}[f(\xi_t^x)] = \int_M f \, \mathrm{d}\rho_t^x$$

where $\{\xi_t^x\}_{t\geqslant 0}$ is the Markov diffusion on M started at x with generator Δ (that is, ξ^x is a Brownian motion on M) and $\rho_t^x = \mathrm{Law}(\xi_t^x)$ is the heat kernel measure. If Ric denotes the Ricci curvature tensor on M, then $\mathrm{Ric} \geqslant 2k$ for $k \in \mathbb{R}$ implies that for all $t > 0$ the estimate (2.1) holds for ρ_t with coefficient $C(t) = 2(1 - e^{-kt/2})/k$, where when $k = 0$, this is interpreted as $C(t) = t$. (A typical proof is via log Sobolev estimates.) See for example Corollary 2.6 and Lemma 6.3 of [Led99] (stated in the case that $k \geqslant 0$, which is the only relevant case here).

For small t the general machinery described above leads to a sharp concentration estimate for heat kernel measure ρ_t^N on $\mathbb{U}(N)$. For large t, the estimates are no longer sharp, but we can improve them using a coupling approach inspired by one in [MM13b]. The following lemma gives the key idea.

Lemma 2.1. — *Let b_t^0 be a real-valued Brownian motion and $z_t := e^{ib_t^0/N}$, and let V_t be a Brownian motion on $\mathbb{SU}(N)$ issued from the identity, with b_t^0 and V_t independent. Then $z_t V_t$ is a Brownian motion on $\mathbb{U}(N)$.*

Proof. — Set $Z_t := z_t I_N$, and note that z_t and Z_t satisfy the stochastic differential equations

$$\mathrm{d}z_t = z_t \frac{idb_t^0}{N} - \frac{1}{2N^2} z_t \,\mathrm{d}t \quad \text{and} \quad \mathrm{d}Z_t = Z_t \,\mathrm{d}b_t - \frac{1}{2N^2} Z_t \mathrm{d}t$$

where $b_t = b_t^0 \xi_0$ with $\xi_0 = iI_N/N$. Let $\{\xi_j\}_{j=1}^{N^2-1}$ be an orthonormal basis of $\mathfrak{su}(N)$, and let $\{b_t^j\}_{j=1}^{N^2-1}$ be independent real-valued Brownian motions, independent of b_t^0. Then $\widetilde{W}_t = \sum_{j=1}^{N^2-1} b_t^j \xi_j$ is a Brownian motion on $\mathfrak{su}(N)$, and we may take V_t to be the solution of the stochastic differential equation

$$\mathrm{d}V_t = V_t \circ \mathrm{d}\widetilde{W}_t = V_t \,\mathrm{d}\widetilde{W}_t + \frac{1}{2} V_t \sum_{j=1}^{N^2-1} \xi_j^2 \,\mathrm{d}t = V_t \,\mathrm{d}\widetilde{W}_t - \left(\frac{N^2-1}{2N^2}\right) V_t \,\mathrm{d}t.$$

(Here $\circ$ denotes a Stratonovich integral, which is then expressed as an Itô integral via the usual calculus.) Choosing V_t in this way, b_t^0 and V_t are indeed independent Brownian motions on $\mathbb{R}$ and $\mathbb{SU}(N)$, respectively.

Now, $\{\xi_j\}_{j=0}^{N^2-1}$ is an orthonormal basis of $\mathfrak{u}(N)$, and $z_t V_t = Z_t V_t \in \mathbb{SU}(N) \rtimes \mathbb{U}(1) \simeq \mathbb{U}(N)$ satisfies

$$\begin{aligned} d(Z_t V_t) &= \left(Z_t \mathrm{d}b_t - \frac{1}{2N^2} Z_t \,\mathrm{d}t\right) V_t + Z_t \left(V_t \,\mathrm{d}\widetilde{W}_t - \left(\frac{N^2-1}{2N^2}\right) V_t \mathrm{d}t\right) \\ &= Z_t V_t \,(\mathrm{d}b_t + \mathrm{d}\widetilde{W}_t) - \frac{1}{2} Z_t V_t \,\mathrm{d}t. \end{aligned}$$

Since $W_t = b_t + \widetilde{W}_t$ is a Brownian motion on $\mathfrak{u}(N)$, this implies that $z_t V_t$ is a Brownian motion on $\mathbb{U}(N)$. □

We use this realization of the Brownian motion on $\mathbb{U}(N)$ along with concentration properties of the laws of z_t and V_t to obtain sub-Gaussian concentration independent of t on $\mathbb{U}(N)$ for large t.

Proposition 2.2. — *Let U_t be distributed according to heat kernel measure on $\mathbb{U}(N)$, and let $F : \mathbb{U}(N) \to \mathbb{R}$ be L-Lipschitz. For any $t, r > 0$,*

$$\mathbb{P}(|F(U_t) - \mathbb{E}F(U_t)| > r) \leqslant 2e^{-\frac{r^2}{tL^2}}.$$

Furthermore, there is a constant $C \in (0, \infty)$ such that for all $t \geqslant 8(\log N)^2$ and $r > 0$

$$\mathbb{P}\left(|F(U_t) - \mathbb{E}F(U_t)| > r\right) \leqslant Ce^{-\frac{r^2}{4L^2}}.$$

Proof. — To prove the first statement, observe that since the Ricci curvature on $\mathbb{U}(N)$ is nonnegative, the comments preceding Lemma 2.1 imply that the desired concentration estimate holds for ρ_t^N with coefficient $C(t) = t$. That is, if $F : \mathbb{U}(N) \to \mathbb{R}$ is L-Lipschitz with $\mathbb{E}|F| < \infty$, then

$$\mathbb{P}(|F(U_t) - \mathbb{E}F(U_t)| > r) \leqslant 2e^{-\frac{r^2}{tL^2}}.$$

To prove the second statement, observe that the representation of U_t in Lemma 2.1 implies that

$$\begin{aligned}\mathbb{P}\left(|F(U_t)-\mathbb{E}F(U_t)|>r\right)&=\mathbb{P}\left(|F(z_tV_t)-\mathbb{E}F(z_tV_t)|>r\right)\\ &\leqslant \mathbb{E}\left[\mathbb{P}\left[\left|F(z_tV_t)-\mathbb{E}\Big[F(z_tV_t)\Big|z_t\Big]\right|>\frac{r}{2}\middle|z_t\right]\right]\\ &\quad+\mathbb{P}\left(\left|\mathbb{E}\Big[F(z_tV_t)\Big|z_t\Big]-\mathbb{E}F(z_tV_t)\right|>\frac{r}{2}\right).\end{aligned}\tag{2.2}$$

Now for the first term, measure concentration for V_t follows again from curvature considerations: following for example Proposition E.15 and Lemma F.27 of [AGZ10], one may compute the Ricci curvature on $\mathbb{SU}(N)$ with respect to the given inner product as

$$\operatorname{Ric}(X,X)=\frac{1}{2}\langle X,X\rangle_N.$$

Thus, by the discussion preceding Lemma 2.1, $\operatorname{Law}(V_t)$ on $\mathbb{SU}(N)$ satisfies the following concentration estimate: if $G:\mathbb{SU}(N)\to\mathbb{R}$ is L-Lipschitz, then

$$\mathbb{P}(|G(V_t)-\mathbb{E}G(V_t)|>r)\leqslant 2e^{-\frac{r^2}{L^2c(t)}},$$

where $c(t):=4(1-e^{-t/4})$. For z_t fixed, $G=F(z_t\cdot)$ is an L-Lipschitz function on $\mathbb{SU}(N)$, and so the first term of (2.2) is bounded by $2e^{-\frac{r^2}{4L^2}}$.

For the second term of (2.2), let $K=K(z_t)$ be the random variable taking values in $\{0,\dots,N-1\}$ such that, on $\{K=k\}$, $z_t\in[e^{\frac{2\pi ik}{N}},e^{\frac{2\pi i(k+1)}{N}})$. Conditioning on K, we have

$$\begin{aligned}&\mathbb{P}\left(\left|\mathbb{E}\Big[F(z_tV_t)\Big|z_t\Big]-\mathbb{E}F(z_tV_t)\right|>\frac{r}{2}\right)\\ &\quad=\mathbb{E}\left(\mathbb{P}\left[\left|\mathbb{E}\Big[F(z_tV_t)\Big|z_t\Big]-\mathbb{E}F(z_tV_t)\right|>\frac{r}{2}\middle|K\right]\right)\\ &\quad\leqslant\mathbb{E}\left(\mathbb{P}\left[\left|\mathbb{E}\Big[F(z_tV_t)\Big|z_t\Big]-\mathbb{E}\Big[F(z_tV_t)\Big|K\Big]\right|>\frac{r}{4}\middle|K\right]\right)\\ &\qquad+\mathbb{P}\left(\left|\mathbb{E}\Big[F(z_tV_t)\Big|K\Big]-\mathbb{E}F(z_tV_t)\right|>\frac{r}{4}\right).\end{aligned}\tag{2.3}$$

For the first term in (2.3), let $\mathbb{E}_{V_t}$ denote integration over V_t only, $\mathbb{E}_{z_t}$ integration over z_t only, and let $\mathbb{E}_{z_t|K=k}$ denote integration over z_t conditional on $K=k$. Observe that by independence of V_t and z_t

$$\begin{aligned}&\left|\mathbb{E}\Big[F(z_tV_t)\Big|z_t\Big]-\mathbb{E}\Big[F(z_tV_t)\Big|K=k\Big]\right|\\ &\quad=\left|\mathbb{E}_{V_t}\left[F(z_tV_t)\right]-\mathbb{E}_{V_t}\mathbb{E}_{z_t|K=k}\left[F(z_tV_t)\right]\right|\\ &\quad\leqslant\mathbb{E}_{V_t}\left|F(z_tV_t)-\mathbb{E}_{z_t|K=k}\left[F(z_tV_t)\right]\right|\\ &\quad=\int_{\mathbb{SU}(N)}\left|F(z_tV)-\mathbb{E}_{z_t|K=k}[F(z_tV)]\right|h_t^{\mathbb{SU}(N)}(V)\mathrm{d}V,\end{aligned}$$

where $h_t^{\mathbb{SU}(N)}$ denotes the density of V_t with respect to Haar measure on $\mathbb{SU}(N)$. Now, for V fixed, $F(\cdot V)$ is an NL-Lipschitz function on $\mathbb{S}^1$. So, conditioned on $K=k$,

$F(z_tV)$ can only fluctuate by as much as $2\pi L$. Thus if $\frac{r}{4} > 2\pi L$, the first term is zero. For $\frac{r}{4} \leqslant 2\pi L$, we may just use the trivial bound of 1 and choose C in the statement of the proposition so that $C \geqslant e^{(8\pi)^2/4}$.

For the second term in (2.3), note that we can replace V_t with a Haar-distributed random matrix V for t sufficiently large. Indeed, letting $\mathrm{d}V$ denote integration with respect to Haar measure on $\mathbb{SU}(N)$, and assuming without loss in generality that $F(I_N) = 0$,

$$\begin{aligned}\left|\mathbb{E}\left[F(z_tV_t) - F(z_tV)\middle|z_t\right]\right| &\leqslant \int_{\mathbb{SU}(N)} \left|F(z_tV)\right|\left|h_t^{\mathbb{SU}(N)}(V) - 1\right|\mathrm{d}V \\ &\leqslant LN\|h_t^{\mathbb{SU}(N)} - 1\|_1,\end{aligned} \tag{2.4}$$

since the diameter of $\mathbb{U}(N)$ is N. A sharp estimate of the time to equilibrium of V_t was proved in Theorem 1.2 of [SC94], from which it follows (see the discussion preceding the theorem in [SC94], and note that the normalization here differs by a factor of 2 from the one used there) that

$$\|h_t^{\mathbb{SU}(N)} - 1\|_1 \leqslant e^{-\frac{t(1+o(1))}{8\log N}}. \tag{2.5}$$

Thus if $t \geqslant 8(\log N)^2$, replacing V_t by V will only affect the constants.

Consider therefore

$$\mathbb{P}\left[\left|\mathbb{E}\left[F(z_tV)\middle|K\right] - \mathbb{E}[F(z_tV)]\right| > \frac{r}{4}\right],$$

and write $z_t = \omega_t e^{\frac{2\pi iK}{N}}$, with ω_t in the arc from 1 to $e^{\frac{2\pi i}{N}}$.

Observe that, by Fubini's theorem and the translation invariance of Haar measure on $\mathbb{SU}(N)$,

$$\mathbb{E}[F(z_tV)] = \mathbb{E}_{z_t}\mathbb{E}_V[F(\omega_t e^{\frac{2\pi iK}{N}}V)] = \mathbb{E}_{z_t}\mathbb{E}_V[F(\omega_tV)] = \mathbb{E}_V\mathbb{E}_{z_t}[F(\omega_tV)],$$

and similarly

$$\begin{aligned}\mathbb{E}[F(z_tV)|K=k] &= \mathbb{E}_{z_t|K=k}\mathbb{E}_V[F(\omega_t e^{\frac{2\pi iK}{N}}V)] \\ &= \mathbb{E}_{z_t|K=k}\mathbb{E}_V[F(\omega_tV)] = \mathbb{E}_V\mathbb{E}_{z_t|K=k}[F(\omega_tV)].\end{aligned}$$

Thus

$$\left|\mathbb{E}\left[F(z_tV)\middle|K\right] - \mathbb{E}[F(z_tV)]\right| \leqslant \mathbb{E}_V\left|\mathbb{E}_{z_t}[F(\omega_tV)] - \mathbb{E}_{z_t|K}[F(\omega_tV)]\right| \leqslant 2\pi L,$$

where we have used again that for fixed V, $F(\omega V)$ is an NL-Lipschitz function of ω, and here ω lies within an arc of length $\frac{2\pi}{N}$. The estimate now follows as in the first term. □

3. Concentration of μ_t^N

Armed with the concentration inequality for heat kernel measure, the proof of Theorem 1.1 is an application of the program laid out in [MM17] for estimating the Wasserstein distance between the empirical spectral measure of a random matrix and the ensemble average, in the presence of measure concentration. Since it is relatively brief, we include the detailed argument here for completeness.

The first step is to bound the "average distance to average" $\mathbb{E}W_1(\mu_t^N,\overline{\mu}_t^N)$ as follows.

Proposition 3.1. — *There is a constant $c \in (0,\infty)$ such that for all $N \in \mathbb{N}$ and $t > 0$*

$$\mathbb{E}W_1(\mu_t^N,\overline{\mu}_t^N) \leqslant c\left(\frac{t}{N^2}\right)^{1/3},$$

and for all $N \in \mathbb{N}$ and $t \geqslant 8(\log N)^2$

$$\mathbb{E}W_1(\mu_t^N,\overline{\mu}_t^N) \leqslant \frac{c}{N^{2/3}}.$$

Proof. — We will give the proof of the first statement only, which applies the first half of Proposition 2.2; the proof of the second statement is identical using only instead the second half of Proposition 2.2.

Recall that

$$W_1(\mu_t^N,\overline{\mu}_t^N) = \sup_{|f|_L \leqslant 1}\left(\int f\mathrm{d}\mu_t^N - \int f\mathrm{d}\overline{\mu}_t^N\right),$$

where $|f|_L \leqslant 1$. That is, our task is to estimate the expected supremum of the centered stochastic process $\{X_f\}_{|f|_L\leqslant 1}$, with

$$X_f := \int f\mathrm{d}\mu_t^N - \int f\mathrm{d}\overline{\mu}_t^N = \int f\mathrm{d}\mu_t^N - \mathbb{E}\int f\mathrm{d}\mu_t^N.$$

Note that without loss we may choose the indexing set to be 1-Lipschitz functions on the circle with $f(1) = 0$; write $Lip_0(1)$ for the set of all such functions. Now, if f is a fixed Lipschitz function and μ_U denotes the spectral measure of U, then

$$U \longmapsto \left(\int f\mathrm{d}\mu_U - \int f\mathrm{d}\overline{\mu}_t^N\right)$$

is $\frac{|f|_L}{N}$-Lipschitz (see Lemma 2.3 of [MM13a], and note that the different normalization of the metric on matrices used there means that the Lipschitz constant quoted there must be multiplied by $N^{-\frac{1}{2}}$ in the present context), and so by Proposition 2.2,

$$\mathbb{P}\left(|X_f - X_g| > x\right) = \mathbb{P}\left(|X_{f-g}| > x\right) \leqslant 2e^{-\frac{N^2x^2}{t|f-g|_L^2}}.$$

That is, the stochastic process $\{X_f\}_{f\in Lip_0(1)}$ satisfies a sub-Gaussian increment condition.

Now, if $\{X_v\}_{\|v\|=1}$ is a centered stochastic process indexed by the unit ball of a *finite-dimensional* normed space V, and $\{X_v\}$ satisfies the increment condition

$$\mathbb{P}(|X_u - X_v| > x) \leqslant ae^{-\frac{x^2}{K^2\|u-v\|^2}}$$

for each $x > 0$, then it is a consequence of Dudley's entropy bound (see [MM17] for a detailed proof) that

$$\mathbb{E}\left(\sup_{\|v\|=1} X_v\right) \leqslant aK\sqrt{\dim V}. \tag{3.1}$$

The index set $Lip_0(1)$ is the unit ball of an infinite-dimensional normed space, but Lipschitz test functions may be approximated by piecewise linear functions coming from a finite-dimensional space. Specifically, for $m \in \mathbb{N}$, let $A_0^{(m)}$ be the set of $f : [0, 2\pi] \to \mathbb{R}$ such that

- $f(0) = f(2\pi) = 0$,
- $|f|_L \leqslant 1$, and
- f is piecewise linear, with changes in slope occurring only at the values $\frac{2\pi k}{m}$, $1 \leqslant k \leqslant m-1$.

For any $f \in Lip_0(1)$, there is $f^{(m)} \in A_0^{(m)}$ such that $\|f - f^{(m)}\|_\infty \leqslant \frac{\pi}{m}$, and so

$$\left|X_f - X_{f^{(m)}}\right| = \left|\int (f - f^{(m)}) \mathrm{d}\mu_t^N - \int (f - f^{(m)}) \mathrm{d}\overline{\mu}_t^N\right| \leqslant \frac{2\pi}{m}.$$

The space of functions for which $A_0^{(m)}$ is the unit ball is $(m-1)$-dimensional, and so it follows from (3.1) that

$$\begin{aligned}\mathbb{E}\left(\sup_{f \in Lip_0(1)} X_f\right) &\leqslant \frac{2\pi}{m} + \mathbb{E}\left(\sup_{f \in A_0^{(m)}} X_f\right) \\ &\leqslant \frac{2\pi}{m} + C'\left(\frac{\sqrt{t}}{N}\right)\sqrt{m-1}.\end{aligned}$$

Choosing $m = \left(\frac{N^2}{t}\right)^{1/3}$ completes the proof. $\square$

The proof of Theorem 1.1 is completed via the concentration of $W_1(\mu_t^N, \overline{\mu}_t^N)$ about its mean, as follows.

Proposition 3.2. — *For all $t > 0$, $N \in \mathbb{N}$, and $x > 0$,*

$$\mathbb{P}\left(W_1(\mu_t^N, \overline{\mu}_t^N) > \mathbb{E}W_1(\mu_t^N, \overline{\mu}_t^N) + x\right) \leqslant 2e^{-\frac{N^2x^2}{t}},$$

and there exists $C \in (0, \infty)$ such that for all $t \geqslant 8(\log N)^2$, $N \in \mathbb{N}$, and $x > 0$,

$$\mathbb{P}\left(W_1(\mu_t^N, \overline{\mu}_t^N) > \mathbb{E}W_1(\mu_t^N, \overline{\mu}_t^N) + x\right) \leqslant Ce^{-N^2x^2/4}.$$

Proof. — Again, we prove only the first statement and the proof of the second is analogous.

Consider the mapping $F : \mathbb{U}(N) \to \mathbb{R}$ given by

$$F(U) = W_1(\mu_U, \overline{\mu}_t^N),$$

where μ_U is the spectral measure of U and $\overline{\mu}_t^N$ is the ensemble-averaged empirical spectral measure of U_t^N as before. The function F is a $\frac{1}{N}$-Lipschitz function of U (again, see Lemma 2.3 of [MM13a]), and so by Proposition 2.2, for all $t > 0$ and all $x > 0$,

$$\mathbb{P}\left(W_1(\mu_t^N, \overline{\mu}_t^N) - \mathbb{E}W_1(\mu_t^N, \overline{\mu}_t^N) > x\right) \leqslant 2e^{-\frac{N^2x^2}{t}}.$$

$\square$

From the tail estimate of Proposition 3.2 together with Proposition 3.1, it follows that for any $t, x > 0$,

$$\mathbb{P}\left(W_1(\mu_t^N, \overline{\mu}_t^N) > c\left(\frac{t}{N^2}\right)^{1/3} + x\right) \leqslant 2e^{-\frac{N^2x^2}{t}}. \tag{3.2}$$

In particular, an application of the Borel–Cantelli lemma with $x_N = c\left(\frac{t}{N^2}\right)^{1/3}$ completes the proof of the first statement of Theorem 1.1. The second statement follows in the same way.

4. Convergence to ν_t

The previous section established a bound on the distance between the (random) spectral measure μ_t^N and the ensemble average $\overline{\mu}_t^N$. The picture is completed by obtaining a rate of convergence of $\overline{\mu}_t^N$ to the limiting measure ν_t. The following is relevant for moderate t.

Theorem 4.1. — *There is a constant $C \in (0,\infty)$ such that for all $N \in \mathbb{N}$ and $t > 0$*

$$W_1(\overline{\mu}_t^N, \nu_t) \leqslant C\frac{t^{2/5}\log N}{N^{2/5}}.$$

Proof. — The proof is via Fourier analysis and classical approximation theory, following the approach of Theorem 2.1 in [MM13a]. The key ingredient of this proof is the bound (4.1) below, which was proved in [CDK18].

Given $f : \mathbb{S}^1 \to \mathbb{R}$ a 1-Lipschitz function, let

$$S_m(z) := \sum_{|k|<m} \hat{f}(k) z^k,$$

where $\hat{f}(k) = \int_{\mathbb{S}^1} f(z) z^{-k} \mathrm{d}\nu(z)$, with ν denoting the uniform probability measure on $\mathbb{S}^1$. Observe that

$$\int z^k \mathrm{d}\overline{\mu}_t^N = \frac{1}{N}\mathbb{E}[\mathrm{tr}(U_t^k)]$$

where U_t is a Brownian motion on $\mathbb{U}(N)$. Since f is 1-Lipschitz, it is known that $|\hat{f}(k)| \leqslant \frac{C}{k}$ for $k \geqslant 1$ (in fact, $C = \frac{\pi}{2}$; see, for example, Theorem 4.6 of [Kat04]), and so

$$\left|\int S_m \,\mathrm{d}\overline{\mu}_t^N - \int S_m \,\mathrm{d}\nu_t\right| = \left|\sum_{1\leqslant|k|<m} \hat{f}(k)\left(\frac{1}{N}\mathbb{E}[\mathrm{tr}(U_t^k)] - \int z^k \,\mathrm{d}\nu_t\right)\right|$$

Now, by Theorem 1.3 of [CDK18], for t and k fixed,

$$\left|\frac{1}{N}\mathbb{E}[\mathrm{tr}(U_t^k)] - \int z^k \mathrm{d}\nu_t\right| \leqslant \frac{t^2k^4}{N^2}. \tag{4.1}$$

Thus,

$$\left|\int S_m \,\mathrm{d}\overline{\mu}_t^N - \int S_m \,\mathrm{d}\nu_t\right| \leqslant C\sum_{1\leqslant|k|<m} \frac{1}{k}\frac{t^2k^4}{N^2} \leqslant C\frac{t^2m^4}{N^2}.$$

The proof now proceeds exactly as in Theorem 2.1 of [MM13a]. A theorem of Lebesgue implies that

$$\|f - S_m\|_\infty \leqslant C' \log m \left(\inf_g \|f - g\|_\infty\right)$$

where the infimum is over all trigonometric polynomials $g(z) = \sum_{|k|<m} a_k z^k$; see for example Theorem 2.2 of [Riv81]. Combining this with Jackson's theorem (Theorem 1.4 of the same reference) implies that $\|f - S_m\|_\infty \leqslant C' \frac{\log m}{m}$, and thus

$$\begin{aligned}\left|\int f \,\mathrm{d}\overline{\mu}_t^N - \int f \,\mathrm{d}\nu_t\right| &\leqslant \left|\int f \,\mathrm{d}\overline{\mu}_t^N - \int S_m \,\mathrm{d}\overline{\mu}_t^N\right| + \left|\int S_m \,\mathrm{d}\overline{\mu}_t^N - \int S_m \,\mathrm{d}\nu_t\right| \\ &\quad + \left|\int S_m \,\mathrm{d}\nu_t - \int f \,\mathrm{d}\nu_t\right| \\ &\leqslant C'' \left(\frac{\log m}{m} + \frac{t^2 m^4}{N^2}\right).\end{aligned}$$

Choosing $m = (N/t)^{2/5}$ then gives the stated bound. □

The bound above decays if and only if $t = o(N/((\log N)^{5/2}))$. But for sufficiently large t, both $\overline{\mu}_t^N$ and ν_t are close to the uniform measure on the circle. This is not reflected in the bound above, which gets worse for large t. The following propositions treat the large t case by appealing to convergence to stationarity.

Proposition 4.2. — *Let $\overline{\mu}_t^N$ denote the ensemble-averaged spectral measure of a random matrix U_t distributed according to heat kernel measure on $\mathbb{U}(N)$, and let ν denote the uniform probability measure on $\mathbb{S}^1$. There are constants $C, c \in (0,\infty)$ so that for all $N \in \mathbb{N}$ and $t > 0$*

$$W_1(\overline{\mu}_t^N, \nu) \leqslant e^{-\frac{t(1+o(1))}{8\log(N)}} + \frac{2\pi}{N}.$$

Proof. — First recall again that, as in the proof of Proposition 3.1, if μ_U denotes the spectral measure of U, then for fixed $f : \mathbb{S}^1 \to \mathbb{R}$ with $|f|_L \leqslant 1$, the function

$$F(U) = \int f \mathrm{d}\mu_U$$

is $\frac{1}{N}$-Lipschitz on $\mathbb{U}(N)$. Since ν is the spectral measure of a Haar-distributed random unitary matrix U on $\mathbb{U}(N)$, this means that

$$\int f \mathrm{d}\mu_t^N - \int f \mathrm{d}\nu \leqslant \frac{\|U_t - U\|_N}{N},$$

where $\|\cdot\|_N$ is the norm induced by the scaled inner product $\langle\cdot,\cdot\rangle_N$, and this holds for any coupling (U_t, U) of heat kernel measure and Haar measure. Taking expectation gives

$$\int f \mathrm{d}\overline{\mu}_t^N - \int f \mathrm{d}\nu = \mathbb{E}\left(\int f \mathrm{d}\mu_t^N - \int f \mathrm{d}\nu\right) \leqslant \frac{\mathbb{E}\|U_t - U\|_N}{N}.$$

Taking the supremum over f gives that

$$W_1(\overline{\mu}_t^N, \nu) \leqslant \frac{\mathbb{E}\|U_t - U\|_N}{N},$$

and now taking infimum over couplings we have

$$W_1(\overline{\mu}_t^N, \nu) \leqslant \inf_{(U_t,U)} \frac{\mathbb{E}\|U_t - U\|_N}{N} = \frac{1}{N} W_1(U_t, U), \tag{4.2}$$

where $W_1(U_t, U)$ now denotes the L_1-Wasserstein distance between the heat kernel and Haar measures on $\mathbb{U}(N)$.

Now consider the coupling $U_t \stackrel{d}{=} z_t V_t$ from Lemma 2.1, where $z_t = e^{ib_t^0/N}$ for b_t^0 a standard Brownian motion on $\mathbb{R}$ and V_t an independent Brownian motion on $\mathbb{SU}(N)$ with $V_0 = I_N$. One can similarly obtain Haar measure on the unitary group from uniform measure on an interval and Haar measure on $\mathbb{SU}(N)$: if $z = e^{i\theta/N}$ with θ uniform in $[0, 2\pi)$ and V is independent of θ and distributed according to Haar measure on $\mathbb{SU}(N)$, then zV is distributed according to Haar measure on $\mathbb{U}(N)$; see for example Lemma 16 of [MM13b]. Moreover, by the translation invariance of Haar measure, θ could also be distributed uniformly on $[2\pi k, 2\pi(k+1))$ for any $k \in \mathbb{Z}$, or indeed be distributed according to any mixture of uniform measure on such intervals, as long as the mixing measure is independent of V.

Given any such z_t, z, V_t, and V, for any $F : \mathbb{U}(N) \to \mathbb{R}$ a 1-Lipschitz function, we have that

$$\begin{aligned}(4.3)\qquad \left|\mathbb{E}F(U_t) - \mathbb{E}F(U)\right| &= \left|\mathbb{E}F(z_tV_t) - \mathbb{E}F(zV)\right| \\ &\leqslant \mathbb{E}\left|\mathbb{E}\left[F(z_tV_t) - F(z_tV)\middle|z_t\right]\right| + \left|\mathbb{E}\left[F(z_tV) - F(zV)\right]\right|\end{aligned}$$

The first term of (4.3) was already bounded in the course of the proof of Proposition 2.2:

$$(4.4)\qquad \mathbb{E}\left|\mathbb{E}\left[F(z_tV_t) - F(z_tV)\middle|z_t\right]\right| \leqslant Ne^{-\frac{t(1+o(1))}{8\log(N)}}.$$

To treat the second term, we may as in the proof of Proposition 2.2 write $z_t = \omega_t e^{\frac{2\pi i K}{N}}$, with ω_t in the arc from 1 to $e^{\frac{2\pi i}{N}}$ and $K \in \{0, \dots, N-1\}$, and similarly $z = \omega e^{\frac{2\pi i K}{N}}$ the second term of (4.3) can be bounded as

$$\begin{aligned}\mathbb{E}\left[F(z_tV) - F(zV)\right] &= \mathbb{E}\left[F(\omega_t e^{\frac{2\pi i K_t}{N}}V)\right] - \mathbb{E}\left[F(\omega e^{\frac{2\pi i K}{N}}V)\right] \\ &= \mathbb{E}\left[F(\omega_tV) - F(\omega V)\right] \leqslant \frac{2\pi}{N}\cdot N,\end{aligned}$$

where the second equality follows from the independence of V with (z, z_t) and Fubini's theorem, and the inequality uses the fact that, for V fixed, $F(\omega V)$ is N-Lipschitz as a function of ω, with ω, ω_t lying in an arc of length $\frac{2\pi}{N}$.

Combining this last estimate with (4.2), (4.3), and (4.4) implies that

$$\begin{aligned}W_1(\overline{\mu}_t^N, \nu) &\leqslant \frac{1}{N}W_1(U_t, U) \\ &= \frac{1}{N}\sup_{|F|_L\leqslant 1}\left|\mathbb{E}F(U_t) - \mathbb{E}F(U)\right| \\ &\leqslant e^{-\frac{t(1+o(1))}{8\log N}} + \frac{2\pi}{N}.\end{aligned}$$

□

Finally, we compare the limiting (large N) measure ν_t to the uniform measure ν. We restate and prove here Proposition 1.4.

Proposition 1.4. — *For ν_t and ν defined as above, there is a constant $C \in (0, \infty)$ so that for all $t \geqslant 1$*

$$W_1(\nu_t, \nu) \leqslant Ct^{3/2}e^{-t/4}.$$

Observe in particular that for large t, $t^{3/2}e^{-t/4} \leqslant e^{-\frac{t}{8\log(N)}}$, and so Theorem 1.2 follows from Theorem 4.1 and Propositions 1.4 and 4.2 together with the triangle inequality.

Proof of Proposition 1.4. — The measure ν_t is symmetric, and the moments of ν_t for $k \geqslant 1$ are given by

$$\int_{\mathbb{S}^1} z^k \mathrm{d}\nu_t(z) = Q_k(t)e^{-\frac{kt}{2}},$$

where

$$Q_k(t) := \sum_{j=0}^{k-1} \frac{(-tk)^j}{(j+1)!}\binom{k-1}{j};$$

see [Bia97a]. As in the proof of Theorem 4.1, for a fixed 1-Lipschitz test function $f : \mathbb{S}^1 \to \mathbb{R}$, let

$$S_m(z) := \sum_{|k|<m} \hat{f}(k)z^k$$

and we have that $|\hat{f}(k)| \leqslant \frac{C}{k}$ for all $k \geqslant 1$. Then since both ν_t and ν are probability measures on $\mathbb{S}^1$ and $\int_{\mathbb{S}^1} z^j \mathrm{d}\nu(z) = 0$ if $j \neq 0$,

$$\begin{aligned}\left|\int S_m(z)\,\mathrm{d}\nu_t(z) - \int S_m(z)\,\mathrm{d}\nu(z)\right| &= \left|\sum_{1\leqslant|k|\leqslant m} \hat{f}(k)\int z^k\,\mathrm{d}\nu_t(z)\right| \\ &\leqslant C\sum_{1\leqslant k\leqslant m} \frac{1}{k}\left|Q_k(t)\right|e^{-\frac{kt}{2}}. \end{aligned} \tag{4.5}$$

Let

$$A_k(t) := Q_k(-t) = \sum_{j=0}^{k-1} \frac{(tk)^j}{(j+1)!}\binom{k-1}{j},$$

so that $|Q_k(t)| \leqslant A_k(t)$. Now,

$$\begin{aligned} A_{k+1}(t) &= 1 + \sum_{j=1}^{k} \frac{[t(k+1)]^j}{(j+1)!}\binom{k}{j} \\ &= 1 + tk(k+1)\sum_{j=1}^{k}\left(\frac{\left(1+\frac{1}{k}\right)^{j-1}}{j(j+1)}\right)\left[\frac{(tk)^{j-1}}{k(j-1)!}\binom{k}{j}\right] \end{aligned}$$

and note that

$$A_k(t) = \sum_{j=0}^{k-1} \frac{(tk)^j}{(j+1)!}\binom{k-1}{j} = \sum_{j=0}^{k-1} \frac{(tk)^j}{kj!}\binom{k}{j+1} = \sum_{\ell=1}^{k} \frac{(tk)^{\ell-1}}{k(\ell-1)!}\binom{k}{\ell}.$$

Since $\frac{\left(1+\frac{1}{k}\right)^{\ell-1}}{\ell(\ell+1)}$ is decreasing as a function of ℓ on $\{1,\ldots,k\}$, it follows that

$$A_{k+1}(t) \leqslant 1 + \left(\frac{tk(k+1)}{2}\right)A_k(t) \leqslant tk(k+1)A_k(t),$$

since $t, k \geqslant 1$. By induction and the fact that $A_1(t) = 1$, this implies that

$$|Q_k(t)| \leqslant A_k(t) \leqslant t^{k-1}k[(k-1)!]^2.$$

It now follows from (4.5) that

$$\left|\int S_m(z)\,\mathrm{d}\nu_t(z) - \int S_m(z)\,\mathrm{d}\nu(z)\right| \leqslant \sum_{k=1}^{m} t^{k-1}[(k-1)!]^2 e^{-\frac{kt}{2}}$$
$$\leqslant e^{-t/2}\sum_{k=1}^{m}\left(t(k-1)^2 e^{-t/2}\right)^{k-1} \leqslant e^{-t/2}\sum_{k=1}^{m}\left(tm^2 e^{-t/2}\right)^{k-1}.$$

Choose $m = \left\lfloor \frac{1}{\sqrt{2t}} e^{t/4} \right\rfloor$, so that $tm^2 e^{-t/2} \leqslant \frac{1}{2}$. Then

$$\left|\int S_m(z)\,\mathrm{d}\nu_t(z) - \int S_m(z)\,\mathrm{d}\nu(z)\right| \leqslant 2e^{-t/2}.$$

As in the proof of Theorem 4.1, we have that $\|S_m - f\|_\infty \leqslant C' \frac{\log m}{m}$, which for the chosen value of m yields

$$\|S_m - f\|_\infty \leqslant C'' t^{3/2} e^{-t/4}.$$

Combining these estimates completes the proof. □

5. Convergence of paths

This section is devoted to the proof of Theorem 1.3. The idea is to first discretize the interval $[0, T]$ and apply the bound from Proposition 3.2 at the discretization points, then move from approximation at this discrete set of points to approximation along an entire path via a continuity property of the family of measures $\{\nu_t\}_{t>0}$.

The following tail bound is used in both parts of the argument.

Proposition 5.1. — *Let $\{U_t\}_{t\geqslant 0}$ denote Brownian motion in $\mathbb{U}(N)$ with $U_0 = I_N$, and let d_g denote the geodesic distance on $\mathbb{U}(N)$ induced by $\langle\cdot,\cdot\rangle_N$. Then for all $\delta, r, s > 0$,*

$$\mathbb{P}\left(\sup_{0<t<\delta} d_g(U_t, I_N) \geqslant r + 2s\right) \leqslant 16\left(1 + \frac{r}{s}\right)^{N^2} e^{-\frac{r^2}{2\delta}}.$$

Proof. — For any $U \in \mathbb{U}(N)$ such that $d_g(U, I_n) < s$, left invariance of the metric and the triangle inequality imply that

$$d_g(U_t, I_N) = d_g(UU_t, U) \leqslant d_g(UU_t, I_n) + s.$$

Thus,

$$\mathbb{P}\left(\sup_{0<t<\delta} d_g(U_t, I_N) \geqslant 2s + r\right) \leqslant \inf_{d_g(U,I)\leqslant s} \mathbb{P}\left(\sup_{0<t<\delta} d_g(UU_t, I_N) \geqslant s + r\right).$$

Applying the bound in Equation (9.20) of [Gri99] with $M = \mathbb{U}(N)$ and $K = \overline{B(I_N, s)}$ (the closed geodesic ball of radius s about I_N) gives that

$$\inf_{d_g(U,I)\leqslant s} \mathbb{P}\left(\sup_{0<t<\delta} d_g(UU_t, I_N) \geqslant s + r\right) \leqslant 16 \frac{\mathrm{vol}(B(I_N, s + r))}{\mathrm{vol}(B(I_N, s))} e^{-\frac{r^2}{2\delta}}.$$

Then, recalling again that $\mathrm{Ric} \geqslant 0$ on $\mathbb{U}(N)$, the Bishop–Gromov comparison theorem allows us to control the volume of balls in $\mathbb{U}(N)$ by the volume of balls in $\mathbb{R}^{N^2}$ (see for example Theorem 3.16 of [Gro87]); in particular,

$$\frac{\mathrm{vol}\left(B\left(I_N, s+r\right)\right)}{\mathrm{vol}\left(B\left(I_N, s\right)\right)} \leqslant \left(1+\frac{r}{s}\right)^{N^2},$$

which completes the proof. □

The following lemma gives the required continuity for the family of measures $\{\nu_t\}$.

Lemma 5.2. — *For all* $0 < s < t$

$$W_1(\nu_t, \nu_s) \leqslant 3\sqrt{t-s}.$$

Proof. — The triangle inequality for W_1 and Theorem 4.1 imply that for any N

$$\begin{aligned} W_1(\nu_t, \nu_s) &\leqslant W_1(\nu_t, \overline{\mu}_t^N) + W_1(\nu_s, \overline{\mu}_s^N) + W_1(\overline{\mu}_t^N, \overline{\mu}_s^N) \\ &\leqslant C\frac{(t^{2/5}+s^{2/5})\log N}{N^{2/5}} + W_1(\overline{\mu}_t^N, \overline{\mu}_s^N). \end{aligned}$$

Moreover, recall that

$$W_1(\overline{\mu}_t^N, \overline{\mu}_s^N) = \sup_{|f|_L \leqslant 1} \mathbb{E}\left[\int f \mathrm{d}\mu_t^N - \int f \mathrm{d}\mu_s^N\right] \leqslant \frac{\mathbb{E}\|U_t - U_s\|_N}{N},$$

since $U \mapsto \int f \mathrm{d}\mu_U$ is $\frac{|f|_L}{N}$-Lipschitz. Trivially, for any $U, V \in \mathbb{U}(N)$, $\|U-V\|_N \leqslant d_g(U,V)$. So, using the stationarity of increments together with Proposition 5.1 with $r = 2s = \frac{c}{2}N\sqrt{t-s}$,

$$\begin{aligned} \mathbb{E}\|U_t - U_s\|_N = \mathbb{E}\|I_N - U_{t-s}\|_N &\leqslant \mathbb{E}d_g(I_N, U_{t-s}) \\ &\leqslant cN\sqrt{t-s} + N\mathbb{P}\left(d_g(I, U_{t-s}) > cN\sqrt{t-s}\right) \\ &\leqslant cN\sqrt{t-s} + 3^{N^2}Ne^{-c^2N^2/8}. \end{aligned}$$

Choosing $c = 3$ implies that $\log 3 + \frac{\log N}{N^2} - \frac{c^2}{8} < 0$ for all $N \geqslant 4$, so that for $N \geqslant 4$,

$$\mathbb{E}\|U_t - U_s\|_N \leqslant 3N\sqrt{t-s} + 1$$

and thus

$$W_1(\nu_t, \nu_s) \leqslant C\frac{(t^{2/5}+s^{2/5})\log N}{N^{2/5}} + 3\sqrt{t-s} + \frac{1}{N}.$$

Letting $N \to \infty$ completes the proof. □

Proof of Theorem 1.3. — Let $m \in \mathbb{N}$, and for $j = 1, \ldots, m$, let $t_j := \frac{jT}{m}$.
By Lemma 5.2,

$$\sup_{\substack{0 \leqslant s,t \leqslant T \\ |s-t| < \frac{T}{m}}} W_1(\nu_t, \nu_s) \leqslant 3\sqrt{\frac{T}{m}},$$

so that if $x > 9\sqrt{\frac{T}{m}}$, then

$$\mathbb{P}\left(\sup_{0\leqslant t\leqslant T} W_1(\mu_t^N,\nu_t) > x\right) \leqslant \mathbb{P}\left(\max_{1\leqslant j\leqslant m}\sup_{|t-t_j|<\frac{T}{m}} W_1(\mu_t^N,\mu_{t_j}^N) > \frac{x}{3}\right) + \mathbb{P}\left(\max_{1\leqslant j\leqslant m} W_1(\mu_{t_j}^N,\nu_{t_j}) > \frac{x}{3}\right). \tag{5.1}$$

Using again that $W_1(\mu_t^N,\mu_s^N) \leqslant \frac{\|U_t-U_s\|_N}{N}$, we have that for any $A\subseteq[0,T]^2$

$$\begin{aligned}\mathbb{P}\left(\sup_{(s,t)\in A} W_1(\mu_t^N,\mu_s^N) > \frac{x}{3}\right) &\leqslant \mathbb{P}\left(\sup_{(s,t)\in A}\|U_t-U_s\| > \frac{Nx}{3}\right)\\ &= \mathbb{P}\left(\sup_{(s,t)\in A}\|I_N-U_t^{-1}U_s\| > \frac{Nx}{3}\right)\\ &= \mathbb{P}\left(\sup_{(s,t)\in A}\|I_N-U_{t-s}\| > \frac{Nx}{3}\right),\end{aligned}$$

where the first equality is because $U_t\in\mathbb{U}(N)$ and the second is by the stationarity of the increments of Brownian motion. It follows from this and (5.1) that

$$\mathbb{P}\left(\sup_{0\leqslant t\leqslant T} W_1(\mu_t^N,\nu_t) > x\right) \leqslant m\mathbb{P}\left(\sup_{|t|<\frac{T}{m}}\|I_N-U_t\| > \frac{Nx}{3}\right) + m\max_{1\leqslant j\leqslant m}\mathbb{P}\left(W_1(\mu_{t_j}^N,\nu_{t_j}) > \frac{x}{3}\right).$$

Applying Proposition 5.1 to the first term with $2s = r = \frac{Nx}{6}$ gives that

$$\begin{aligned}\mathbb{P}\left(\sup_{|t|<\frac{T}{m}}\|I_N-U_t\|_N > \frac{Nx}{3}\right) &\leqslant \mathbb{P}\left(\sup_{|t|<\frac{T}{m}} d_g(U_t,I_N) > \frac{Nx}{3}\right)\\ &\leqslant 3^{N^2}e^{-\frac{N^2x^2m}{72T}}.\end{aligned}$$

For the second term, applying the estimate (3.2) together with Theorem 4.1, if $x \geqslant 3C\frac{T^{2/5}\log(N)}{N^{2/5}} > 18\left(\frac{T}{N^2}\right)^{1/3}$, then

$$\max_{1\leqslant j\leqslant m}\mathbb{P}\left(W_1(\mu_{t_j}^N,\nu_{t_j}) > \frac{x}{3}\right) \leqslant 2e^{-\frac{N^2x^2}{T}}.$$

We thus have that, for any $m\in\mathbb{N}$ and $x \geqslant 3C\frac{T^{2/5}\log(N)}{N^{2/5}}$,

$$\mathbb{P}\left(\sup_{0\leqslant t\leqslant T} W_1(\mu_t^N,\nu_t) > x\right) \leqslant m3^{N^2}e^{-\frac{N^2x^2m}{72T}} + 2me^{-\frac{N^2x^2}{T}}.$$

Choosing $m = 81\left(\frac{T}{x^2}+1\right)$ (note that the condition $x > 9\sqrt{\frac{T}{m}}$ appearing prior to equation 5.1 is then satisfied) completes the proof of the first claim; the second follows by choosing $x = 3C\frac{T^{2/5}\log(N)}{N^{2/5}}$ and applying the Borel–Cantelli lemma. □

BIBLIOGRAPHY

[AGZ10] Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni, *An Introduction to Random Matrices*, Cambridge Studies in Advanced Mathematics, vol. 118, Cambridge University Press, 2010. ↑253

[Bia97a] Philippe Biane, *Free Brownian motion, free stochastic calculus and random matrices*, Free probability theory (Waterloo, ON, 1995), Fields Institute Communications, vol. 12, American Mathematical Society, 1997, pp. 1–19. ↑248, 249, 260

[Bia97b] ———, *Segal-Bargmann transform, functional calculus on matrix spaces and the theory of semi-circular and circular systems*, J. Funct. Anal. **144** (1997), no. 1, 232–286. ↑248

[CDK18] Benoît Collins, Antoine Dahlqvist, and Todd Kemp, *The spectral edge of unitary Brownian motion*, Probab. Theory Relat. Fields **170** (2018), no. 1-2, 49–93. ↑248, 249, 250, 257

[DHK13] Bruce K. Driver, Brian C. Hall, and Todd Kemp, *The large-N limit of the Segal-Bargmann transform on* $\mathbb{U}_N$, J. Funct. Anal. **265** (2013), no. 11, 2585–2644. ↑248

[Gri99] Alexander Grigor'yan, *Analytic and geometric background of recurrence and non-explosion of the Brownian motion on Riemannian manifolds*, Bull. Am. Math. Soc. **36** (1999), no. 2, 135–249. ↑261

[Gro87] Karsten Grove, *Metric differential geometry*, Differential geometry (Lyngby, 1985), Lecture Notes in Mathematics, vol. 1263, Springer, 1987, pp. 171–227. ↑262

[Kat04] Yitzhak Katznelson, *An introduction to harmonic analysis*, third ed., Cambridge Mathematical Library, Cambridge University Press, 2004. ↑257

[Kem15] Todd Kemp, *Heat Kernel Empirical Laws on* $\mathbb{U}_N$ *and* $\mathbb{GL}_N$, J. Theor. Probab. **30** (2015), no. 2, 1–55. ↑248

[Led99] Michel Ledoux, *Concentration of measure and logarithmic Sobolev inequalities*, Séminaire de probabilités de Strasbourg XXXIII, Lecture Notes in Mathematics, vol. 1709, Springer, 1999, pp. 120–216. ↑251

[Lév08] Thierry Lévy, *Schur–Weyl duality and the heat kernel measure on the unitary group*, Adv. Math. **218** (2008), no. 2, 537–575. ↑248

[LM10] Thierry Lévy and Mylène Maïda, *Central limit theorem for the heat kernel measure on the unitary group*, J. Funct. Anal. **259** (2010), no. 12, 3163–3204. ↑248

[MM13a] Elizabeth S. Meckes and Mark W. Meckes, *Concentration and convergence rates for spectral measures of random matrices*, Probab. Theory Relat. Fields **156** (2013), no. 1-2, 145–164. ↑255, 256, 257

[MM13b] ———, *Spectral measures of powers of random matrices*, Electron. Commun. Probab. **18** (2013), 78 (13 pages), Art. ID 78, 13 pages. ↑251, 259

[MM17] ———, *Rates of convergence for empirical spectral measures: a soft approach*, Convexity and concentration, The IMA Volumes in Mathematics and its Applications, vol. 161, Springer, 2017, pp. 1–21. ↑248, 254, 255

[Rai97] Eric M. Rains, *Combinatorial properties of Brownian motion on the compact classical groups*, J. Theor. Probab. **10** (1997), no. 3, 659–679. ↑248

[Riv81] Theodore J. Rivlin, *An Introduction to the Approximation of Functions*, Dover Books on Advanced Mathematics, Dover Publications, 1981, Corrected reprint of the 1969 original. ↑258

[RV10] Mark Rudelson and Roman Vershynin, *Non-asymptotic theory of random matrices: extreme singular values*, Proceedings of the International Congress of Mathematicians. Volume III, Hindustan Book Agency, 2010, pp. 1576–1602. ↑248

[SC94] Laurent Saloff-Coste, *Precise estimates on the rate at which certain diffusions tend to equilibrium*, Math. Z. **217** (1994), no. 4, 641–677. ↑254

[Vil09] Cédric Villani, *Optimal transport. Old and new*, Grundlehren der Mathematischen Wissenschaften, vol. 338, Springer, 2009. ↑250

[Xu97] Feng Xu, *A random matrix model from two-dimensional Yang-Mills theory*, Commun. Math. Phys. **190** (1997), no. 2, 287–307. ↑248

Manuscript received on 16th February 2018,
revised on 1st November 2018,
accepted on 13th December 2018.

Recommended by Editor N. Privault.

Elizabeth MECKES
Department of Mathematics, Applied Mathematics, and Statistics, Case Western Reserve University, 10900 Euclid Ave., Cleveland, OH 44016, USA
elizabeth.meckes@case.edu

Tai MELCHER
Department of Mathematics, University of Virginia, Charlottesville, VA 22904, USA
melcher@virginia.edu