Brownian Motion and the Classical Groups

Anthony D'Aristotile
SUNY at Plattsburgh

Persi Diaconis
Stanford University

Charles M. Newman
Courant Inst. of Math. Sciences

Abstract

Let Γ be chosen from the orthogonal group O_n according to Haar measure, and let A be an $n \times n$ real matrix with non-random entries satisfying $TrAA^t = n$. We show that $TrA\Gamma$ converges in distribution to a standard normal random variable as $n \to \infty$ uniformly in A. This extends a theorem of E. Borel. The result is applied to show that if entries $\beta_1, \cdots, \beta_{k_n}$ are selected from Γ where $k_n \to \infty$ as $n \to \infty$, then $\sqrt{\frac{n}{k_n}} \sum_{j=1}^{[k_n t]} \beta_j, 0 \leq t \leq 1$ converges to Brownian motion. Partial results in this direction are obtained for the unitary and symplectic groups.

Keywords: Brownian motion; sign-symmetry; classical groups; random matrix; Haar measure

1 Introduction

Let O_n be the group of $n \times n$ orthogonal matrices, and let Γ be chosen from the uniform distribution (Haar measure) on O_n. There are various senses in which the elements of $\sqrt{n}\Gamma$ behave like independent standard Gaussian random variables to good approximation when n is large.

To begin with, a classical theorem of Borel [6] shows that $P\{\sqrt{n}\Gamma_{11} \leq x\} \to \Phi(x)$ where $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt$. Theorems 2.1 and 2.2 below refine this, showing that an arbitrary linear combination of the elements of Γ is approximately normal: as $n \to \infty$,

$$\sup_{\substack{A \neq 0 \\ -\infty < x < \infty}} \left| P\left\{\frac{Tr(A\Gamma)}{\sqrt{||A||/\sqrt{n}}} \leq x\right\} - \Phi(x) \right| \to 0. \tag{1.1}$$

Here A ranges over all non-zero $n \times n$ matrices and $||A|| = Tr(AA^t)$; thus the normal approximation result is uniform in A. Borel's theorem follows by taking A to have a one in the one-one position and zeros elsewhere. When A above is the identity matrix, Diaconis and Mallows (see [11]) proved that $Tr\Gamma$ is approximately normal; this follows by taking A as the identity. As A varies, it follows that linear combinations of elements of Γ are also approximately normal. Interpolating between these facts and Borel's result, we prove that linking appropriately normalized entries from Γ yields in the limit standard Brownian motion. This is stated precisely in Theorem 3 below.

We give a little history. Borel's result is usually stated thus: Let X be the first entry of a point randomly chosen from the n-dimensional unit sphere. Then $P\{\sqrt{n}X \leq x\} \to \Phi(x)$ as n tends to ∞. Since the first row(or column) of a uniformly chosen orthogonal matrix is uniformly distributed on the unit sphere, Theorem 2.1 includes Borel's theorem. Borel, following earlier work by Mehler [31] and Maxwell [28, 29], proved the result as a rigorous version of the equivalence of ensembles in statistical mechanics. This says that features of the microcanonical ensemble (uniform on the sphere) are captured by the canonical ensemble (product Gauss measure). These results are often mistakenly attributed to Poincaré. See [15] for a careful history, rates of convergence, and applications to de Finetti type theorems for orthogonally invariant processes. The present project may be seen in the same light: the conditional distribution of an $n \times n$ matrix M with independent standard Gaussian coordinates, conditioned on $MM^T = I$ is Haar measure on the orthogonal group.

Borel also studied the joint distribution of several coordinates of $\sqrt{n}\ \Gamma$. His work was extended by Levy [24, 25, 26], Olshanski and Vershik [33] and Diaconis-Eaton-Lauritzen [13]. These last authors show that any $n^{\frac{1}{3}} \times n^{\frac{1}{3}}$ block of $\sqrt{n}\Gamma$ converges to product Gauss measure in total variation. They also give applications to versions of deFinettti's theorem suitable for regression and the analysis of variance. Extensions by McKean to infinite dimensions are in [30]; he writes that "It is fruitful to think of Wiener space as an infinite-dimensional sphere of radius $\sqrt{\infty}$." Our Theorem 3 gives one rigorous version of this fantastic statement. These ideas were developed by Hida [21]; see Kuo [23] for a recent account.

The study of global functionals such as the trace is carried out in [14, 16, 33]. In particular, the joint limiting distribution of $Tr(\Gamma), Tr(\Gamma^2), \cdots, Tr(\Gamma^k)$ is determined as that of independent normal variables. This turns out to be equivalent to a celebrated theorem of Szego and allows further study of the eigenvalues of Γ; see [7].

The eigenvalues of such random matrix models arise in dozens of situations and are currently being intensely studied. Mehta [32] gives a book length treatment. The area is in active development; see [12] for a recent survey. Interestingly, the eigenvalues of a Gaussian matrix have very different behavior from the eigenvalues of a random orthogonal matrix. In the first case they fill out the inside of the unit circle with order $\sqrt{n}$ of them on the real axis [2, 17]; in the second case the eigenvalues lie on the unit circle.

Brownian limits for partial traces are established in [10] and by Rains [34]. This last paper does much more, establishing results for partial traces of random matrices with law invariant under conjugation by O_n. This includes powers of Haar distributed matrices. One recent global result of Jiang [22] shows that the maximum entry of $\sqrt{n}\ \Gamma$ has the same limiting distribution as the maximum of n^2 standard normal variables. His method of proof gives an approximate coupling between the first J columns of Γ and J columns of standard normals for J of order $n/(\log n)^2$.

The uniform Gaussian limit for linear combinations of the entries of a random orthogonal matrix is proved in Section 2. This is used to prove Brownian motion limits in Section 3. The unitary and symplectic groups are treated in Sections 4 and 5. While we cannot prove completely parallel results, we can show that the sequences of partial sums along the diagonal, suitably normalized, converge to complex Brownian motions.

2 A Refinement of a Theorem of Borel

Our main tool will be obtained by extending a theorem of Borel [6]. A key to the analysis is that a Haar distributed element of the orthogonal group has entries that are invariant under the sign-change group. If Γ is an $n \times n$ orthogonal matrix and M is a random diagonal matrix with ± 1 chosen uniformly down the diagonal, then the diagonal entries of $M\Gamma$ are $\pm\Gamma_{ii}$. Under mild conditions on Γ_{ii}, sums of such entries are close to Gaussian by classical theory. If Γ is uniform on O_n, then $M\Gamma$ has the same law as Γ. The following result both makes this precise and more general.

Theorem 2.1. *For each positive integer n, choose any $n \times n$ real nonrandom matrix A with $||A|| = n$ (Here $||A|| = TrAA^t$), and let Γ be a random Haar distributed $n \times n$ orthogonal matrix. Then $TrA\Gamma$ converges in distribution to $N(0,1)$ as $n \to \infty$.*

Remark. The matrix A above depends on n. We have suppressed this in the notation. See Mallows [27] for further discussion of this method quantifying joint convergence of a growing vector to a vector of independent normals.

Proof. By singular value decomposition [20], there are orthogonal $n \times n$ matrices U and V such that $UAV = W$ where $W = Diag(a_1, ..., a_n)$ and $a_1 \geq a_2 \geq ... \geq a_n \geq 0$. Now

$$\begin{aligned} TrA\Gamma &= TrAVV^{-1}\Gamma = TrU(AVV^{-1}\Gamma)U^{-1} \\ &= Tr(UAV)(V^{-1}\Gamma U^{-1}). \end{aligned} \tag{2.2}$$

However, UAV is diagonal with non-negative, non-increasing entries and $V^{-1}\Gamma U^{-1}$ is random orthogonal by the invariance of Haar measure. We thus assume for the rest of the proof that A is diagonal with nonincreasing entries a_j and $\|A\| = n$.

If we write X_j for Γ_{jj}, then we have $TrA\Gamma = \Sigma_1^n a_j X_j$ which we may also write as S_n. We will show that $|E(e^{irS_n}) - e^{-\frac{r^2}{2}}|$ converges to 0. To do this, it is enough to demonstrate that for each real r there is a constant $L \geq 0$ such that, for each ϵ in $(0,1)$,

$$\limsup_{n\to\infty} |E(e^{irS_n} - e^{-\frac{r^2}{2}})| \leq L\epsilon. \tag{2.3}$$

This last assertion will hold if, given any subsequence n_l of the positive integers, there is a further subsequence n_{l_u} such that $|E(e^{irS_{n_{l_u}}}) - e^{-\frac{r^2}{2}})|$ is eventually less than or equal to $L\epsilon$.

Given $\epsilon > 0$, choose a positive integer $m \geq \frac{1}{\epsilon^2}$ so that $\frac{a_j^2}{n} \leq \epsilon^2$ for $j > m$. This is possible since by induction one can show that $a_j^2 \leq \frac{n}{j}$ for all j and all n (recall that a_i is non-increasing in i). Since $a_j \leq \sqrt{n}$, it is possible to choose n_{l_u} which satisfies

$$\frac{a_j}{\sqrt{n_{l_u}}} \to \alpha_j \text{ as } u \to \infty \text{ for } j = 1, \cdots, m. \tag{2.4}$$

Here $0 \leq \alpha_j \leq 1$. We must consider $E(e^{irS_{n_{l_u}}})$ but shall henceforth replace n_{ℓ_u} by n to simplify notation. Now $|E(e^{irS_n}) - e^{-\frac{r^2}{2}}|$ is less than or equal to the sum of the following 3 terms:

$$|E(e^{irS_n}) - e^{-\frac{r^2}{2}\frac{1}{n}\sum_{j=m+1}^n a_j^2} E(e^{ir\sum_{j=1}^m a_j X_j})|, \tag{2.5}$$

$$|e^{-\frac{r^2}{2}\frac{1}{n}\sum_{j=m+1}^n a_j^2} E(e^{ir\sum_{j=1}^m a_j X_j}) - e^{-\frac{r^2}{2}\frac{1}{n}\sum_{j=m+1}^n a_j^2} e^{-\frac{r^2}{2}\sum_{j=1}^m \alpha_j^2}|, \tag{2.6}$$

and

$$|e^{-\frac{r^2}{2}\frac{1}{n}\sum_{j=m+1}^n a_j^2 - \frac{r^2}{2}\sum_{j=1}^m \alpha_j^2} - e^{-\frac{r^2}{2}}|. \tag{2.7}$$

To bound (2.5) , first of all note that

$$E(e^{irS_n}) = E(e^{ir\sum_{j=1}^m a_jX_j} \prod_{j=m+1}^n e^{ira_jX_j}) \tag{2.8}$$

$$= E(e^{ir\sum_{j=1}^m a_jX_j} \prod_{j=m+1}^n (\cos(ra_jX_j) + i\sin(ra_jX_j))) \tag{2.9}$$

$$= E(e^{ir\sum_{j=1}^m a_jX_j} \prod_{j=m+1}^n \cos(ra_jX_j)). \tag{2.10}$$

To pass from (2.9) to (2.10), one should keep in mind the sign-symmetry of the X_j. In addition,

$$| \prod_{j=m+1}^n \cos(ra_jX_j) - e^{-\sum_{j=m+1}^n \frac{r^2}{2}a_j^2E(X_j^2)}| \tag{2.11}$$

$$\leq | \prod_{j=m+1}^n \cos(ra_jX_j) - e^{-\frac{r^2}{2}\sum_{j=m+1}^n a_j^2X_j^2}| +$$

$$|e^{-\frac{r^2}{2}\sum_{j=m+1}^n a_j^2X_j^2} - e^{-\frac{r^2}{2}\sum_{j=m+1}^n a_j^2E(X_j^2)}| \tag{2.12}$$

$$\leq r^4 \sum_{j=m+1}^n a_j^4X_j^4 + \frac{r^2}{2}| \sum_{j=m+1}^n a_j^2(X_j^2 - E(X_j^2))|. \tag{2.13}$$

To see that (2.12) is bounded above by (2.13), first take notice that for complex numbers $z_1, \cdots, z_n, w_1, \cdots, w_n$ of modulus less than or equal to 1, we have

$$|\prod_{j=1}^n z_j - \prod_{j=1}^n w_j| \leq \sum_{j=1}^n |z_j - w_j|.$$

This is easily proved by induction. Also, it is not hard to show that

$$|\cos(t) - e^{-\frac{t^2}{2}}| \leq t^4$$

for all real numbers t. Finally, one observes that $|e^{-a} - e^{-b}| \leq |a - b|$ for non-negative a, b.

In view of (2.8)-(2.10), (2.5) is equal to

$$| \int (e^{ir\sum_{j=1}^m a_jX_j} \prod_{j=m+1}^n \cos(ra_jX_j)$$

$$-e^{-\frac{r^2}{2}\sum_{j=m+1}^n a_j^2E(X_j^2)} e^{ir\sum_{j=1}^m a_jX_j}) \, dP|$$

(as we saw earlier that $E(X_j^2)=\frac{1}{n}$). Using (11)-(13), this is bounded by

$$\begin{aligned}
&\le \int |\prod_{j=m+1}^{n} \cos(ra_jX_j) - e^{-\frac{r^2}{2}\sum_{j=m+1}^{n} a_j^2 E(X_j^2)}|\, dP \\
&\le r^4 \int \sum_{j=m+1}^{n} a_j^4 X_j^4\, dP + \frac{r^2}{2}\int |\sum_{j=m+1}^{n} a_j^2 (X_j^2 - E(X_j^2))|\, dP \\
&= r^4 \sum_{j=m+1}^{n} a_j^4 E(X_j^4) + \frac{r^2}{2}\int |\sum_{j=m+1}^{n} a_j^2 X_j^2 - E(\sum_{j=m+1}^{n} a_j^2 X_j^2)|\, dP \\
&\le r^4 \sum_{j=m+1}^{n} a_j^4 E(X_j^4) + \frac{r^2}{2}(\mathrm{Var}(\sum_{j=m+1}^{n} a_j^2 X_j^2))^{\frac{1}{2}}. \qquad (2.14)
\end{aligned}$$

To obtain (2.14), which is our initial bound for (2.5), keep in mind that

$$\int |Y-EY|\, dP \le (\int |Y-EY|^2\, dP)^{\frac{1}{2}} = (\mathrm{Var} Y)^{\frac{1}{2}}$$

by Hölder's inequality.

We will return to (14) but first we claim that (2.7) converges to zero. Since

$$\begin{aligned}
\frac{1}{n}\sum_{j=m+1}^{n} a_j^2 &= \frac{1}{n}(n - \sum_{j=1}^{m} a_j^2) \\
&= 1 - \frac{1}{n}\sum_{j=1}^{m} a_j^2
\end{aligned}$$

which converges to $1-\sum_{j=1}^{m}\alpha_j^2$, our assertion is clear. It is also the case that (2.6) converges to zero. To see this, first note that since $e^{-\frac{r^2}{2}\frac{1}{n}\sum_{j=m+1}^{n} a_j^2}$ is bounded, it is enough to verify that $E(e^{ir\sum_{j=1}^{m} a_jX_j})$ converges to $e^{-\frac{r^2}{2}\sum_{j=1}^{m}\alpha_j^2}$. But this immediately follows from the fact [13] that the entries of the block matrix $[\sqrt{n}\Gamma_{ij}]_{1\le i,j\le m}$ are in the limit independent, each with the standard normal distribution.

From (2.14), and the previous paragraph, we have

$$\begin{aligned}
&|E(e^{irS_n}) - e^{\frac{r^2}{2}}| \\
&\le r^4 \sum_{j=m+1}^{n} a_j^4 E(X_j^4) + \frac{r^2}{2}(\mathrm{Var}(\sum_{j=m+1}^{n} a_j^2 X_j^2))^{\frac{1}{2}} + B_n
\end{aligned}$$

where $B_n \to 0$ as $n\to\infty$. Since ${X_j}^2$ has a beta distribution with parameters $1, n-1$ and thus $E(X_j^4)=\frac{3}{(n)(n+2)}\le\frac{3}{n^2}$ and $\frac{1}{n}\sum_{m+1}^{n} a_j^2\le 1$, we have

$$\frac{1}{n^2}\sum_{j=m+1}^{n} a_j^4 \le \frac{a_{m+1}^2}{n}\frac{1}{n}\sum_{j=m+1}^{n} a_j^2 \le \epsilon^2(1)=\epsilon^2. \qquad (2.15)$$

Therefore

$$\sum_{j=m+1}^{n} a_j^4 E(X_j^4) \le \frac{3}{n^2}\sum_{j=m+1}^{n} a_j^4 \le 3\epsilon^2.$$

Furthermore,

$$\operatorname{Var}(\sum_{j=m+1}^{n} a_j^2 X_j^2) = \sum_{m+1}^{n} a_j^4 \operatorname{Var}(X_j^2) + \sum_{\substack{j,k=m+1 \\ j\neq k}}^{n} a_j^2 a_k^2 \operatorname{Cov}(X_j^2, X_k^2).$$

Now

$$\begin{aligned}
\sum_{m+1}^{n} a_j^4 \operatorname{Var}(X_j^2) &= \sum_{j=m+1}^{n} a_j^4 (E(X_j^4) - \frac{1}{n^2}). \\
&= \sum_{j=m+1}^{n} a_j^4 E(X_j^4) - \frac{1}{n^2} \sum_{j=m+1}^{n} a_j^4 \\
&\leq \frac{3}{n^2} \sum_{j=m+1}^{n} a_j^4 - \frac{1}{n^2} \sum_{j=m+1}^{n} a_j^4 \\
&= \frac{2}{n^2} \sum_{j=m+1}^{n} a_j^4 \\
&\leq 2\epsilon^2. \qquad (2.16)
\end{aligned}$$

To obtain (2.16) we can appeal to (2.15).

By expanding and taking expectations of both sides of

$$1 = (\sum_{j=1}^{n} \Gamma_{1j}^2)(\sum_{j=1}^{n} \Gamma_{2j}^2),$$

it follows that

$$1 = nE(\Gamma_{11}^2 \Gamma_{21}^2) + n(n-1)E(\Gamma_{11}^2 \Gamma_{22}^2).$$

Thus

$$E(\Gamma_{11}^2 \Gamma_{22}^2) = \frac{1 - nE(\Gamma_{11}^2 \Gamma_{21}^2)}{n(n-1)} \leq \frac{1}{n(n-1)}.$$

Therefore, for $j \neq k$,

$$\begin{aligned}
\operatorname{Cov}(X_j^2, X_k^2) &= E(X_j^2 X_k^2) - \frac{1}{n^2} \\
&\leq \frac{1}{n(n-1)} - \frac{1}{n^2}.
\end{aligned}$$

One then easily verifies that for $n \geq 2$

$$\operatorname{Cov}(X_j^2, X_k^2) \leq \frac{2}{n^3}.$$

Thus

$$\begin{aligned}
&\sum_{\substack{j,k=m+1 \\ j\neq k}}^{n} a_j^2 a_k^2 \operatorname{Cov}(X_j^2, X_k^2) \\
&\leq 2\frac{1}{n}(\frac{1}{n} \sum_{j=m+1}^{n} a_j^2)(\frac{1}{n} \sum_{k=m+1}^{n} a_k^2) \leq \frac{2}{n}
\end{aligned}$$

which converges to zero.

We have

$$|E(e^{irS_n}) - e^{-\frac{r^2}{2}}| \leq r^4 3\epsilon^2 + \frac{r^2}{2}(2\epsilon^2 + \frac{2}{n})^{\frac{1}{2}} + B_n$$

where $B_n \to 0$ as $n \to \infty$.

This yields (2.3) for some L depending on r, as desired, and we are done. □

Our next result shows that the convergence in Theorem 2.2 is uniform in A. We only work with diagonal matrices A here but singular value decomposition says that this suffices. We then find it convenient to think of A as a point of a sphere of radius $\sqrt{n}$.

Theorem 2.2. *Let Γ, $X_j = \Gamma_{jj}$ be as in Theorem 2.1 , and let A_n be the surface of the sphere of radius $\sqrt{n}$ in $\mathbb{R}^n$. For $v = (a_1, ..., a_n) \in A_n$, write $S_n(v)$ for $\sum_{j=1}^n a_j X_j$. Then S_n converges in distribution to $N(0,1)$ uniformly on A_n, i.e., as $n \to \infty$,*

$$\sup_{x \in \mathbb{R},\, v \in A_n} |P(S_n(v) \leq x) - \Phi(x)| \to 0.$$

Proof. We first verify that the family $\mathcal{F} = \{S_n(v) : v \in A_n, \quad n = 1, 2, ...\}$ is tight. Corresponding to any sequence $\mathcal{S}$ of $\mathcal{F}$, either there is a positive integer Y such that $\mathcal{S}$ is contained in the family $\{S_j(v_j) : v_j \in A_j, 1 \leq j \leq Y\}$ or $\mathcal{S}$ has a sub-sequence $S_{n_l}(v_{n_l})$ where $n_l \to \infty$. In the first case, $\mathcal{S}$ has a sub-sequence of the form $S_k(p_{ku})$ where k is a fixed positive integer, $1 \leq k \leq Y$, and $p_{ku} = (a_{1u}, ..., a_{ku}) \in A_k$ for $u = 1, 2, ...$. Choose a sub-sequence u_l of the positive integers such that $a_{ru_l} \to b_r$ for $1 \leq r \leq k$. Plainly $S_k(p_{ku_l}) \Rightarrow S_k(w)$ where $w = (b_1, ..., b_k)$. In the second case, the argument of Theorem 1 shows that $S_{n_l}(v_{n_l}) \Rightarrow N(0,1)$. Thus $\mathcal{F}$ is tight.

It is easy to see that because of tightness, it suffices to show, as we now do, that for any interval $[a, b] \subseteq \mathbb{R}$

$$\lim_{n \to \infty} \sup_{x \in [a,b],\, v \in A_n} |P(S_n(v) \leq x) - \Phi(x)| = 0.$$

If false, there exists an $\epsilon_0 > 0$, a sub-sequence $n_l \to \infty$, points $x_{n_l} \in [a, b]$, and elements $v_{n_l} \in A_{n_l}$ such that

$$|P(S_{n_l}(v_{n_l}) \leq x_{n_l}) - \Phi(x_{n_l})| \geq \epsilon_0.$$

Now x_{n_l} has a non-increasing or non-decreasing sub-sequence $x_{n_{l_u}}$ which converges to $x \in [a, b]$. We assume without loss of generality that $x_{n_{l_u}}$ is non-decreasing. We henceforth work with n_{l_u} but suppress the subsequence notation. Note that

$$P(S_n(v_n) \leq x) = P(S_n(v_n) \leq x_n) + P(x_n < S_n(v_n) \leq x).$$

Since $S_n(v_n) \Rightarrow N(0,1)$, it is clear that $P(x_n < S_n(v_n) \leq x) \to 0$ and hence that $P(S_n(v_n) \leq x_n) \to \Phi(x)$. Since $\Phi(x_n) \to \Phi(x)$, we obtain a contradiction which proves our claim. □

3 Orthogonal Matrices

We use the results of Section 2 to prove the main theorem of this section. This shows that if any growing selection of entries of a random orthogonal matrix are linked together in the classical way, a limiting standard Brownian motion results. To set up our notation, let $\Gamma = (\Gamma_{ij})_{i,j=1}^{n}$ be an $n \times n$ orthogonal matrix distributed by Haar measure. Choose a subset of size k_n from among the entries of Γ. Suppose the entries are $\beta_1, \beta_2, \ldots, \beta_{k_n}$ with β_j corresponding to e.g. lexicographic order of $\Gamma_{rs} : (r,s) < (x,y)$ if $r < x$ or if $r = x$ and $s < y$. To denote this ordering we write $\beta_1 \sim \Gamma_{11}$, $\beta_2 \sim \Gamma_{12}$, ..., $\beta_{n+1} \sim \Gamma_{21}$, etc.

Theorem 3.1. *Let $\beta_1, \beta_2, \ldots, \beta_{k_n}$ be entries of a Haar distributed random matrix in O_n, as above. Assume that $k_n \nearrow \infty$. If for ℓ in $\{1, \ldots, k_n\}$ and t in $[0,1]$,*

$$S_\ell^{(n)} = \sqrt{\frac{n}{k_n}} \sum_{j=1}^{\ell} \beta_j, \quad X_n(t) = S_{[k_n t]}^{(n)},$$

then $X_n \Longrightarrow W$, a standard Brownian motion, as $n \to \infty$.

Proof. We first prove that the finite-dimensional distributions of X_n converge to the corresponding distributions of W. For a single time point t, we must prove that

$$X_n(t) \Longrightarrow N(0,t) = W_t \quad \text{as } n \to \infty.$$

However, this is equivalent to

$$\frac{1}{\sqrt{t}} X_n(t) \Longrightarrow N(0,1).$$

For each n, let $A = (a_{ij})_{i,j=1}^{n}$ be the $n \times n$ real matrix defined as follows :

$$a_{rs} = \begin{cases} \frac{1}{\sqrt{t}}\sqrt{\frac{n}{k_n}} & \text{if } \beta_i \sim \Gamma_{sr}, \\ & \quad \text{for some } i, 1 \le i \le [k_n t] \\ (n - \frac{[k_n t] n}{k_n t})^{\frac{1}{2}} & \text{if } \beta_{[k_n t]+1} \sim \Gamma_{sr} \\ 0 & \text{otherwise} \end{cases}$$

Note that $||A|| = n$ and

$$Tr(A\Gamma) = \frac{1}{\sqrt{t}} X_n(t) + (n - \frac{[k_n t] n}{k_n t})^{\frac{1}{2}} \beta_{[k_n t]+1}$$

which converges to $N(0,1)$ in distribution by Theorem 2.1. However, $0 \le n - \frac{[k_n t] n}{k_n t} \le \frac{n}{k_n t}$ and so, by [5], it suffices to show that $\sqrt{\frac{n}{k_n t}} \beta_{[k_n t]+1} \to 0$ in probability, which folows from $k_n \to \infty$ and the fact [13] that

$$\sqrt{n} \beta_{[k_n t]+1} \Longrightarrow N(0,1).$$

We now consider two time points s and t with $s < t$. By the Cramer-Wold device [5], it is enough to show that

$$aX_n(s) + b(X_n(t) - X_n(s)) \Longrightarrow aW_s + b(W_t - W_s)$$

for any $(a,b) \in \mathbb{R}^2$. However, this is equivalent to showing that

$$\frac{a}{C(s,t)} S_{[k_n s]} + \frac{b}{C(s,t)} (S_{[k_n t]} - S_{[k_n s]}) \Longrightarrow N(0,1)$$

where

$$C(s,t) = (a^2 s + b^2(t-s))^{\frac{1}{2}}.$$

This can be shown by choosing an appropriate sequence of matrices A, as follows, and again applying Theorem 2.1.

First note that

$$\begin{aligned}
&[k_n s]a^2 + ([k_n t] - [k_n s])b^2 \\
&\geq (k_n s - 1)a^2 + ((k_n t - 1) - k_n s)b^2 \\
&= k_n s a^2 - a^2 + k_n t b^2 - b^2 - k_n s b^2 \\
&= k_n C^2(s,t) - (a^2 + b^2)
\end{aligned}$$

Also observe that

$$\begin{aligned}
&[k_n s]a^2 + ([k_n t] - [k_n s])b^2 \\
&\leq k_n s a^2 + k_n t b^2 - (k_n s - 1)b^2 \\
&= k_n s a^2 + k_n t b^2 - k_n s b^2 + b^2 \\
&= k_n C^2(s,t) + b^2
\end{aligned}$$

Combining these facts, we have

$$\begin{aligned}
&-\frac{nb^2}{k_n C^2(s,t)} \\
&\leq (n - \frac{[k_n s]na^2}{k_n C^2(s,t)} - \frac{([k_n t] - [k_n s])nb^2}{k_n C^2(s,t)}) \\
&\leq \frac{n(a^2+b^2)}{k_n C^2(s,t)} .
\end{aligned}$$

With these preliminaries, we define the matrix A in two cases. If $n - \frac{[k_n s]na^2}{k_n C^2(s,t)} - \frac{([k_n t]-[k_n s])nb^2}{k_n C^2(s,t)} \geq 0$, let $A = (a_{i,j})_{i,j=1}^n$ be defined as follows :

$$a_{uv} = \begin{cases} \frac{a}{C(s,t)} \sqrt{\frac{n}{k_n}} & \text{if } \beta_i \sim \Gamma_{vu}, \\ & \text{for some } i, 1 \leq i \leq [k_n s] \\ \frac{b}{C(s,t)} \sqrt{\frac{n}{k_n}} & \text{if } \beta_i \sim \Gamma_{vu}, \text{ for some} \\ & i, [k_n s]+1 \leq i \leq [k_n t] \\ (n - \frac{[k_n s]na^2}{k_n C^2(s,t)} - \frac{([k_n t]-[k_n s])nb^2}{k_n C^2(s,t)})^{\frac{1}{2}} & \text{if } \beta_{[k_n t]+1} \sim \Gamma_{vu} \\ 0 & \text{otherwise} \end{cases}$$

On the other hand, if $n - \frac{[k_n s]na^2}{k_n C^2(s,t)} - \frac{([k_n t]-[k_n s])nb^2}{k_n C^2(s,t)} < 0$,
we define $A = (a_{i,j})_{i,j=1}^n$ by :

$$a_{uv} = \begin{cases} \frac{a}{C(s,t)}\sqrt{\frac{n}{k_n}} & \text{if } \beta_i \sim \Gamma_{vu}, \\ & \text{for some } i, 1 \le i \le [k_n s] \\ \frac{b}{C(s,t)}\sqrt{\frac{n}{k_n}} & \text{if } \beta_i \sim \Gamma_{vu}, \\ & \text{for some } i, [k_n s]+1 \le i \le ([k_n t]-1) \\ (n - \frac{[k_n s]na^2}{k_n C^2(s,t)} - \frac{(([k_n t]-1)-[k_n s])nb^2}{k_n C^2(s,t)})^{\frac{1}{2}} & \text{if } \beta_{[k_n t]} \sim \Gamma_{vu} \\ 0 & \text{otherwise} \end{cases}$$

Note that in either case $||A|| = n$ and so $Tr(A\Gamma)) \Rightarrow N(0,1)$ by Theorem 2.1. However, it is plain that $\frac{a}{C(s,t)}S_{[k_n s]} + \frac{b}{C(s,t)}(S_{[k_n t]} - S_{[k_n s]})$ differs from $Tr(A\Gamma)$ by a quantity in absolute value bounded by $\frac{\sqrt{a^2+b^2}}{C(s,t)}\sqrt{\frac{n}{k_n}}\gamma$ where γ is an entry of the random $n \times n$ random Γ. Thus, as before, what remains is to show that $\frac{\sqrt{a^2+b^2}}{C(s,t)}\sqrt{\frac{n}{k_n}}\gamma$ converges to zero in probability. Thus, given $\epsilon > 0$

$$P(|\frac{\sqrt{a^2+b^2}}{C(s,t)}\sqrt{\frac{n}{k_n}}\gamma| < \epsilon) = P(|\sqrt{n}\gamma| < \frac{C(s,t)}{\sqrt{a^2+b^2}}\sqrt{k_n}\epsilon)$$

which converges to 1 as $n \to \infty$.

A similar argument shows that the higher order finite dimensional distributions behave properly.

We next show that X_n is tight. According to Theorem 15.6 of [5], it is enough to show that for sufficiently large n

$$E(|X_n(t) - X_n(t_1)|^2|X_n(t_2) - X_n(t)|^2) \le K(t_2 - t_1)^2, \quad t_1 \le t \le t_2, \tag{3.17}$$

for K independent of $n, t_1, t,$ and t_2. The left member of the above expression is

$$\frac{n^2}{k_n^2}\sum E(\beta_i\beta_j\beta_k\beta_l)$$

where $[k_n t_1] < i, j \le [k_n t]$ and $[k_n t] < k, l \le [k_n t_2]$. Put $[k_n t] - [k_n t_1] = m_1$ and $[k_n t_2] - [k_n t] = m_2$. The left member of (3.17) is bounded from above by

$$\frac{n^2}{k_n^2}(aE(\Gamma_{11}^2\Gamma_{12}^2) + bE(\Gamma_{11}^2\Gamma_{22}^2))$$

where a and b are both less than or equal to $m_1 m_2$. Here we have used the fact that for distinct entries δ, β, α and σ of a random orthogonal matrix, $E(\delta^2\beta\alpha) = 0$ and $E(\delta\beta\alpha\sigma)$ is non-positive. The first assertion uses the fact that for any (nonrandom) diagonal sign matrix M, the random matrices ΓM and $M\Gamma$ are equidistributed with Γ; the second assertion uses that and also the fact that $E(\gamma_{11}\gamma_{12}\gamma_{21}\gamma_{22}) = -\frac{1}{(n-1)n(n+2)}$ [36]. However, both $n^2E(\Gamma_{11}^2\Gamma_{12}^2)$ and $n^2E(\Gamma_{11}^2\Gamma_{22}^2)$ converge to 1 , and so for all n, both expectations are less than $\frac{L}{n^2}$ for some positive constant L. Combining all of this information, we have that

$$
\begin{aligned}
&E(|X_n(t) - X_n(t_1)|^2 \; |X_n(t_2) - X_n(t)|^2) \\
&\leq L m_1 m_2 \frac{1}{k_n^2} \leq L(\frac{m_1 + m_2}{k_n})^2 \\
&\leq L(\frac{k_n t_2 - (k_n t_1 - 1)}{k_n})^2 \leq L(t_2 - t_1 + \frac{1}{k_n})^2
\end{aligned}
$$

If $\frac{1}{k_n} \leq t_2 - t_1$, then $\frac{m_1+m_2}{k_n} \leq 2(t_2 - t_1)$ while if $\frac{1}{k_n} > t_2 - t_1$, then either $X_n(t) - X_n(t_1)$ or $X_n(t_2) - X_n(t)$ is zero. Thus our claim is established. □

4 Unitary Matrices

We first thought that obtaining unitary analogues of Theorems 1, 2, 3 would be straightforward but then encountered difficulties in translating to the complex case because of the lack of a singular value decomposition. This led us to carefully redo the preliminaries. Our main results are Theorems 5 and 6 below.

For the proof of Theorem 6, it will be necessary to first establish Theorem 4, which is the analogue of Theorem 15.1 of [5]. To this end, let $\mathcal{R}^k$, $\mathcal{D}$ and $\mathcal{E}$ denote respectively the Borel sets of $\mathbb{R}^k$, D and $D \times D$, where D is the Skorokhod space of right-continuous real-valued functions on $[0,1]$ with left limits. For $t_1, \cdots, t_k$ in [0,1], define

$$\pi_{t_1,\cdots,t_k} : D \to \mathbb{R}^k$$

by $\pi_{t_1,\cdots,t_k}(x) = (x(t_1), \cdots, x(t_k))$ for $x \in D$. Following Billingsley [5], sets of the form $\pi^{-1}_{t_1,\cdots,t_k}(H)$ where $H \in \mathcal{R}^k$ are subsets of D and called finite-dimensional sets.

If T_0 is a subset of [0,1], let $\mathcal{F}_{T_0}$ be the collection of sets $\pi^{-1}_{t_1,\cdots,t_k}(H)$ where $k \geq 1, t_i \in T_0$, and $H \in \mathcal{R}^k$. Then $\mathcal{F}_{T_0}$ is an algebra of sets, i.e., $\mathcal{F}_{T_0}$ is closed under finite unions and finite intersections and the empty set $\emptyset \in \mathcal{F}_{T_0}$. See Royden [35] for more details. Obviously, $\mathcal{F}_{[0,1]}$ is the class of finite-dimensional sets. Billingsley has shown (Theorem 14.5 of [5]) that if T_0 contains 1 and is dense in [0,1], then $\mathcal{F}_{T_0}$ generates $\mathcal{D}$.

Extending these ideas, for $s_1, \cdots, s_k$ and $t_1, \cdots, t_l$ in [0,1], define

$$\pi_{s_1,\cdots,s_k;t_1,\cdots,t_l} : D \times D \to \mathbb{R}^k \times \mathbb{R}^l$$

by sending (x, y) to $(x(s_1), \cdots, x(s_k); y(t_1), \cdots, y(t_l))$. Subsets of $D \times D$ of the form

$$\pi^{-1}_{s_1,\cdots,s_k;t_1,\cdots,t_l}(H \times K)$$

where $H \in \mathcal{R}^k$, $K \in \mathcal{R}^l$ are called finite-dimensional sets (of $D \times D$).
If T_0 and T_1 are subsets of [0,1], let $\mathcal{F}_{T_0,T_1}$ be the class of sets

$$\pi^{-1}_{s_1,\cdots,s_k;t_1,\cdots,t_l}(H \times K)$$

where $s_i \in T_0$, $t_j \in T_1$, $k \geq 1$, $l \geq 1$, $H \in \mathcal{R}^k$, and $K \in \mathcal{R}^l$. One can easily verify that $\mathcal{F}_{T_0,T_1}$ is a semi-algebra of sets, i.e., the intersection of any two members of $\mathcal{F}_{T_0,T_1}$ is again in $\mathcal{F}_{T_0,T_1}$ and the complement of any set in $\mathcal{F}_{T_0,T_1}$ is a finite disjoint union of elements of $\mathcal{F}_{T_0,T_1}$. If we let $\mathcal{A}$ be all finite disjoint unions of members of $\mathcal{F}_{T_0,T_1}$, then $\mathcal{A}$ is an algebra of sets in $D \times D$ (any semialgebra generates an algebra in this way [35]).

Suppose T_0 and T_1 are both dense subsets of [0,1] and that $1 \in T_0 \cap T_1$. Let $\mathcal{L}$ be the σ-algebra of subsets of $D \times D$ generated by $\mathcal{F}_{T_0,T_1}$. Sets of the form

$$\pi_{s_1,\cdots,s_k}(H) \times D$$

where H is in $\mathcal{R}^k$ and $s_1, \cdots, s_k \in T_0$ are in $\mathcal{F}_{T_0,T_1}$ and may be identified with $\mathcal{F}_{T_0}$. Since $\mathcal{F}_{T_0}$ generates $\mathcal{D}$, it is clear that $G \times D \in \mathcal{L}$ for all open sets G of D. Similarly $D \times L \in \mathcal{L}$ for all L open in D and so $\mathcal{L}$ contains all sets $G \times L$ where G, L are open in D. It is now plain that $\mathcal{E} \subseteq \mathcal{L}$.

On the other hand, Billingsley has shown that

$$\pi_{s_1,\cdots,s_k} : (D, \mathcal{D}) \to (\mathbb{R}^k, \mathcal{R}^k)$$

is a measurable mapping. In a completely analogous way, it can be shown that

$$\pi_{s_1,\cdots,s_k;t_1,\cdots,t_l} : (D \times D, \mathcal{E}) \to (\mathbb{R}^k \times \mathbb{R}^l, \mathcal{R}^k \times \mathcal{R}^l)$$

is also measurable (here $\mathcal{R}^k \times \mathcal{R}^l$ is the σ-algebra of subsets of $\mathbb{R}^k \times \mathbb{R}^l$ generated by "measurable rectangles" of the form $H \times K$ where $H \in \mathcal{R}^k,\ K \in \mathcal{R}^l$). This σ-algebra is precisely the σ-algebra of Borel sets of $\mathbb{R}^k \times \mathbb{R}^l$ (see [4]). It follows that the finite-dimensional subsets of $D \times D$ lie in $\mathcal{E}$ by definition of measurable mapping. Thus $\mathcal{L} \subseteq \mathcal{E}$ and so we have $\mathcal{L} = \mathcal{E}$.

Suppose P and Q are two probability measures on $(D \times D, \mathcal{E})$ which agree on $\mathcal{F}_{T_0,T_1}$. Then they clearly agree on the σ-algebra $\mathcal{A}$ generated by $\mathcal{F}_{T_0,T_1}$. Since $\mathcal{A}$ generates $\mathcal{E}$, it follows that $P = Q$ on $\mathcal{E}$ by Theorem 3. 2 of [4]. In the language of Billingsley [5], for T_0, T_1 dense in [0,1] with $1 \in T_0 \cap T_1$, $\mathcal{F}_{T_0,T_1}$ is a "determining class."

If P is a probability measure on $(D, \mathcal{D})$, let T_P be the set of all points $t \in [0, 1]$ such that π_t is continuous except on a subset of D which has P-measure 0. Billingsley [5] has shown that T_P contains 0 and 1 and its complement in [0,1] is at most countable. Now let P be a probability measure on $(D \times D, \mathcal{E})$ with marginals R_1 and R_2. If $s_1, \cdots, s_k \in T_{R_1}$ and $t_1, \cdots, t_l \in T_{R_2}$, then $\pi_{s_1,\cdots,s_k}$ is continuous except on a subset A of D of R_1-measure zero. Similarly $\pi_{t_1,\cdots,t_l}$ is continuous except on a subset B of D of R_2-measure zero. Now $(A \times D) \cup (D \times B)$ has P-measure 0 and off this set $\pi_{s_1,\cdots,s_k;t_1,\cdots,t_l}$ is continuous. We will need the following:

Theorem 4.1. *Let P_n, $n = 1, 2, \cdots$, and P be probability measures on $(D \times D, \mathcal{E})$. Suppose R_1 and R_2 are the marginal probability measures of P. If $\{P_n\}$ is tight and if $P_n \pi^{-1}_{s_1,\cdots,s_k;t_1,\cdots,t_l} \Rightarrow P\pi^{-1}_{s_1,\cdots,s_k;t_1,\cdots,t_l}$ holds whenever all the s_i are in T_{R_1} and all the t_j are in T_{R_2}, then $P_n \Rightarrow P$.*

Proof. Since $\{P_n\}$ is tight, each subsequence $\{P_{n'}\}$ contains a further subsequence $\{P_{n''}\}$ converging weakly to some limit Q. By Theorem 2 of [5], it suffices to show that each such Q is equal to P.

Suppose Q_1 and Q_2 are the marginals of Q. If $s_1, \cdots, s_k$ all lie in $T_{R_1} \cap T_{Q_1}$ and $t_1, \cdots, t_l$ all lie in $T_{R_2} \cap T_{Q_2}$, then

$$P_{n''} \pi^{-1}_{s_1,\cdots,s_k;t_1,\cdots,t_l} \Rightarrow P\pi^{-1}_{s_1,\cdots,s_k;t_1,\cdots,t_l}$$

by hypothesis. Also $\pi_{s_1,\cdots,s_k;t_1,\cdots,t_l}$ is continuous except on a subset of $D \times D$ of Q-measure zero by comments preceding the statement of the theorem. Since $P_{n''} \Rightarrow Q$, it follows by Theorem 5.1 of [5] that

$$P_{n''}\pi^{-1}_{s_1,\cdots,s_k;t_1,\cdots,t_l} \Rightarrow Q\pi^{-1}_{s_1,\cdots,s_k;t_1,\cdots,t_l}.$$

Thus

$$P\pi^{-1}_{s_1,\cdots,s_k;t_1,\cdots,t_l} = Q\pi^{-1}_{s_1,\cdots,s_k;t_1,\cdots,t_l}$$

whenever each $s_i \in T_{R_1} \cap T_{Q_1}$ and each $t_j \in T_{R_2} \cap T_{Q_2}$. Let $T_1 = T_{R_1} \cap T_{Q_1}$ and $T_2 = T_{R_2} \cap T_{Q_2}$. Each of T_1 and T_2 is dense in [0,1] and $1 \in T_1 \cap T_2$ and so as we have seen above, $\mathcal{F}_{T_1,T_2}$ is a determining class. The above equality says that P and Q agree on $\mathcal{F}_{T_1,T_2}$ and we are done. □

We are now in a position to establish the complex analogue of Theorem 2.1 (for diagonal A).

Theorem 4.2. *Let $A = Diag(a_1, ..., a_n)$ and $B = Diag(b_1, ..., b_n)$ where $a_1 \geq a_2 \geq ... \geq a_n$ and $b_1 \geq b_2 \geq ... \geq b_n$ and $||A|| = ||B|| = n$, and let $\Delta = \Gamma + i\Lambda$ be an $n \times n$ unitary matrix distributed by Haar measure. Then $(TrA\Gamma, TrB\Lambda) \Rightarrow \frac{1}{\sqrt{2}}(Z_1, Z_2)$ as $n \to \infty$, where Z_1 and Z_2 are i.i.d. standard normal (i.e., $TrA\Gamma + iTrB\Gamma$ converges in distribution to a complex standard normal distribution).*

Proof. By the Cramer-Wold device [5], it suffices to prove that

$$xTrA\Gamma + yTrB\Lambda \Rightarrow x\frac{1}{\sqrt{2}}Z_1 + y\frac{1}{\sqrt{2}}Z_2$$

for arbitrary $(x, y) \in \mathbb{R}^2$. Write X_j for γ_{jj} and Y_j for λ_{jj} with $\gamma_{ij}, \lambda_{ij}$ the entries of Γ and Λ. We will show that

$$|E(e^{ir(x\sum_{j=1}^n a_jX_j + y\sum_{j=1}^n b_jY_j)} - e^{-\frac{1}{4}x^2r^2}e^{-\frac{1}{4}y^2r^2})| \tag{4.18}$$

converges to zero. We follow the proof of Theorem 2.1 and show that there is a constant $L > 0$ such that, for each $\epsilon > 0$, the lim sup of (4.18) is less or equal to $L\epsilon$.

Given $\epsilon > 0$, choose a positive integer $m \geq \frac{1}{\epsilon^2}$ so that $\frac{a_j^2}{n} \leq \epsilon^2$ and $\frac{b_j^2}{n} \leq \epsilon^2$ for $j > m$ and all n. Given any subsequence n_l of the positive integers, choose a subsequence n_{l_s} which satisfies

$$\frac{a_j}{\sqrt{n_{l_s}}} \to \alpha_j, \quad \frac{b_j}{\sqrt{n_{l_s}}} \to \beta_j, \quad \text{as } \mu \to \infty \text{ for } j = 1, 2, ...m. \tag{4.19}$$

As before, we will suppress the subsequence notation.

The quantity (4.18) is less than or equal to the sum of the following three terms

$$|E(e^{ir(x\sum_{j=1}^n a_jX_j + y\sum_{j=1}^n b_jY_j)} - e^{-\frac{r^2}{2}\frac{1}{2n}(x^2\sum_{j=m+1}^n a_j^2 + y^2\sum_{j=m+1}^n b_j^2)} \cdot E(e^{ir(x\sum_{j=1}^m a_jX_j + y\sum_{j=1}^m b_jY_j)})|, \tag{4.20}$$

$$|e^{-\frac{r^2}{2}\frac{1}{2n}(x^2\sum_{j=m+1}^n a_j^2 + y^2\sum_{j=m+1}^n b_j^2)}E(e^{ir(x\sum_{j=1}^m a_jX_j + y\sum_{j=1}^m b_jY_j)}) - e^{-\frac{r^2}{2}\frac{1}{2n}x^2\sum_{j=m+1}^n a_j^2}e^{-\frac{r^2}{4}x^2\sum_{j=1}^m \alpha_j^2}e^{-\frac{r^2}{2}\frac{1}{2n}y^2\sum_{j=m+1}^n b_j^2}e^{-\frac{r^2}{4}y^2\sum_{j=1}^m \beta_j^2}|, \tag{4.21}$$

$$|e^{-\frac{r^2}{2}\frac{1}{2n}x^2\sum_{j=m+1}^n a_j^2}e^{-\frac{r^2}{4}x^2\sum_{j=1}^m \alpha_j^2}e^{-\frac{r^2}{2}\frac{1}{2n}y^2\sum_{j=m+1}^n b_j^2} \cdot e^{-\frac{r^2}{4}y^2\sum_{j=1}^m \beta_j^2}| - e^{-\frac{1}{4}x^2r^2}e^{-\frac{1}{4}y^2r^2}|. \tag{4.22}$$

Since

$$\frac{1}{n}\sum_{j=m+1}^{n} a_j^2 \to 1-\sum_{j=1}^{m}\alpha_j^2 \quad \text{and}$$
$$\frac{1}{n}\sum_{j=m+1}^{n} b_j^2 \to 1-\sum_{j=1}^{m}\beta_j^2,$$

the term (4.22) converges to zero.

By a known result (see, e.g., Lemma 5.3 of [33]),

$$(\sqrt{n}X_1,...,\sqrt{n}X_m,\sqrt{n}Y_1,...,\sqrt{n}Y_m) \Rightarrow \frac{1}{\sqrt{2}}(Z_1,Z_2,...,Z_{2m})$$

where the Z_i are i.i.d. $N(0,1)$. Thus

$$\begin{aligned}&(a_1xX_1,...,a_mxX_m,b_1yY_1,...,b_myY_m)\\ &=(\frac{a_1}{\sqrt{n}}\sqrt{n}xX_1,...,\frac{a_m}{\sqrt{n}}\sqrt{n}xX_m,\frac{b_1}{\sqrt{n}}\sqrt{n}yY_1,...,\frac{b_m}{\sqrt{n}}\sqrt{n}yY_m)\\ &\Rightarrow \frac{1}{\sqrt{2}}(\alpha_1xZ_1,\alpha_2xZ_2,...,\alpha_mxZ_m,\beta_1yZ_{m+1},...,\beta_myZ_{2m})\end{aligned}$$

and so

$$E(e^{ir(x\sum_{j=1}^m a_jX_j+y\sum_{j=1}^m b_jY_j)}) \to e^{-\frac{r^2}{4}x^2\sum_{j=1}^m\alpha_j^2}\, e^{-\frac{r^2}{4}y^2\sum_{j=1}^m\beta_j^2}$$

and hence (4.21) converges to zero.

To bound (4.20), we first claim that

$$\begin{aligned}&E(e^{ir(x\sum_{j=1}^n a_jX_j+y\sum_{j=1}^n b_jY_j)})\\ &=E(e^{irx\sum_{j=1}^m a_jX_j}\, e^{iry\sum_{j=1}^m b_jY_j}\prod_{j=m+1}^{n}\cos(rxa_jX_j)\prod_{j=m+1}^{n}\cos(ryb_jY_j)).\end{aligned}$$

To see this, let

$$G=e^{irx\sum_{j=1}^m a_jX_j}\, e^{iry\sum_{j=1}^m b_jY_j}$$

and note that

$$e^{ir(x\sum_{j=1}^n a_jX_j+y\sum_{j=1}^n b_jY_j)}=G(\prod_{j=m+1}^{n}\cos(rxa_jX_j)\prod_{j=m+1}^{n}\cos(ryb_jY_j))$$

plus a sum of products of the form GJ where J is a product of sines and cosines involving at least one sine term.

To establish our claim, it is enough to verify that the expectation of any such GJ is zero. First suppose J contains the factor $\sin(rxa_jX_j)$ but not the factor $\sin(ryb_jY_j)$. Then $E(GJ)=0$ by the sign-symmetry of the diagonal elements of Δ. Next consider a product GJ containing a factor $\sin(rxa_jX_j)\sin(ryb_jY_j)$. The diagonal elements of Δ are also exchangeable, and so we can assume $j=m+1$. Write

$$\begin{aligned}&GJ=H\sin(rxa_{m+1}X_{m+1})\sin(ryb_{m+1}Y_{m+1}),\\ &X_{m+1}+iY_{m+1}=se^{i\gamma},\quad \text{and}\\ &\int_{U_n}GJ\;d\mu_n=I\end{aligned}$$

where U_n is the unitary group and μ_n is Haar measure. For $\theta \in [0, 2\pi]$, let $D(\theta)$ be the $n \times n$ diagonal matrix $Diag(1, 1, ..., 1, e^{i\theta}, 1, ..., 1)$ where $e^{i\theta}$ is in position $m+1$. By the invariance of Haar measure, $D(\theta)\Delta$ has the same distribution as Δ, and so

$$\int_{U_n} H \sin(rxa_{m+1}(s\cos(\gamma+\theta)))\ \sin(ryb_{m+1}(s\sin(\gamma+\theta)))\ d\mu_n = I.$$

Thus

$$\int_0^{2\pi}\int_{U_n} H \sin(rxa_{m+1}(s\cos(\gamma+\theta)))\ \sin(ryb_{m+1}(s\sin(\gamma+\theta)))\ d\mu_n\ d\theta = 2\pi I.$$

By Fubini's Theorem [35], we have

$$\int_{U_n} H \int_0^{2\pi} \sin(rxa_{m+1}(s\cos(\gamma+\theta)))\ \sin(ryb_{m+1}(s\sin(\gamma+\theta)))\ d\theta\ d\mu_n = 2\pi I.$$

Next let $l(\theta) = \sin(rxa_{m+1}(s\cos\theta))\ \sin(ryb_{m+1}(s\sin\theta))$. Now, l is periodic with period 2π and shifting l by γ units yields a functions whose integral over $[0, 2\pi]$ coincides with the integral of l over that same interval. Thus

$$\int_{U_n} H \int_0^{2\pi} l(\theta)\ d\theta\ d\mu_n = 2\pi I.$$

However, l is an odd function and so

$$\int_0^{2\pi} l(\theta)\ d\theta = \int_{-\pi}^{\pi} l(\theta)\ d\theta = 0.$$

It follows that $I = 0$ and our claim is established.

Using this fact and arguing as we did in the proof of Theorem 2.1 , we have that the expression in (4.20) does not exceed the value

$$\begin{aligned}\int_{U_n} \Big| \prod_{j=m+1}^{n} \cos(rxa_jX_j) \prod_{j=m+1}^{n} \cos(ryb_jY_j) \\ -e^{-\frac{r^2x^2}{2}\sum_{j=m+1}^{n} a_j^2E(X_j^2)}\, e^{-\frac{r^2y^2}{2}\sum_{j=m+1}^{n} b_j^2E(Y_j^2)}\Big|\ d\mu_n \\ \le r^4x^4 \sum_{j=m+1}^{n} a_j^4E(X_j^4) + \frac{r^2x^2}{2}(\mathrm{Var}(\sum_{j=m+1}^{n} a_j^2X_j^2))^{\frac{1}{2}} \\ + r^4y^4 \sum_{j=m+1}^{n} b_j^4E(Y_j^4) + \frac{r^2y^2}{2}(\mathrm{Var}(\sum_{j=m+1}^{n} b_j^2Y_j^2))^{\frac{1}{2}}.\end{aligned}$$

We can bound this last expression as in the proof of Theorem 1, which leads us to a proper choice of L and completes the proof of Theorem 4.2. □

It is natural to ask if Theorem 3.1 has complex and symplectic analogues. We believe this is the case but thus far, like in the case of Theorem 2.1, we are able to prove a result of this type only for elements of the diagonals of these classes of matrices. In doing so, we obviously lean heavily on the preceding theorem.

Theorem 4.3. *Let $\Omega_n = U_n$ be the unitary group of $n \times n$ complex matrices, and let $\Delta = \Gamma + i\Lambda$ be an element of Ω_n distributed according to Haar measure. Let $d_j = \gamma_{jj} + i\lambda_{jj}$ and let $S_k^n = \sum_{j=1}^k d_j$. If*

$$Z_n(t, \omega) = S_{[nt]}(\omega), \ t \in [0, 1],$$

then $Z_n \Rightarrow \bar{W}$ converges to $\bar{W}$ where $\bar{W}$ is standard complex-valued Brownian motion ($\bar{W} = W_t^{(1)} + iW_t^{(2)}$ where $W^{(1)}$ and $W^{(2)}$ are independent one-dimensional Brownian motions with drift 0 and diffusion coefficient $\frac{1}{2}$).

Proof. We appeal to Theorem 5. One can easily adapt the argument for tightness given in Theorem 3.1 to show that ReZ_n is tight. Here $E(\gamma_{11}^2) = \frac{1}{2n}$ and $E(\gamma_{rr}\gamma_{ss}\lambda_{uu}\lambda_{vv}) = 0$ for distinct r, s, u, and v. Similarly, ImZ_n is tight and hence P_n is tight where P_n is the law of (ReZ_n, ImZ_n).

By Theorem 4.1, it remains to show that

$$P_n\pi^{-1}_{s_1,\cdots,s_k;t_1,\cdots,t_l} \Rightarrow P\pi^{-1}_{s_1,\cdots,s_k;t_1,\cdots,t_l} \tag{4.23}$$

where P is the law of $(W^{(1)}, W^{(2)})$. We consider time points s_1, s_2, t_1, and t_2 where $s_1 < s_2$ and $t_1 < t_2$, and one may easily verify that the general case can be handled analogously.

Letting $X_n = ReZ_n$ and $Y_n = ImZ_n$, we wish to prove that

$$(X_n(s_1), X_n(s_2), Y_n(t_1), Y_n(t_2)) \Rightarrow (W_{s_1}^{(1)}, W_{s_2}^{(1)}, W_{t_1}^{(2)}, W_{t_2}^{(2)}).$$

However, this statement would follow if

$$(X_n(s_1), X_n(s_2) - X_n(s_1), Y_n(t_1), Y_n(t_2) - Y_n(t_1))$$

converges in distribution to

$$(W_{s_1}^{(1)}, W_{s_2}^{(1)} - W_{s_1}^{(1)}, W_{t_1}^{(2)}, W_{t_2}^{(2)} - W_{t_1}^{(2)}).$$

Appealing as before to the Cramer-Wold device [5], it suffices to show that

$$aX_n(s_1) + b(X_n(s_2) - X_n(s_1)) + cY_n(t_1) + d(Y_n(t_2) - Y_n(t_1))$$

converges in distribution to

$$aW_{s_1}^{(1)} + b(W_{s_2}^{(1)} - W_{s_1}^{(1)}) + cW_{t_1}^{(2)} + d(W_{t_2}^{(2)} - W_{t_1}^{(2)})$$

for any $(a, b, c, d) \in \mathbb{R}^4$. The remainder of the proof follows by applying Theorem 4.2 in essentially the same way as Theorem 2.1 is applied in the proof of Theorem 3.1. $\square$

5 Symplectic matrices

Recall (see [8]) that the group of symplectic matrices $Sp(n)$ may be identified with the subgroup of $U(2n)$ of the form

$$\begin{bmatrix} A & -\bar{B} \\ B & \bar{A} \end{bmatrix} \in U(2n), \tag{5.24}$$

where A, B are complex $n \times n$ matrices. The trace of random matrices from this group is studied in [14, 16]. As shown there, if Θ is chosen according to Haar measure in $Sp(n)$, then $Tr(\Theta)$, $Tr(\Theta^2)$, ... , $Tr(\Theta^k)$ are asymptotically independent normal random variables. We now study the extent to which the diagonal entries of a random symplectic matrix generate Brownian motion.

Random matrices in $Sp(n)$ can be generated in the following way. Fill the real and imaginary entries of A and B in with real, standard normal i.i.d. random variables. Apply the Gram-Schmidt process to the n complex column vectors of dimension $2n$ which result. We now have a new A and B and we complete the right half of our matrix by following the pattern of (5.24). The matrix obtained in this way is distributed according to Haar measure in $Sp(n)$. To see this, one can adapt the argument given for the construction of a random orthogonal matrix. See for example Proposition 7.2 of [17]. We now have

Theorem 5.1. *Let $Sp(n)$ be the symplectic group of $2n \times 2n$ complex matrices of the form (5.24) , and let Θ be an element of $Sp(n)$ chosen according to Haar measure μ_n. Let $A = (a_{ij})_{i,j=1}^n$ be the upper left $n \times n$ block of Θ, and let $d_i = a_{ii}$, $1 \le i \le n$, and let $S_k^n = \sum_{i=1}^k d_i$. If*

$$Z_n(t, \omega) = S_{[nt]}(\omega), \quad t \in [0, 1]$$

then $Z_n \Rightarrow \frac{1}{\sqrt{2}}\bar{W}$ where $\bar{W}$ is standard complex-valued Brownian motion.

Proof. We are working with complex matrices and so we can follow the arguments of Theorems 4.2 and 4.3. We first need the symplectic analogue of Theorem 4.2. To accomplish this, only one change in the proof of Theorem 4.2 is required. In place of the diagonal matrix $D(\theta)$, we use instead the $2n \times 2n$ diagonal matrix $D_1(\theta) = Diag(1, ..., 1, e^{i\theta}, 1, ..., 1, e^{-i\theta}, 1, ..., 1)$ where $e^{i\theta}$ and $e^{-i\theta}$ occur in positions number $m + 1$ and $n + m + 1$ respectively. The rest of the arguments for the analogues of Theorems 4.2 and 4.3 are clear. □

It should be noted that we cannot link all $2n$ diagonal entries to obtain Brownian motion. If we were to try, note that $Z_n(\frac{1}{2})$ and $Z_n(1) - Z_n(\frac{1}{2})$ would tend to limits which are complex conjugates of one another and hence dependent.

Acknowledgement. The authors thank Harry Kesten for explaining how sign-symmetry could be used to show that the trace of a random orthogonal matrix converges to a standard normal distribution at the Bowdoin Conference on random matrices in 1985. They also thank Francis Comets for comments on earlier drafts of this paper. The first author thanks the Department of Statistics of Stanford University for warm hospitality extended to him during the summers of 1994-1997. He also thanks Jeff Rosenthal and Patrick Billingsley for some useful conversations. In addition, he acknowledges support from the Research Foundation of the State University of New York in the form of a PDQ Fellowship. The second and third authors acknowledge research support from the Division of Mathematical Sciences of the National Science Foundation.

Anthony D'Aristotile
Dept. of Mathematics
SUNY at Plattsburgh
Plattsburgh, NY 12901

Persi Diaconis
Depts. of Mathematics and Statistics
Stanford University
Stanford, CA 94305

Charles M. Newman
Courant Inst. of Math. Sciences
New York University
251 Mercer Street
New York, NY 10012

Bibliography

[1] Arratia, R., Goldstein, L., and Gordon, L., Poisson Approximation and the Chen-Stein Method, *Stat. Science* **5**, 403-434, 1990.

[2] Bai, Z.D., Methodologies in Special Analysis of Large Dimensional Random Matrices. A review, *Statist. Sinica* **9**, 611-677, 1994.

[3] Bhattacharya, R. and Waymire, E., *Stochastic Processes with Applications*, John Wiley and Sons, 1990.

[4] Billingsley, P. J., *Probability and Measure*, Second Edition, John Wiley and Sons, 1986.

[5] Billingsley, P. J., *Convergence of Probability Measures*, John Wiley and Sons, 1968.

[6] Borel, E., Sur les principes de la theorie cinétique des gaz, *Annales de l'ecole normale sup.* **23**, 9-32, 1906.

[7] Bump, D. and Diaconis, P., Toeplitz minors, *Jour. Combin. Th. A.* **97**, 252-271, 2001.

[8] Bröckner, T. and tom Dieck, J., *Representation of Compact Lie Groups*, Springer Verlag, 1985.

[9] Daffer, P., Patterson, R., and Taylor, R., *Limit Theorems for Sums of Exchangeable Random Variables*, Rowman and Allanhold, 1985.

[10] D'Aristotile, A., An Invariance Principle for Triangular Arrays, *Jour. Theoret. Probab.* **13**, 327-342, 2000.

[11] Diaconis, P., Application of the method of moments in probability and statistics. In H.J. Landau, ed., *Moments in Mathematics*, 125-142, Amer. Math. Soc., Providence, 1987.

[12] Diaconis, P., Patterns in eigenvalues, To appear *Bull. Amer. Math. Soc.*, 2002.

[13] Diaconis, P., Eaton, M., and Lauritzen, L., Finite de Finetti theorems in linear models and multivariate analysis, *Scand. J. Stat.* **19**, 289-315, 1992.

[14] Diaconis, P. and Evans, S., Linear functions of eigenvalues of random matrices, *Trans. Amer. Math. Soc.* **353**, 2615-2633, 2001.

[15] Diaconis, P. and Freedman, D., A dozen de Finetti-style results in search of a theory, *Ann. Inst. Henri Poincare* Sup au n. 2 **23**, 397-423, 1987.

[16] Diaconis, P. and Shahshahani, M., On the eigenvalues of random matrices, *J. Appl. Prob.* **31A**, 49-62, 1994.

[17] Eaton, M., *Multivariate Statistics*, John Wiley and Sons, 1983.

[18] Edelman, A., Kostlan, E., and Shub, M., How many eigenvalues of a random matrix are real?, *Jour. Amer. Math. Soc.* **7**, 247-267, 1999.

[19] Feller, W., *An Introduction to Probability Theory and Its Applications*, Vol. II, John Wiley and Sons, 1971.

[20] Golub, R. and Van Loan, C., *Matrix Computations*, 2nd Ed., Johns Hopkins Press, 1993.

[21] Hida, T., A role of Fourier transform in the theory of infinite dimensional unitary group, *J. Math. Kyoto Univ.* **13**, 203-212, 1973.

[22] Jiang, T.F., Maxima of entries of Haar distributed matrices, Technical Report, Dept. of Statistics, Univ. of Minnesota., 2002.

[23] Kuo, H.H., *White Noise Distribution Theory*, CRC Press, Boca Raton, 1996.

[24] Levy, P., *Leçons d'Analyse Fonctionnelle*, Gauthiers-Villars, Paris, 1922.

[25] Levy, P., *Analyse Fonctionnelle*, Memorial des Sciences Mathematiques, Vol. 5, Gauthier-Villars, Paris, 1925.

[26] Levy, P., *Problemes Concrets d'Analyse Fonctionnelle*, Gauthier-Villars, Paris, 1931.

[27] Mallows, C., A Note on asymptotic joint normality, *Ann. Math. Statist.* **43**, 508-515, 1972.

[28] Maxwell, J.C., *Theory of Heat*, 4th ed., Longmans, London, 1875.

[29] Maxwell, J.C., On Boltzmann's theorem on the average distribution of energy in a system of material points, *Cambridge. Phil. Soc. Trans.* **12**, 547-575, 1878.

[30] McKean, H.P., Geometry of Differential Space, *Ann. Prob.* **1**, 197-206, 1973.

[31] Mehler, F.G., Ueber die Entwicklung einer Function von beliebig vielen. Variablen nach Laplaschen Functionen höherer Ordnung, *Crelle's Journal* **66**, 161-176, 1866.

[32] Mehta, M., *Random Matrices*, Academic Press, 1991.

[33] Olshanski, G., Unitary representations of infinite-dimensional pairs (G, K) and the formalism of R. Howe. In A.M. Vershik and D.P. Zhelobenko, eds. *Representation of Lie Grops and Related Topics*, Adv. Studies in Contemp. Math. **7**, 269-463, Gordon and Breach, New York, 1990.

[34] Rains, E., Normal limit theorems for asymmetric random matrices, *Probab. Th. Related Fields* **112**, 411-423, 1998.

[35] Royden, H. L., *Real Analysis*, 2nd Edition, The Macmillan Company, 1968.

[36] Stein, C., The accuracy of the normal approximation to the distribution of the traces of powers of random orthogonal matrices, Technical Report No. 470, Stanford University, 1995.