CONVERGENCE OF THE EMPIRICAL SPECTRAL MEASURE OF UNITARY BROWNIAN MOTION

ELIZABETH MECKES† AND TAI MELCHER∗

ABSTRACT. Let \( \{U^N_t\}_{t \geq 0} \) be a standard Brownian motion on \( U(N) \). For fixed \( N \in \mathbb{N} \) and \( t > 0 \), we give explicit bounds on the \( L_1 \)-Wasserstein distance of the empirical spectral measure of \( U^N_t \) to both the ensemble-averaged spectral measure and to the large-\( N \) limiting measure identified by Biane. We are then able to use these bounds to control the rate of convergence of paths of the measures on compact time intervals. The proofs use tools developed by the first author to study convergence rates of the classical random matrix ensembles, as well as recent estimates for the convergence of the moments of the ensemble-averaged spectral distribution.

CONTENTS

1. Introduction 1
2. A concentration inequality for heat kernel measure 4
3. Concentration of \( \mu^N_t \) 7
4. Convergence to \( \nu_t \) 9
5. Convergence of paths 13
References 16

1. INTRODUCTION

This paper studies the convergence of the empirical spectral measure of Brownian motion on the unitary group \( U(N) \) 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 natural in the context of random matrices, there has been particular focus on the asymptotic behavior (as \( N \) tends to infinity) of the spectral measure of unitary Brownian motions; see for example \[16, 19, 2, 3, 11, 12, 5, 9, 4\] and the references therein.

Of course, many tools have been developed to study the spectral distributions of random matrices in high dimension in a variety of contexts. Among them is an approach developed by the first author with M. Meckes (see \[15\] 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 \cite{4} to prove asymptotically almost sure rates of convergence. We then use these bounds to control the rate of convergence of paths of the measures on compact time intervals.

**Statement of results.** Let $\mathbb{U}(N)$ denote the unitary group and $u(N)$ its Lie algebra of skew-Hermitian matrices equipped with the scaled (real) inner product $\langle U, V \rangle_N := N \operatorname{tr}(UV^*)$. This is the unique scaling that gives meaningful limiting behavior as $N \to \infty$; see for example Remark 3.4 of \cite{5}. The inner product on $u(N)$ induces a left-invariant Riemannian metric on $\mathbb{U}(N)$, and we may define Brownian motion on $\mathbb{U}(N)$ as the Markov diffusion $\{U_t^N\}_{t \geq 0}$ 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

$$dU_t^N = U_t^N dW_t^N - \frac{1}{2} U_t^N dt$$

with $U_0^N = I_N$, where $W_t$ is a standard Brownian motion on $u(N)$ (for example, take $\{\xi_k\}_{k=0}^{N^2-1}$ an orthonormal basis of $u(N)$ with respect to the given inner product and $W_t = \sum_{j=0}^{N^2-1} b_t^j \xi_j$, 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 $\rho_t^N = \operatorname{Law}(U_t^N)$ denote the end point distribution of Brownian motion; $\rho_t^N$ 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 $e^{i\theta_1}, \ldots, e^{i\theta_N}$ (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 $f \in C(\mathbb{S}^1)$

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

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 \cite{2}, Biane showed that the random probability measure $\mu_t^N$ converges weakly almost surely to a deterministic probability measure, which we denote by $\nu_t$: that is, for all $f \in C(\mathbb{S}^1)$,

$$\lim_{N \to \infty} \int_{\mathbb{S}^1} f \, d\mu_t^N = \int_{\mathbb{S}^1} f \, d\nu_t \text{ a.s.}$$
The measure $\nu_t$ represents in some sense the spectral distribution of a “free unitary Brownian motion”. For $t > 0$, $\nu_t$ possesses a continuous density that is symmetric about $1 \in S^1$. When $0 < t < 4$, $\nu_t$ is supported on an arc strictly contained in the circle; for $t \geq 4$, $\text{supp}(\nu_t) = S^1$. The paper [4] presents a nice brief summary of these and other properties of $\nu_t$ 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 $\mu^N_t$ and its limiting spectral measure $\nu_t$, where for probability measures $\mu$ and $\nu$ on $C$, the $L_1$-Wasserstein distance is defined by

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

We will also make use of the equivalent dual representation of $W_1$ due to Kantorovich and Rubenstein:

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

where $|f|_L$ denotes the Lipschitz constant of $f$.

The main results of this paper are the following.

**Theorem 1.** Let $\{U^N_t\}_{t \geq 0}$ be a Brownian motion on $U(N)$. For $t > 0$, let $\mu^N_t$ denote the empirical spectral measure $U_t$ as above, and let $\overline{\mu^N_t}$ denote the ensemble-averaged spectral measure of $U^N_t$ defined by

$$\int_{S^1} f \, d\overline{\mu^N_t} := E \int_{S^1} f \, d\mu^N_t.$$ 

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

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

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

$$W_1(\mu^N_t, \overline{\mu^N_t}) \leq C \frac{1}{N^{2/3}}.$$ 

**Theorem 2.** Let $\nu_t$ 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 $t > 0$

$$W_1(\overline{\nu^N_t}, \nu_t) \leq C \min \left\{ \frac{t^{2/3} \log N}{N^{2/5}}, e^{-\frac{t(1+o(1))}{8\log(N)}}, \frac{1}{N} \right\}.$$ 

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 $\mu^N_t$ itself; previously the only known convergence rates were for moments of the ensemble-averaged spectral measure $\overline{\mu^N_t}$ [4].

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 3. Let $T \geq 0$. There are constants $c, C$ such that for all $x \geq c T^{2/5} \log(N)$, $\rho(1)$
$$\mathbb{P}\left(\sup_{0 \leq t \leq T} W_1(\mu^N_t, \nu_t) > x\right) \leq C \left(\frac{T}{x^2} + 1\right) e^{-\frac{c^2 x^2}{2}}.$$ 
In particular, with probability one for $N$ sufficiently large
$$\sup_{0 \leq t \leq T} W_1(\mu^N_t, \nu_t) \leq c T^{2/5} \log(N) / N^{2/5}.$$ 

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

Proposition 4. Let $\nu_t$ denote the limiting spectral measure and $\nu$ the uniform measure on $S^1$. Then there is a constant $C \in (0, \infty)$ so that for all $t \geq 1$
$$W_1(\nu_t, \nu) \leq C t^{3/2} e^{-t/4}.$$ 

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 $S^1$ and $\mathbb{U}(N)$. These estimates are then used in Section 3 to prove Theorem 3. 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 $\nu_t$ as in Theorem 3. In this section, we also give the proof of Proposition 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 $\{\nu_t\}_{t \geq 0}$, which are then both used to give the proof of Theorem 3.

2. A CONCENTRATION INEQUALITY FOR HEAT KERNEL MEASURE

In this section, we will consider concentration of measure results for Lipschitz functions of the following form. Let $(X, d)$ be a metric space equipped with Borel probability measure $\rho$. Then, under some conditions, there exists $C > 0$ such that, for all $r > 0$ and $F : X \to \mathbb{R}$ Lipschitz with Lipschitz constant $L$ and $\mathbb{E}[|F|] < \infty$,
$$\rho(|F - \mathbb{E}[F]| \geq r) \leq 2 e^{-r^2/L^2 C}.$$ 

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 $\Delta$ 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 $f : M \to \mathbb{R}$,
$$P_t f(x) = \mathbb{E}[f(\xi^x_t)] = \int_M f \, d\rho^x_t$$
where $\{\xi^x_t\}_{t \geq 0}$ is the Markov diffusion on $M$ started at $x$ with generator $\Delta$ (that is, $\xi^x_t$ is a Brownian motion on $M$) and $\rho^x_t = \text{Law}(\xi^x_t)$ is the heat kernel measure. If $\text{Ric}$ denotes the Ricci curvature tensor on $M$, then $\text{Ric} \geq 2k$ for $k \in \mathbb{R}$ implies that for all $t > 0$ the estimate (1) holds for $\rho_t$ with coefficient $C(t) = 2(1 - e^{-kt/2})/k$, where when $k = 0$, we interpret this to be $C(t) = t$. (A typical proof is via log Sobolev estimates.) See for example Corollary 2.6 and Lemma 6.3 of [10] (stated in the case that $k \geq 0$, 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 $\rho^N_t$ 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 \[14\]. The following lemma gives the key idea.

**Lemma 5.** Let \( b^0 \) be a real-valued Brownian motion and \( z_t := e^{ib^0_t/N} \), and let \( V_t \) be a Brownian motion on \( \mathbb{SU}(N) \) issued from the identity. 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

\[
dz_t = \frac{i b^0_t}{N} - \frac{1}{2N^2} z_t \, dt \quad \text{and} \quad dZ_t = Z_t \, dB_t - \frac{1}{2N^2} Z_t \, dt
\]

where \( b_t = b^0_t \xi_0 \) with \( \xi_0 = iI_N/N \). Let \( \{\xi_j\}_{j=1}^{N^2-1} \) be an orthonormal basis of \( \mathfrak{su}(N) \), and let \( \{b_j\}_{j=1}^{N^2-1} \) be independent real-valued Brownian motions. Then

\[
\tilde{W}_t = \sum_{j=1}^{N^2-1} b_j \xi_j \quad \text{is a Brownian motion on} \quad \mathfrak{su}(N), \quad \text{and} \quad V_t \quad \text{satisfies the stochastic differential equation}
\]

\[
dV_t = V_t \circ d\tilde{W}_t = V_t d\tilde{W}_t + \frac{1}{2} V_t \sum_{\xi \in \beta} \xi^2 \, dt = V_t d\tilde{W}_t - \left( \frac{N^2 - 1}{2N^2} \right) V_t \, dt.
\]

(Here \( \circ \) denotes a Stratonovich integral, which is then expressed as an Itô integral via the usual calculus.)

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

\[
d(Z_t V_t) = \left( Z_t dB_t - \frac{1}{2} Z_t \, dt \right) V_t + Z_t \left( V_t d\tilde{W}_t - \left( \frac{N^2 - 1}{2N^2} \right) V_t \right)
\]

\[
= Z_t V_t (dB_t + d\tilde{W}_t) - \frac{1}{2} Z_t V_t \, dt.
\]

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

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 6.** 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 \( t, r > 0 \),

\[
\mathbb{P}( |F(U_t) - \mathbb{E}F(U_t)| > r ) \leq 2 e^{-\frac{cr^2}{\pi t}}.
\]

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

\[
\mathbb{P}( |F(U_t) - \mathbb{E}F(U_t)| > r ) \leq Ce^{-\frac{cr^2}{\pi t}}.
\]

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

\[
\mathbb{P}( |F(U_t) - \mathbb{E}F(U_t)| > r ) \leq 2 e^{-\frac{cr^2}{\pi t}}.
\]
To prove the second statement, observe that the representation of $U_t$ in Lemma 5 implies that
\[ \mathbb{P}(|F(U_t) - \mathbb{E}F(U_t)| > r) = \mathbb{P}(|F(z_t V_t) - \mathbb{E}F(z_t V_t)| > r) \]
\[ \leq \mathbb{E} \left[ \mathbb{P} \left( \left| F(z_t V_t) - \mathbb{E}[F(z_t V_t)|z_t] \right| > \frac{r}{2} \right) \right] + \mathbb{P} \left( \left| \mathbb{E}[F(z_t V_t)|z_t] - \mathbb{E}F(z_t V_t) \right| > \frac{r}{2} \right). \]
(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 \[1\], one may compute the Ricci curvature on $\text{SU}(N)$ with respect to the given inner product as
\[ \text{Ric}(X, X) = \frac{1}{2} \langle X, X \rangle_N. \]

Thus, by the discussion preceding Lemma 5, Law($V_t$) on $\text{SU}(N)$ satisfies the following concentration estimate: if $G : \text{SU}(N) \to \mathbb{R}$ is $L$-Lipschitz, then
\[ \mathbb{P}(|G(V_t) - \mathbb{E}G(V_t)| > r) \leq 2e^{-\frac{c(t)^2}{4}}, \]
where $c(t) := 4(1 - e^{-t/4})$. For $z_t$ fixed, $G = F(z_t)$ is an $L$-Lipschitz function on $\text{SU}(N)$, and so the first term of (2) is bounded by $2e^{-\frac{c(t)^2}{4}}$.

For the second term of (2), let $K = K(z_t)$ be the random variable taking values in $\{0, \ldots, N - 1\}$ such that, on $\{K = k\}$, $z_t \in \left[e^{2\pi i/k}, e^{2\pi i(k+1)/N}\right)$. Conditioning on $K$, we have
\[ \mathbb{P} \left( \left| \mathbb{E}[F(z_t V_t)|z_t] - \mathbb{E}F(z_t V_t) \right| > \frac{r}{2} \right) \]
\[ = \mathbb{E} \left[ \mathbb{P} \left( \left| \mathbb{E}[F(z_t V_t)|z_t] - \mathbb{E}F(z_t V_t) \right| > \frac{r}{2} \bigg| K \right) \right] \]
\[ \leq \mathbb{E} \left[ \mathbb{P} \left( \left| \mathbb{E}[F(z_t V_t)|z_t] - \mathbb{E}[F(z_t V_t)|K] \right| > \frac{r}{4} \bigg| K \right) \right] \]
\[ + \mathbb{P} \left( \left| \mathbb{E}[F(z_t V_t)|K] - \mathbb{E}F(z_t V_t) \right| > \frac{r}{4} \right). \]
(3)

To deal with the first term in (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 $z_t$
\[ \left| \mathbb{E}[F(z_t V_t)|z_t] - \mathbb{E}[F(z_t V_t)|K = k] \right| = \left| \mathbb{E}_{V_t} [F(z_t V_t)] - \mathbb{E}_{V_t} \mathbb{E}_{z_t|K=k} [F(z_t V_t)] \right| \]
\[ \leq \mathbb{E}_{V_t} \left| F(z_t V_t) - \mathbb{E}_{z_t|K=k} [F(z_t V_t)] \right| \]
\[ = \int_{\text{SU}(N)} \left| F(z_t V) - \mathbb{E}_{z_t|K=k} [F(z_t V)] \right| d\mu_{\text{SU}(N)}(V), \]
where $\mu_{\text{SU}(N)}$ denotes the density of $V_t$ with respect to Haar measure on $\text{SU}(N)$.

Now, for $V$ fixed, $F(V)$ is an $NL$-Lipschitz function on $S^1$. So, conditioned on $K = k$, $F(z_t V)$ can only fluctuate by as much as $2\pi L$. Thus if $\frac{r}{4} > 2\pi L$, the first term is zero. For $\frac{r}{4} \leq 2\pi L$, we may just use the trivial bound of 1 and choose $C$ in the statement of the proposition so that $C \geq e^{(8\pi^2)^2/4}$.

To deal with the second term in (3), note that we can replace $V_t$ with a Haar-distributed random matrix $V$ for $t$ sufficiently large. Indeed, letting $dV$ denote
integration with respect to Haar measure on $\SU(N)$, and assuming without loss in generality that $F(I_N) = 0$,
\begin{equation}
\|E[F(z_t V) - F(z_t V)]\|_{L^1} \leq \int_{\SU(N)} |F(z_t V)| \left\| h_t^{\SU(N)}(V) - 1 \right\| dV
\end{equation}
\begin{equation}
\leq LN\|h_t^{\SU(N)} - 1\|_1,
\end{equation}
since the diameter of $U(N)$ is $N$. A sharp estimate of the time to equilibrium of $V_t$ was proved in Theorem 1.2 of [1S], from which it follows (see the discussion preceding the theorem in [1S], and note that the normalization here differs by a factor of 2 from the one used there) that
\begin{equation}
\|h_t^{\SU(N)} - 1\|_1 \leq e^{-\frac{t(1+\omega_1^2)}{8\log N}}.
\end{equation}
Thus if $t \geq 8(\log N)^2$, replacing $V_t$ by $V$ will only affect the constants.

Consider therefore
\begin{equation}
P \left[ \left| E[F(z_t V)|K] - E[F(z_t V)] \right| > \frac{t}{4} \right],
\end{equation}
and write $z_t = \omega_t e^{\frac{2\pi i K}{N}}$, with $\omega_t$ 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 $\SU(N)$,
\begin{equation}
E[F(z_t V)] = E_{z_t}E_V[F(\omega_t e^{\frac{2\pi i K}{N}} V)] = E_{z_t}E_V[F(\omega_t V)] = E_VE_{z_t}[F(\omega_t V)],
\end{equation}
and similarly
\begin{equation}
E[F(z_t V)|K = k] = E_{z_t|K = k}E_V[F(\omega_t e^{\frac{2\pi i K}{N}} V)] = E_VE_{z_t|K = k}[F(\omega_t V)].
\end{equation}
Thus
\begin{equation}
\left| E \left[ F(z_t V)|K \right] - E[F(z_t V)] \right| \leq E_V \left| E_{z_t}[F(\omega_t V)] - E_{z_t|K}[F(\omega_t V)] \right| \leq 2\pi L,
\end{equation}
where we have used again that for fixed $V$, $F(\omega V)$ is an $NL$-Lipschitz function of $\omega$, and here $\omega$ lies within an arc of length $\frac{2\pi}{N}$. The estimate now follows as in the first term.

3. Concentration of $\mu_t^N$

Armed with the concentration inequality for heat kernel measure, the proof of Theorem [11] is an application of the program laid out in [15] 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” $E W_1(\mu_t^N, \overline{\mu}_t^N)$ as follows.

**Proposition 7.** There is a constant $c \in (0, \infty)$ such that for all $N \in \mathbb{N}$ and $t > 0$
\begin{equation}
E W_1(\mu_t^N, \overline{\mu}_t^N) \leq c \left( \frac{t}{N^2} \right)^{1/3},
\end{equation}
and for all $N \in \mathbb{N}$ and $t \geq 8(\log N)^2$
\begin{equation}
E W_1(\mu_t^N, \overline{\mu}_t^N) \leq \frac{c}{N^{2/3}}.
\end{equation}
Proof. We will give the proof of the first statement only, which applies the first half of Proposition 6, the proof of the second statement is identical using only instead the second half of Proposition 6.

Recall that

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

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

\[ X_f := \int f \, d\mu_t^N - \int f \, d\overline{\mu}_t^N = \int f \, d\mu_t^N - \mathbb{E} \int f \, d\mu_t^N. \]

Note that without loss we may choose the indexing set to be 1-Lipschitz functions on the circle with \( f(1) = 0 \); write \( \text{Lip}_0(1) \) for the set of all such functions. Now, if \( f \) is a fixed Lipschitz function and \( \mu_U \) denotes the spectral measure of \( U \), then

\[ \mathbb{P}(\sup_{|v|_\infty = 1} |X_v - X_{-v}| > x) = \mathbb{P}(\sup_{|v|_\infty = 1} |X_v - g| > x) \leq 2e^{-\frac{x^2}{\sqrt{\dim V}}}. \]

That is, the stochastic process \( \{X_f\}_{f \in \text{Lip}_0(1)} \) satisfies a sub-Gaussian increment condition.

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

\[ \mathbb{P}(\sup_{\|v\|_V = 1} |X_u - X_v| > x) \leq 2e^{-\frac{x^2}{\sqrt{\dim V}}}, \]

for each \( x > 0 \), then it is a consequence of Dudley’s entropy bound (see [15] for a detailed proof) that

\[ \mathbb{E} \left( \sup_{\|v\|_V = 1} X_v \right) \leq aK\sqrt{\dim V}. \]

The index set \( \text{Lip}_0(1) \) 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

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

For any \( f \in \text{Lip}_0(1) \), there is \( f^{(m)} \in A_0^{(m)} \) such that \( \|f - f^{(m)}\|_\infty \leq \frac{2\pi}{m} \), and so

\[ |X_f - X_{f^{(m)}}| = \left| \int (f - f^{(m)}) \, d\mu_t^N - \int (f - f^{(m)}) \, d\overline{\mu}_t^N \right| \leq \frac{2\pi}{m}. \]
The space of functions for which $A_{0}^{(m)}$ is the unit ball is $(m - 1)$-dimensional, and so it follows from (6) that
\[
E \left( \sup_{f \in L^{0}(1)} X_{f} \right) \leq \frac{2\pi}{m} + E \left( \sup_{f \in A_{0}^{(m)}} X_{f} \right) \leq \frac{2\pi}{m} + C + \left( \frac{\sqrt{t}}{N} \right) \sqrt{m - 1}.
\]
Choosing $m = \left( \frac{N^{2}}{t} \right)^{1/3}$ completes the proof.

The proof of Theorem 1 is completed via the concentration of $W_{1}(\mu_{N}^{t} \cdot \overline{\nu}_{t}^{N})$ about its mean, as follows.

**Proposition 8.** For all $t > 0$, $N \in \mathbb{N}$, and $x > 0$,
\[
\mathbb{P} \left( W_{1}(\mu_{t}^{N} \cdot \overline{\nu}_{t}^{N}) > EW_{1}(\mu_{t}^{N} \cdot \overline{\nu}_{t}^{N}) + x \right) \leq 2e^{-N^{2}x^{2}/4},
\]
and there exists $C \in (0, \infty)$ such that for all $t \geq 8(\log N)^{2}$, $N \in \mathbb{N}$, and $x > 0$,
\[
\mathbb{P} \left( W_{1}(\mu_{t}^{N} \cdot \overline{\nu}_{t}^{N}) > EW_{1}(\mu_{t}^{N} \cdot \overline{\nu}_{t}^{N}) + x \right) \leq Ce^{-N^{2}x^{2}/4}.
\]

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

Consider the mapping $F : U(N) \to \mathbb{R}$ given by
\[
F(U) = W_{1}(\mu_{U} \cdot \overline{\nu}_{t}^{N}),
\]
where $\mu_{U}$ is the spectral measure of $U$ and $\overline{\nu}_{t}^{N}$ 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 [13]), and so by Proposition 6 for all $t > 0$ and all $x > 0$,
\[
\mathbb{P} \left( W_{1}(\mu_{t}^{N} \cdot \overline{\nu}_{t}^{N}) > EW_{1}(\mu_{t}^{N} \cdot \overline{\nu}_{t}^{N}) + x \right) \leq 2e^{-N^{2}x^{2}/4}.
\]

From the tail estimate of Proposition 8 together with Proposition 7 it follows that for any $t, x > 0$,
\[
\mathbb{P} \left( W_{1}(\mu_{t}^{N} \cdot \overline{\nu}_{t}^{N}) > c \left( \frac{t}{N^{2}} \right)^{1/3} + x \right) \leq 2e^{-N^{2}x^{2}/4}.
\]
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. The second statement follows in the same way.

### 4. Convergence to $\nu_{t}$

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

**Theorem 9.** There is a constant $C \in (0, \infty)$ such that for all $N \in \mathbb{N}$ and $t > 0$

$$W_1(\mathbb{P}^N_t, \nu_t) \leq C \frac{t^{2/5} \log N}{N^{2/5}}.$$  

*Proof.* The proof is via Fourier analysis and classical approximation theory, following the approach of Theorem 2.1 in [13]. The key ingredient of this proof is the bound (7) below, which was proved in [4]. Let

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

and observe that

$$\int z^k d\mathbb{P}^N_t = \frac{1}{N} \mathbb{E}[\text{tr}(U^k_t)]$$

where $U_t$ is a Brownian motion on $U(N)$. Given $f: \mathbb{S}^1 \to \mathbb{R}$ a 1-Lipschitz function, it is known that $|\hat{f}(k)| \leq \frac{C}{k}$ for $k \geq 1$ (in fact, $C = \frac{\pi}{2}$; see, for example, Theorem 4.6 of [3]), and so

$$\left| \int S_m d\mathbb{P}^N_t - \int S_m dv_t \right| = \left| \sum_{1 \leq |k| \leq m} \hat{f}(k) \left( \frac{1}{N} \mathbb{E}[\text{tr}(U^k_t)] - \int z^k dv_t \right) \right|$$

Now, by Theorem 1.3 of [4], for $t$ and $k$ fixed,

$$\left| \frac{1}{N} \mathbb{E}[\text{tr}(U^k_t)] - \int z^k dv_t \right| \leq \frac{t^2 k^4}{N^2}.$$  

Thus,

$$\left| \int S_m d\mathbb{P}^N_t - \int S_m dv_t \right| \leq C \sum_{1 \leq |k| \leq m} \frac{t^2 k^4}{k N^2} \leq C t^2 m^4 N^{-2}.$$  

The proof now proceeds exactly as in Theorem 2.1 of [13]. A theorem of Lebesgue implies that

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

where the infimum is over all trigonometric polynomials $g(z) = \sum_{|k| \leq m} a_k z^k$; see for example Theorem 2.2 of [17]. Combining this with Jackson’s theorem (Theorem 1.4 of the same reference) implies that $\|f - S_m\|_\infty \leq C' \frac{\log m}{m}$, and thus

$$\left| \int f d\mathbb{P}^N_t - \int f dv_t \right| \leq \left| \int f d\mathbb{P}^N_t - \int S_m d\mathbb{P}^N_t \right| + \left| \int S_m d\mathbb{P}^N_t - \int S_m dv_t \right| + \left| \int S_m dv_t - \int f dv_t \right| \leq C' \left( \frac{\log m}{m} + \frac{t^2 m^4}{N^2} \right).$$

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

The bound above decays if and only if $t = o(N/((\log N)^{5/2}))$. But for sufficiently large $t$, both $\mathbb{P}^N_t$ and $\nu_t$ 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 10. Let $\overline{\mu}_t^N$ denote the ensemble-averaged spectral measure of a random matrix $U_t$, distributed according to heat kernel measure on $U(N)$, and let $\nu$ denote the uniform probability measure on $S^1$. There are constants $C, c \in (0, \infty)$ so that for all $N \in \mathbb{N}$ and $t > 0$

$$W_1(\overline{\mu}_t^N, \nu) \leq e^{-\frac{(1+c)(1+\log(N))}{N}} + \frac{2\pi}{N}.$$ 

Proof. First recall again that, as in the proof of Proposition 7, if $\mu_U$ denotes the spectral measure of $U$, then for fixed $f : S^1 \to \mathbb{R}$ with $|f|_L \leq 1$, the function

$$F(U) = \int fd\mu_U$$

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

$$\int fd\mu_t^N - \int fd\nu \leq \|U_t - U\|_N,$$

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

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

Taking the supremum over $f$ gives that

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

and now taking infimum over couplings we have

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

Now consider the coupling $U_t \overset{d}{=} z_tV_t$ from Lemma 2, where $z_t = e^{ib_t/N}$ for $b_t$ a standard Brownian motion on $\mathbb{R}$ and $V_t$ an independent Brownian motion on $SU(N)$ 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 $SU(N)$: if $z = e^{i\theta/N}$ with $\theta$ uniform in $[0, 2\pi)$ and $V$ is independent of $\theta$ and distributed according to Haar measure on $SU(N)$, then $zV$ is distributed according to Haar measure on $U(N)$; see for example Lemma 16 of [14]. Moreover, by the translation invariance of Haar measure, $\theta$ 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 $F : U(N) \to \mathbb{R}$ a 1-Lipschitz function, we have that

$$|EF(U_t) - EF(U)| = |EF(z_tV_t) - EF(zV)| \leq \mathbb{E}\left|F(z_tV_t) - F(zV)\right| = \mathbb{E}\left|F(z_tV_t) - F(z_tV)\right| + \mathbb{E}\left|F(z_tV) - F(zV)\right|,$$

The first term of (8) was already bounded in the course of the proof of Proposition 6

$$\mathbb{E}\left|F(z_tV_t) - F(z_tV)\right| \leq N e^{-\frac{(1+c)(1+\log(N))}{2N}}.$$
To treat the second term, we may as in the proof of Proposition 6 write \( z_t = \omega_t e^{2\pi i K/N} \), with \( \omega_t \) in the arc from 1 to \( e^{2\pi i/N} \) and \( K \in \{0, \ldots, N-1\} \), and similarly \( z = \omega e^{2\pi i K/N} \), the second term of (9) can be bounded as

\[
E[J(F(z_t V) - F(z V))] = E[J(F(\omega_t e^{2\pi i K t/N} V)] - E[J(F(\omega e^{2\pi i K/N} V)] = E[J(F(\omega_t V) - F(\omega V))],
\]

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 \( \omega \), with \( \omega, \omega_t \) lying in an arc of length \( 2\pi/N \).

Combining this last estimate with (8), (9), (4), and (5) implies that

\[
W_1(\mu_N t, \nu) \leq C t^{3/2} e^{-t/4}.
\]

Finally, we compare the limiting (large \( N \)) measure \( \nu_t \) to the uniform measure \( \nu \). We restate and prove here Proposition 4.

**Proposition 4.** For \( \nu_t \) and \( \nu \) defined as above, there is a constant \( C \in (0, \infty) \) so that for all \( t \geq 1 \)

\[
W_1(\nu_t, \nu) \leq C t^{3/2} e^{-t/4}.
\]

Observe in particular that as \( t \to \infty \), \( t^{3/2} e^{-t/4} \leq e^{-8 \log(N)} \), and so Theorem 2 follows from Propositions 10 and 4 together with the triangle inequality.

**Proof of Proposition 4.** The measure \( \nu_t \) is symmetric, and the moments of \( \nu_t \) for \( k \geq 1 \) are given by

\[
\int_{S^1} z^k d\nu_t(z) = Q_k(t) e^{-t/4},
\]

where

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

see [2]. As in the proof of Theorem 4 for a fixed 1-Lipschitz test function \( f : S^1 \to \mathbb{R} \), let

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

and we have that \( |\hat{f}(k)| \leq C \frac{1}{k} \) for all \( k \geq 1 \). Then since both \( \nu_t \) and \( \nu \) are probability measures on \( S^1 \) and \( \int_{S^1} z^j d\nu(z) = 0 \) if \( j \neq 0 \),

\[
\left| \int S_m(z) d\nu_t(z) - \int S_m(z) d\nu(z) \right| = \left| \sum_{1 \leq |k| \leq m} \hat{f}(k) \int z^k d\nu_t(z) \right| \leq C \sum_{1 \leq |k| \leq m} \frac{1}{k} |Q_k(t)| e^{-t/4}.
\]
Let
\[ A_k(t) := Q_k(-t) = \sum_{j=0}^{k-1} \frac{(tk)^j}{(j+1)!} j! \binom{k-1}{j}, \]
so that \(|Q_k(t)| \leq A_k(t)\). Now,
\[ A_{k+1}(t) = 1 + \sum_{j=1}^{k} \frac{[t(k+1)]^j}{(j+1)!} j! \binom{k}{j} \]
\[ = 1 + tk(k+1) \sum_{j=1}^{k} \left( \frac{(1+\frac{1}{tk})^{j-1}}{j!(j+1)!} \right) \binom{k}{j} \]
\[ = 1 + tk(k+1) \sum_{j=1}^{k} \left( \frac{(1+\frac{1}{tk})^{j-1}}{j!(j+1)!} \right) \binom{k}{j} \]
and note that
\[ A_k(t) = \sum_{j=0}^{k-1} \frac{(tk)^j}{(j+1)!} j! \binom{k-1}{j} \]
\[ = \sum_{j=0}^{k-1} \frac{(tk)^j}{k^j!(j+1)!} j! \binom{k}{j} \]
\[ = \sum_{j=0}^{k} \frac{(tk)^j}{k^j!(j+1)!} j! \binom{k}{j} \]
\[ = \sum_{\ell=1}^{k} \frac{(tk)^{\ell-1}}{k^{\ell-1}!(\ell)!} \binom{k}{\ell}. \]

Since \( (1+\frac{1}{tk})^{j-1} \) is decreasing as a function of \( \ell \) on \( \{1, \ldots, k\} \), it follows that
\[ A_{k+1}(t) \leq 1 + \left( \frac{tk(k+1)}{2} \right) A_k(t) \leq tk(k+1)A_k(t), \]
since \( t, k \geq 1 \). By induction and the fact that \( A_1(t) = 1 \), this implies that
\[ |Q_k(t)| \leq A_k(t) \leq tk^{-1}k![(k-1)!]^2. \]

It now follows from (10) that
\[ \left| \int S_m(z) d\nu_t(z) - \int S_m(z) d\nu(z) \right| \leq \sum_{k=1}^{m} t^{k-1}[(k-1)!]^2 e^{-\frac{tk}{4}} \]
\[ \leq e^{-t/2} \sum_{k=1}^{m} \left( t(k-1)^2 e^{-t/2} \right)^{k-1} \leq e^{-t/2} \sum_{k=1}^{m} \left( tm^2 e^{-t/2} \right)^{k-1}. \]

Choose \( m = \left\lfloor \sqrt{\frac{e}{2t}} \right\rfloor \), so that \( tm^2 e^{-t/2} \leq \frac{1}{2} \). Then
\[ \left| \int S_m(z) d\nu_t(z) - \int S_m(z) d\nu(z) \right| \leq 2e^{-t/2}. \]
As in the proof of Theorem 9 we have that \( \|S_m - f\|_\infty \leq C \frac{\log m}{m} \), which for the chosen value of \( m \) yields
\[ \|S_m - f\|_\infty \leq C'' t^{3/2} e^{-t/4}. \]
Combining these estimates completes the proof. \( \Box \)

5. Convergence of paths

This section is devoted to the proof of Theorem 8. The idea is to first discretize the interval \([0, T]\) and apply the bound from Proposition 8 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 11. Let \( \{U_t\}_{t>0} \) denote Brownian motion in \( U(N) \) with \( U_0 = I_N \), and let \( d_g \) denote the geodesic distance on \( U(N) \) induced by \( \langle \cdot, \cdot \rangle_N \). Then for all \( \delta, r, s > 0 \),
\[
\mathbb{P} \left( \sup_{0 < t < \delta} d_g(U_t, I_N) \geq r + 2s \right) \leq 16 \left( 1 + \frac{r}{s} \right)^N e^{-\frac{s^2}{2}}.
\]

Proof. If \( d_g(U_t, I_N) < s \), then by left invariance of the metric and the triangle inequality
\[
d_g(U_t, I_N) = d_g(UU_t, U) \leq d_g(UU_t, I_N) + s.
\]
Thus,
\[
\mathbb{P} \left( \sup_{0 < t < \delta} d_g(U_t, I_N) \geq 2s + r \right) \leq \inf_{d_g(U, I) \leq s} \mathbb{P} \left( \sup_{0 < t < \delta} d_g(UU_t, I_N) \geq s + r \right).
\]
Applying the bound in Equation (9.20) of [5] with \( M = U(N) \) and \( K = B(I_N, s) \) (the closed geodesic ball of radius \( s \) about \( I_N \)) gives that
\[
\inf_{d_g(U, I) \leq s} \mathbb{P} \left( \sup_{0 < t < \delta} d_g(UU_t, I_N) \geq s + r \right) \leq 16 \frac{\text{vol}(B(I_N, s + r))}{\text{vol}(B(I_N, s))} e^{-\frac{s^2}{2}}.
\]
Then, recalling again that \( \text{Ric} \geq 0 \) on \( U(N) \), the Bishop–Gromov comparison theorem allows us to control the volume of balls in \( U(N) \) by the volume of balls in \( \mathbb{R}^{N^2} \) (see for example Theorem 3.16 of [7]); in particular,
\[
\frac{\text{vol}(B(I_N, s + r))}{\text{vol}(B(I_N, s))} \leq \left( 1 + \frac{r}{s} \right)^N,
\]
which completes the proof. \qed

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

Lemma 12. There is a constant \( c \) such that for all \( 0 < s < t \)
\[
W_1(\nu_t, \nu_s) \leq c \sqrt{t - s}.
\]

Proof. The triangle inequality for \( W_1 \) and Theorem 3 imply that for any \( N \)
\[
W_1(\nu_t, \nu_s) \leq W_1(\nu_t, \mu^N_t) + W_1(\nu_s, \mu^N_s) + W_1(\mu^N_t, \mu^N_s) \leq C \frac{(2^{2/5} + 2^{2/5}) \log N}{N^{2/5}} + W_1(\mu^N_t, \mu^N_s).
\]
Moreover, recall that
\[
W_1(\mu^N_t, \mu^N_s) = \sup_{|f| \leq 1} \mathbb{E} \left[ \left| \int f \, d\mu^N_t - \int f \, d\mu^N_s \right| \right] \leq \frac{\|U_t - U_s\|_N}{N},
\]
since \( U_t \) \( N \)-Lipschitz. Trivially, for any \( U, V \in U(N) \), \( \|U - V\|_N \leq d_g(U, V) \). So, using the stationarity of increments together with Proposition 11 with \( r = 2s = \frac{cN}{2} \sqrt{t - s} \),
\[
\mathbb{E}\|U_t - U_s\|_N = \mathbb{E}\|I_N - U_{t-s}\|_N \leq \mathbb{E}d_g(I_N, U_{t-s}) \leq cN \sqrt{t - s} + N \mathbb{P} \left( d_g(I, U_{t-s}) > cN \sqrt{t - s} \right) \leq cN \sqrt{t - s} + 3N^2 Ne^{-cN^2/8}.
\]
Choosing \( c \) large enough that \( \log 3 + \frac{\log N}{N^{2/5}} - \frac{x^2}{7} < 0 \) for all \( N \), this gives that
\[
\mathbb{E}\|U_t - U_s\|_N \leq cN\sqrt{t-s} + 1
\]
and thus
\[
W_1(\nu_t, \nu_s) \leq C\left(\frac{t^{2/5} + s^{2/5}}{N^{2/5}}\log N + c\sqrt{t-s} + \frac{1}{N}\right).
\]
Since this holds for any \( N \), the result follows. \( \square \)

**Proof of Theorem 3.** Let \( m \in \mathbb{N} \) such that \( \frac{T}{m} \leq 1 \), and for \( j = 1, \ldots, m \), let \( t_j := \frac{jT}{m} \).

By Lemma 12
\[
\sup_{0 \leq s,t \leq T \atop |s-t| < \frac{T}{m}} W_1(\nu_t, \nu_s) \leq c\sqrt{\frac{T}{m}},
\]
so that if \( x > 3c\sqrt{\frac{T}{m}} \), then
\[
(11) \quad \mathbb{P}\left( \sup_{0 \leq t \leq T} W_1(\mu_t^N, \nu_t) > x \right) \leq \mathbb{P}\left( \max_{1 \leq j \leq m} \sup_{|t-j| < \frac{T}{m}} W_1(\mu_t^N, \mu_j^N) > \frac{x}{3} \right) + \mathbb{P}\left( \max_{1 \leq j \leq m} W_1(\mu_t^N, \nu_j) > \frac{x}{3} \right).
\]

Using again that \( W_1(\mu_t^N, \mu_j^N) \leq \frac{\|U_t - U_j\|}{N} \), we have that for any \( A \subseteq [0,T]^2 \)
\[
\mathbb{P}\left( \sup_{(s,t) \in A} W_1(\mu_t^N, \mu_s^N) > \frac{x}{3} \right) \leq \mathbb{P}\left( \sup_{(s,t) \in A} \|U_t - U_s\| > \frac{Nx}{3} \right) = \mathbb{P}\left( \sup_{(s,t) \in A} \|I_N - U_t^{-1}U_s\| > \frac{Nx}{3} \right) = \mathbb{P}\left( \sup_{(s,t) \in A} \|I_N - U_{t-s}\| > \frac{Nx}{3} \right),
\]
where the first equality is because \( U_t \in \text{U}(N) \) and the second is by the stationarity of the increments of Brownian motion. It follows from this and (11) that
\[
\mathbb{P}\left( \sup_{0 \leq t \leq T} W_1(\mu_t^N, \nu_t) > x \right) \leq m\mathbb{P}\left( \sup_{|t| < \frac{T}{m}} \|I_N - U_t\| > \frac{Nx}{3} \right) + m\max_{1 \leq j \leq m} \mathbb{P}\left( W_1(\mu_j^N, \nu_t) > \frac{x}{3} \right).
\]

Applying Proposition 11 to the first term with \( 2s = r = \frac{Nx}{m} \) gives that
\[
\mathbb{P}\left( \sup_{|t| < \frac{T}{m}} \|I_N - U_t\| > \frac{Nx}{3} \right) \leq \mathbb{P}\left( \sup_{|t| < \frac{T}{m}} d_{\mathcal{F}}(U_t, I_N) > \frac{Nx}{3} \right) \leq 3N^2 e^{-\frac{N^2x^2}{2r^2}}.
\]

For the second term, applying the estimate following Proposition 8 together with Theorem 6 if \( x \geq 3C\frac{T^{2/5}\log(N)}{N^{2/5}} \) gives that
\[
\mathbb{P}\left( \sup_{1 \leq j \leq m} W_1(\mu_j^N, \nu_t) > \frac{x}{3} \right) \leq 2e^{-\frac{N^2x^2}{r^2}}.
\]
We thus have that, for any $m \in \mathbb{N}$ such that $\frac{T}{m} \leq 1$ and $x \geq 3C \frac{T^{2/5} \log(N)}{N^{2/5}}$,
\[
P \left( \sup_{0 \leq t \leq T} W_1(\mu_t^N, \nu_t) > x \right) \leq m^3 N^2 e^{-\frac{N^{2/5} x^2}{2}} + 2me^{-\frac{N^{2/5} x^2}{2}}.
\]
Choosing $m = \lceil 72 \left( \frac{T \log 3}{x^2} + 1 \right) \rceil$ 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. 

\[\Box\]

**References**

[1] G. W. Anderson, A. Guionnet, and O. Zeitouni. *An Introduction to Random Matrices*, volume 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010.

[2] P. Biane. Free Brownian motion, free stochastic calculus and random matrices. In *Free probability theory (Waterloo, ON, 1995)*, volume 12 of *Fields Inst. Commun.*, pages 1–19. Amer. Math. Soc., Providence, RI, 1997.

[3] P. Biane. Segal-Bargmann transform, functional calculus on matrix spaces and the theory of semi-circular and circular systems. *J. Funct. Anal.*, 144(1):232–286, 1997.

[4] B. Collins, A. Dahlqvist, and T. Kemp. The spectral edge of unitary Brownian motion. *Probability Theory and Related Fields*, pages 1–45, 2017.

[5] B. K. Driver, B. C. Hall, and T. Kemp. The large-$N$ limit of the Segal-Bargmann transform on $U_N$. *J. Funct. Anal.*, 265(11):2585–2644, 2013.

[6] A. Grigor’yan. Analytic and geometric background of recurrence and non-explosion of the Brownian motion on Riemannian manifolds. *Bull. Amer. Math. Soc. (N.S.)*, 36(2):135–249, 1999.

[7] K. Grove. Metric differential geometry. In *Differential geometry (Lyngby, 1985)*, volume 1263 of *Lecture Notes in Math.*, pages 171–227. Springer, Berlin, 1987.

[8] Y. Katznelson. *An Introduction to Harmonic Analysis*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, third edition, 2004.

[9] T. Kemp. Heat kernel empirical laws on $U_n$ and $GL_n$. *Journal of Theoretical Probability*, pages 1–55, 2015.

[10] E. S. Meckes and M. W. Meckes. Concentration and convergence rates for spectral measures of random matrices. *Probab. Theory Related Fields*, 156(1-2):145–164, 2013.

[11] E. S. Meckes and M. W. Meckes. Spectral measures of powers of random matrices. *Electron. Commun. Probab.*, 18:no. 78, 7, 2013.

[12] E. S. Meckes and M. W. Meckes. Rates of convergence for empirical spectral measures: a soft approach. *To appear in the IMA Volume on Discrete Structures*, pages 1–21, 2016.

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

[14] T. J. Rivlin. *An Introduction to the Approximation of Functions*. Dover Publications, Inc., New York, 1981. Corrected reprint of the 1969 original, Dover Books on Advanced Mathematics.

[15] L. Saloff-Coste. Precise estimates on the rate at which certain diffusions tend to equilibrium. *Math. Z.*, 217(4):641–677, 1994.

[16] F. Xu. A random matrix model from two-dimensional Yang-Mills theory. *Comm. Math. Phys.*, 190(2):287–307, 1997.

E-mail address: ese3@case.edu
E-mail address: melcher@virginia.edu
