Weak Convergence of a Collection of Random Functions Defined by the Eigenvectors of Large Dimensional Random Matrices

by

Jack W. Silverstein
Department of Mathematics Box 8205
North Carolina State University
Raleigh, NC 27605-8205, USA

Mathematics Subject Classification 2020. Primary 60F05, 15A18; Secondary 62H99.

Key Words and Phrases. Weak convergence on $D[0,1]$ Haar measure, eigenvectors of random matrices, Brownian bridge, Haar measure.

Abstract

For each n, let U_n be Haar distributed on the group of $n \times n$ unitary matrices. Let $\mathbf{x}_{n,1}, \ldots, \mathbf{x}_{n,m}$ denote orthogonal nonrandom unit vectors in $\mathbb{C}^n$ and let $\mathbf{u}_{n,k} = (u_k^1, \ldots, u_k^n)^* = U_n^* \mathbf{x}_{n,k}$, $k = 1, \ldots, m$. Define the following functions on [0,1]: $X_n^{k,k}(t) = \sqrt{n} \sum_{i=1}^{[nt]} (|u_k^i|^2 - \frac{1}{n})$, $X_n^{k,k'}(t) = \sqrt{2n} \sum_{i=1}^{[nt]} \bar{u}_k^i u_{k'}^i$, $k < k'$. Then it is proven that $X_n^{k,k}$, $\Re X_n^{k,k'}$, $\Im X_n^{k,k'}$, considered as random processes in $D[0,1]$, converge weakly, as $n \to \infty$, to m^2 independent copies of Brownian bridge.

The same result holds for the $m(m+1)/2$ processes in the real case, where O_n is real orthogonal Haar distributed and $\mathbf{x}_{n,i} \in \mathbb{R}^n$, with $\sqrt{n}$ in $X_n^{k,k}$ and $\sqrt{2n}$ in $X_n^{k,k'}$ replaced with $\sqrt{\frac{n}{2}}$ and $\sqrt{n}$, respectively. This latter result will be shown to hold for the matrix of eigenvectors of $M_n = (1/s) V_n V_n^T$ where V_n is $n \times s$ consisting of the entries of $\{v_{ij}\}$, $i, j = 1, 2, \ldots$, i.i.d. standardized and symmetrically distributed, with each $\mathbf{x}_{n,i} = \{\pm 1/\sqrt{n}, \ldots, \pm 1/\sqrt{n}\}$, and $n/s \to y > 0$ as $n \to \infty$. This result extends the result in J.W. Silverstein *Ann. Probab.* **18** 1174-1194.

These results are applied to the detection problem in sampling random vectors mostly made of noise and detecting whether the sample includes a nonrandom vector. The matrix $B_n = \theta \mathbf{v}_n \mathbf{v}_n^* + S_n$ is studied where S_n is Hermitian or symmetric and nonnegative definite with either its matrix of eigenvectors being Haar distributed, or $S_n = M_n$, $\theta > 0$ nonrandom, and $\mathbf{v}_n$ is a nonrandom unit vector. Results are derived on the distributional behavior of the inner product of vectors orthogonal to $\mathbf{v}_n$ with the eigenvector associated with the largest eigenvalue of B_n.

1. Introduction Let $\{v_{ij}\}$, $i,j=1,2,\ldots$ be i.i.d. real valued standardized random variables with finite fourth moment, and for each n let $M_n=\frac{1}{s}V_nV_n^T$, where $V_n=(v_{ij})$, $i=1,2,\ldots,n$, $j=1,2,\ldots,s=s(n)$, and $n/s\to y>0$ as $n\to\infty$. This paper is essentally an extension of results in [16], where it is shown that random elements in $D[0,1]$, the space of r.c.l.l. function on $[0,1]$ embodied with the Skorohod metric, defined by the eigenvectors of M_n converge weakly to Brownian bridge under the assumption $v_{i\,j}$ is symmetrically distributed. Specifically, denote by $O_n\Lambda_nO_n^T$ the spectral decomposition of M_n, where the eigenvalues of M_n are arranged along the diagonal of Λ_n in nondecreasing order, and the columns of the orthogonal matrix O_n, are the corresponding eigenvectors (a unique determination of O_n is outlined in Section 2 of [16]). For each n let $\mathbf{x}_n\in\mathbb{R}^n$ be a nonrandom unit vector, and let $\mathbf{y}_n=(y_1,y_2,\ldots,y_n)^T=O_n^T\mathbf{x}_n$. Define for $t\in[0,1]$

$$X_n(t)\equiv\sqrt{\tfrac{n}{2}}\sum_{i=1}^{[nt]}(y_i^2-\tfrac{1}{n})\qquad([a]\equiv\text{greatest integer}\le a). \tag{1.1}$$

The main result in [16] is that when $v_{i\,j}$ is symmetrically distributed, for $\mathbf{x}_n=(\pm\frac{1}{\sqrt{n}},\pm\frac{1}{\sqrt{n}},\ldots,\pm\frac{1}{\sqrt{n}})^T$.

$$X_n\to_D W^\circ\qquad\text{as } n\to\infty \tag{1.2}$$

(D denoting weak convergence in $D[0,1]$) where W° is Brownian bridge ([3], p. 64).

This result is a partial answer to the question of how the matrix of eigenvectors of M_n are related to the Haar measure on the group $\mathcal{O}_n$ of $n\times n$ orthogonal matrices, which occurs when v_{11} is mean 0 Gaussian, That is, when M_n is a matrix of Wishart type. The question is originally raised in [13] where it is conjectured that for arbitrary centered v_{11} the distribution of O_n in $\mathcal{O}_n$ is near in some way to the Haar measure ([13][14],[15],[16], see also [12]). This resulted in [13] to an investigation in the behavior of (1.1). When O_n is Haar distributed $\mathbf{y}$ is uniformly distributed over the unit sphere in $\mathbb{R}^n$, being the same as the normalized vector, $(\zeta_1,\ldots,\zeta_n)^T$, of i.i.d mean-zero Gaussian entries. (1.1) can then be written as

$$X_n(t)=\frac{\sqrt{n}}{\sqrt{2}}\left(\frac{\sum_{i=1}^{[nt]}\zeta_i^2}{\sum_{i=1}^{n}\zeta_i^2}-\frac{[nt]}{n}\right)=\frac{n}{\sum_{i=1}^{n}\zeta_i^2}\frac{1}{\sqrt{2}}\frac{1}{\sqrt{n}}\left(\sum_{i=1}^{[nt]}(\zeta_i^2-1)-\frac{[nt]}{n}\sum_{i=1}^{n}(\zeta_i^2-1)\right). \tag{1.3}$$

Using the fact that the fourth moment of a standard normal random variable is 3, we apply Donsker's theorem ([3], Theorem 16.1) along with standard results on weak convergence of random functions on $D[0,1]$ to arrive at (1.2).

In [14] and [15] it is shown that a necessary condition for (1.2) to hold for all unit vectors $\mathbf{x}_n$ is that when $\mathsf{E}(v_{1\,1}^2)=1$ we must have $\mathsf{E}(v_{1\,1}^4)=3$. Indeed, it is shown in [15] that when $\mathsf{E}(v_{11}^2) = 1$ but $\mathsf{E}(v_{11}^4) \neq 3$, there exist sequences $\{\mathbf{x}_n\}$ of unit vectors such that $\{X_n\}$ fails to converge weakly. This result suggests a strong relationship needs to exist between the distribution of $v_{1\,1}$ and Gaussian in order for (1.2) to hold for all sequences of unit vectors, and leaves open the possibility that this is true only when v_{11} is Gaussian.

However, the result in [16] indicates some similarity of the distribution of O_n to Haar measure, at least when v_{ij} is symmetrically distributed and the entries of $\mathbf{x}_n$ are equally weighted.

In this paper another property of the Haar measure on $\mathcal{O}_n$ is derived and is shown to be true for v_{11} symmetrically distributed and on unit vectors considered in [16]. In order to provide a more complete setting, the property is stated and derived on $\mathcal{U}_n$, the group of $n \times n$ unitary matrices. The corresponding statements and steps in the verification for the real case will be specified in the proof.

Let for $d \geq 2$ an integer, and $b \geq 1$, $D_d^b = \Pi_{i=1}^d D[0,b]$, and $\mathcal{T}_d^b$ denote the smallest σ-field on D_d^b in which convergence of elements in D_d^b is equivalent to component-wise convergence. We will prove the following:

Theorem 1.1 For each n, let U_n be Haar distributed on $\mathcal{U}_n$. Let $\mathbf{x}_{n,1}, \ldots, \mathbf{x}_{n,m}$ denote orthogonal nonrandom unit vectors in $\mathbb{C}^n$ and let $\mathbf{u}_{n,k} = (u_k^1, \ldots, u_k^n)^* = U^* \mathbf{x}_{n,k}$, $k = 1, \ldots, m$. Define the following functions on [0,1]:

$$(1.4) \qquad X_n^{k,k}(t) = \sqrt{n} \sum_{i=1}^{[nt]} (|u_k^i|^2 - \tfrac{1}{n}), \qquad X_n^{k,k'}(t) = \sqrt{2n} \sum_{i=1}^{[nt]} \bar{u}_k^i u_{k'}^i \quad k < k'$$

("$^-$" denoting complex conjugate). Then $X_n^{k,k}, \Re X_n^{k,k'}, \Im X_n^{k,k'}$ $k < k'$, considered as random processes in $D[0,1]$, converge weakly in $D_{m^2}^1$ to independent copies of Brownian bridge.

The fact that $X_n^{k,k}$ converges weakly to W° follows along the same lines as in (1.2) where now we use the fact that a vector uniformly distributed on the unit sphere in $\mathbb{C}^n$ can be achieved by normalizing an i.i.d. vector, $(z_1, \ldots, z_n)^T$, where each z_i is standard complex normal (real and imaginary parts i.i.d. $N(0,1/2)$), and subsequently $\mathsf{E}|z_1|^2 = 1$, $\mathsf{E}|z_1|^4 = 2$. The reason why $\Re X_n^{k,k'}, \Im X_n^{k,k'}$ $k < k'$ converge weakly to W° will be seen in the proof. It follows from how the proof is approached, by creating the $\mathbf{u}_{n,k}$ after applying the Gram-Schmidt orthogonalization process on a matrix of i.i.d. standard complex Gaussians, resulting in a Haar distributed unitary matrix.

The real case is stated in the following

Theorem 1.2 For each n, let O_n be Haar distributed on $\mathcal{O}_n$. Let $\mathbf{x}_{n,1}, \ldots, \mathbf{x}_{n,m}$ denote orthogonal nonrandom unit vectors in $\mathbb{R}^n$ and let $\mathbf{y}_k = (y_{k,1}, \ldots, y_{k,n})^T = O_n^T \mathbf{x}_{n,k}$,

$k = 1, \ldots, m$. For each of these k define X_n^k, a random element in $D[0,1]$ to be (1.1) with y_i replaced with $y_{k,i}$ For $1 \le j < k \le m$ define Y_n^{jk}, a random element of $D[0,1]$, to be

$$Y_n^{jk}(t) = \sqrt{n} \sum_{i=1}^{[nt]} y_{j,i} y_{k,i}, \tag{1.5}$$

Then the random functions X_n^k, Y_n^{jk}, $1 \le j < k \le m$ converge weakly in D_d^1, $d = m(m+1)/2$, to independent Brownian Bridges.

The extension of the result in [16] is the following:

THEOREM 1.3 Assume v_{11} is symmetrically distributed about 0, $\mathsf{E}v_{11}^4 < \infty$, and the m orthogonal vectors $\mathbf{x}_{n,k} = (\pm 1/\sqrt{n}, \ldots, \pm 1/\sqrt{n})^T$ (this of course necessitates the n's to be restiricted to multiples of 2^m). Then, with O_n being the orthogonal matrix of eigenvectors of $M_n = \frac{1}{s} V_n V_n^T$, the conclusion of Theorem 1.2 holds.

The motivation behind studying these quantities is to analyze the detection problem in sampling random vectors mostly made of noise, and determining whether the sample includes multiples of a nonrandom vector. For example, reading off the values a bank of antennas is receiving at discrete intervals of time. If the values consist of pure Gaussian noise, then the matrix forming the sample correlation matrix S_n is modeled by a Wishart matrix, and its matrix of eigenvectors would be Haar distributed, either in $\mathcal{O}_n$ or $\mathcal{U}_n$. Suppose at certain periods of time multiples of a nonrandom unit vector $\mathbf{v}_n$ appear, resulting in the matrix

$$B_n = \theta \mathbf{v}_n \mathbf{v}_n^* + S_n \quad \theta > 0 \quad \text{nonrandom.} \tag{1.6}$$

It is straightforward to verity that λ_n^1, the largest eigenvalue of B_n, is the unique value which solves

$$\mathbf{v}_n^* (\lambda I - S_n)^{-1} \mathbf{v}_n = 1/\theta \quad \text{for } \lambda > \lambda_{\max}(S_n) \tag{1.7}$$

where I is the $n \times n$ identity matrix and $\lambda_{\max}(S_n)$ is the largest eigenvalue of S_n. Moreover, a multiple of the corresponding eigenvector is

$$(\lambda_n^1 I - S_n)^{-1} \mathbf{v}_n. \tag{1.8}$$

The goal is to understand the random behavior of this largest eigenvector for n large in order to infer as much as possible the nature of $\mathbf{v}_n$. We will place S_n in a more general setting.

Let, for each n, S_n be a Hermitian nonnegative definite random matrix whose matrix of eigenvectors is Haar distributed in $\mathcal{U}_n$. Let F_n denote the empirical distribution function of the eigenvalues of S_n, that is, for $x \ge 0$, $F_n(x) = \frac{1}{n}$(number of eigenvalues of S_n $\le x$).

Suppose with probability one F_n converges in distribution to F, a nonrandom probability distribution function, continuous on $[0,\infty)$, where the largest eigenvalue of S_n converges almost surely to $\lambda_{\max} > 0$.

We will prove the following:

THEOREM 1.4 Suppose for all $\lambda > \lambda_{\max}$, $\int(\lambda - x)^{-1}dF(x) \ \leq 1/\theta$ (integral being over $[0, \lambda_{\max}]$). Then with probability one $\lambda_n^1 \to \lambda_{\max}$ as $n \to \infty$ and knowledge of the limiting behavior of (1.8) is beyond the scope of this paper.

However, if there exists $\lambda > \lambda_{\max}$ such that $\int(\lambda - x)^{-1}dF(x) > 1/\theta$, then, since $\int(\lambda - x)^{-1}dF(x)$ decreases to zero, there exists a unique $\lambda_1 > \lambda_{\max}$ such that

$$\int(\lambda_1 - x)^{-1}dF(x) = 1/\theta, \tag{1.9}$$

and $\lambda_n^1 \xrightarrow{a.s.} \lambda_1$.

For any $\mathbf{x}_{n,1}, \ldots, \mathbf{x}_{n,m-1}$ unit vectors orthogonal to $\mathbf{v}_n$

$$\begin{aligned}\sqrt{2n}\mathbf{x}_{n,k}^*(\lambda_n^1 I - S_n)^{-1}\mathbf{v}_n & \\ \to_D \int(\lambda_1 - x)^{-1}dW_{k,r}^0(F(x)) &+ i\int(\lambda_1 - x)^{-1}dW_{k,i}^0(F(x)),\end{aligned} \tag{1.10}$$

where $W_{k,r}^0$, $W_{k,i}^0$, $k \leq m-1$, are independent copies of Brownian bridge, and I_A is the indicator function on the set A . Thus the limits are iid mean zero Gaussians, and it is straightforward to show their common variance is

$$\int(\lambda_1 - x)^{-2}dF(x) - \left(\int(\lambda_1 - x)^{-1}dF(x)\right)^2. \tag{1.11}$$

Moreover, the norm of the eigenvector (1.8) satisfies

$$\|(\lambda_n^1 I - S_n)^{-1}\mathbf{v}_n\| \xrightarrow{a.s.} \left(\int(\lambda_1 I - x)^{-2}dF(x)\right)^{1/2}. \tag{1.12}$$

With Theorems 1.2 and 1.3 come the analogous results in the real case, with (1.10) becoming

$$\sqrt{n}\mathbf{x}_{n,k}^*(\lambda_n^1 I - S_n)^{-1}\mathbf{v}_n \to_D \int(\lambda_1 - x)^{-1}dW_k^0(F(x)). \tag{1.13}$$

For the matrix $S_n = M_n$ in Theorem 1.3 the vectors $\mathbf{v}_n$ and $\mathbf{x}_{n,i}$ are all orthonormal vectors of the form $(\pm 1/\sqrt{n}, \ldots, \pm 1/\sqrt{n})^T$. There is a limiting F in this case, described below.

These results can aid in detecting the presence of a particular signal by establishing the distributional behavior of inner products of the eigenvector of B_n associated with the largest eigenvalue with vectors orthogonal to $\mathbf{v}_n$. Knowledge of eigenvalue behavior of S_n can aid in the detection. For example, if $S_n = M_n$ where the v_{ij} are $N(0,1)$, F_y is known to be the Marčenko-Pastur distribution ([10], [7], [18], [8], [19], [17]), proven in [19] under the assumption of finite second moment of v_{11}, where, with $a = (1-\sqrt{y})^2$ $b = (1+\sqrt{y})^2$, for $y \le 1$, F_y has density

$$f_y(x) = \begin{cases} \frac{\sqrt{(x-a)((b-x)}}{2\pi y x} & a < x < b \\ 0 & \text{otherwise,} \end{cases}$$

and, for $y > 1$, F has mass $1 - 1/y$ at 0, and density $f_y(x)$ on $((1-\sqrt{y})^2, (1+\sqrt{y})^2)$.

These results have connections to the spike model ([1], [2], [11])where a sample covariance matrix is studied with several of its population eigenvalues being altered, not enough of them to change the limiting empirical spectral distribution, but enough of a change in values to reveal individual sample eigenvalues associated with them. For B_n the size of θ in relation to the function $f(\lambda) = \int (\lambda - x)^{-1} dF(x)$ on $(\lambda_{\max}, \infty)$ determines whether a spike sample eigenvalue is revealed.

The next sections contain proofs of these results. Section 2 contains the proofs of Theorems 1.1 and 1.2, Section 3 has the proof of Theorem 1.3, and Section 4 has the proof of Theorem 1.4

2. Proofs of Theorem 1.1 and 1.2. We concentrate on the proof of Theorem 1.1 and indicate the analogous results in the real case.

We begin with understanding the relationship between $u_{n,k}$ and $u_{n,k'}$ $k \ne k'$. Let U be any unitary matrix having $\mathbf{x}_{n,k}$ $k \le m$ for its first m columns. We know that the matrix $U_n^* U$ is also Haar distributed, so we see that $\mathbf{u}_{n,k}$ $k \le m$, have the same distribution as the first m columns of a Haar distributed matrix. The following lemma will enable us to express their relationship in a simple way.

Lemma 2.1. Let $Z = (z_{ij})$ be $n \times n$ consisting of i.i.d. complex Gaussian entries ($z_{11} = z_r + i z_i$ z_r, z_i independent $N(0, 1/2)$). Form the $n \times n$ unitary matrix U by performing the Gram-Schmidt process on the columns of Z. Then U is Haar distributed in $\mathcal{U}_n$, the group of $n \times n$ unitary matrices.

Proof: Let $\mathbf{z}_k, \mathbf{u}_k$ be the $k-th$ column of Z, U, respectively. Then

$$\mathbf{u}_1 = f_1(\mathbf{z}_1) \equiv \left(\frac{1}{\|\mathbf{z}_1\|}\right) \mathbf{z}_1,$$

and recursively

$$\mathbf{u}_k = f_k(\mathbf{z}_1, \dots, \mathbf{z}_k) \equiv \frac{1}{\|\mathbf{z}_k - \mathbf{p}_k\|} (\mathbf{z}_k - \mathbf{p}_k),$$

where

$$\mathbf{p}_k \equiv (\mathbf{u}_1^*\mathbf{z}_k)\mathbf{u}_1 + \cdots + (\mathbf{u}_{k-1}^*\mathbf{z}_k)\mathbf{u}_{k-1}.$$

Let $Q \in \mathcal{U}_n$. We will show for $k = 1, \ldots, n$

$$Q\mathbf{u}_k = Qf_k(\mathbf{z}_1, \ldots, \mathbf{z}_k) = f_k(Q\mathbf{z}_1, \ldots, Q\mathbf{z}_k). \tag{2.1}$$

We use induction. $k = 1$ is obvious. Assume it is true for $\ell = 1, 2, \ldots, k-1$. Then

$$Q\mathbf{u}_k = \frac{1}{\|Q\mathbf{z}_k - Q\mathbf{p}_k\|}(Q\mathbf{z}_k - Q\mathbf{p}_k),$$

and

$$Q\mathbf{p}_k = ((Qf_1(\mathbf{z}_1))^*Q\mathbf{z}_k)Qf_1(\mathbf{z}_1) + \cdots + ((Qf_{k-1})^*Q\mathbf{z}_k)Qf_{k-1}(\mathbf{z}_1, \ldots, \mathbf{z}_{k-1})$$

$$= ((f_1(Q\mathbf{z}_1))^*Q\mathbf{z}_k)f_1(Q\mathbf{z}_1) + \cdots + ((f_{k-1}(Q\mathbf{z}_1, \ldots, Q\mathbf{z}_{k-1}))^*Q\mathbf{z}_k)f_{k-1}(Q\mathbf{z}_1, \ldots, Q\mathbf{z}_{k-1}),$$

by the inductive hypothesis. Therefore we get (2.1).

We use now the fact that $QZ \sim Z$ to conclude

$$QU = (Qf_1(\mathbf{z}_1), Qf_2(\mathbf{z}_1, \mathbf{z}_2), \ldots, Qf_n(\mathbf{z}_1, \ldots, \mathbf{z}_n))$$

$$= (f_1(Q\mathbf{z}_1), f_2(Q\mathbf{z}_1, Q\mathbf{z}_2), \ldots, f_n(Q\mathbf{z}_1, \ldots, Q\mathbf{z}_n))$$

$$\sim (f_1(\mathbf{z}_1), f_2(\mathbf{z}_1, \mathbf{z}_2), \ldots, f_n(\mathbf{z}_1, \ldots, \mathbf{z}_n)) = U,$$

and we are done.

We will use Lemma 2.1 after we establish the framework for considering the m^2 processes on a common probability space.

We assume the reader is familiar with the basic concepts of probability, including: the notion of a measure space $\{\Omega, \mathcal{F}\}$, where $\mathcal{F}$ is a σ-field of subsets of Ω, and a probabilty space $\{\Omega, \mathcal{F}, \mathsf{P}\}$, where P is a probability measure defined on $\mathcal{F}$. Given two measurable spaces $\{\Omega_1, \mathcal{F}_1\}$, $\{\Omega_2, \mathcal{F}_2\}$, a mapping $T : \Omega_1 \to \Omega_2$, is *measurable* $\mathcal{F}_1/\mathcal{F}_2$ if $T^{-1}A_2 = \{\omega \in \Omega_1 : T\omega \in A_2\} \in \mathcal{F}_1$ for each $A_2 \in \mathcal{F}_2$. For any collection $\mathcal{A}$ of subsets of a set Ω, $\sigma(\mathcal{A})$ denotes the smallest σ-field containing $\mathcal{A}$.

We also assume the reader is also familiar with the material in [3],[5] on weak convergence of probability measures on metric spaces, most notably the metric space $D = D[0,1]$ consisting of real valued functions on $[0,1]$ that are right continuous with left-hand limits, the σ-field $\mathcal{D}$, defined by the Skorohod topology on D. For $0 \le t_1 < \cdots < t_k \le 1$, let $\pi_{t_1 \cdots t_k}$ denote the natural projection from D to $\mathbb{R}^k$:

$$\pi_{t_1 \cdots t_k}(x) = (x(t_1), \ldots, x(t_k)),$$

for any $x \in D$. Let $\mathcal{D}_f$ denote the collection, $\pi^{-1}_{t_1 \cdots t_k} H$, for any k, $0 \leq t_1 < \cdots < t_k \leq 1$, and $H \in \mathcal{R}^k$, the σ-field of Borel sets in $\mathbb{R}^k$, called the class of finite-dimensional sets. In [5] it is shown that $\mathcal{D}_f$ is a π-system (closed under intersections) and $\sigma(\mathcal{D}_f) = \mathcal{D}$. Therefore (Theorem 3.3 of [4] $\mathcal{D}_f$ is a separating class for probability measures on $(D, \mathcal{D})$: if probability measures P_1, P_2 agree on $\mathcal{D}_f$ then they are identical. Thus, showing weak convergence of a sequence, $\{P_n\}$, of probability measures on $(D, \mathcal{D})$ to a probability measure P (denoted by $\mathsf{P}_n \Rightarrow \mathsf{P}$) amounts to verifying $\{\mathsf{P}_n\}$ is tight (that is, for any $\epsilon > 0$ there exists a compact set $A_\epsilon \in \mathcal{D}$ such that $\mathsf{P}_n(A_\epsilon) > 1 - \epsilon$ for all n), and $\mathsf{P}_n(A) \to \mathsf{P}(A)$ for all $A \in \mathcal{D}_f$.

We wish to extend this criterion of weak convergence to the product space $D_d = \Pi_{i=1}^d D$ with the product topology $\mathcal{T}_d$, the smallest σ-field in which convergence of elements in D_d is equivalent to component-wise convergence. Since $(D, \mathcal{D})$ is separable, it follows from natural extensions to the material in M10 of [5], $(D_d, \mathcal{T}_d)$ is separable, which implies

$$\mathcal{T}_d = \sigma(\{\Pi_{i=1}^d A_i : \text{ each } A_i \in \mathcal{D}\}). \tag{2.2}$$

Let $B = \{\Pi_{i=1}^d A_i : \text{ each } A_i \in \mathcal{D}_f\})$. It is clear that B is also a π-system. We also have

Lemma 2.2. $\sigma(B) = \mathcal{T}_d$.

Proof: We have $\sigma(B) \subset \mathcal{T}_d$. Let $T_1(x_1, \ldots, x_d) = x_1$, and define

$$C = \{A \in \mathcal{D} : T_1^{-1} A \in \sigma(B)\}.$$

We have obviously $D \in C$, and $A \in C$ for each $A \in \mathcal{D}_f$, since $T_1^{-1} A = A \otimes \Pi_{i=1}^{d-1} D \in \sigma(B)$. For $A \in C$, $T_1^{-1} A^c = (T_1^{-1} A)^c \in \sigma(B)$, which implies $A^c \in C$. For $\{A_n\} \subset C$, $T_1^{-1} \cup A_n = \cup T_1^{-1} A_n \in \sigma(B)$, implying $\cup A_n \in C$. Therefore, C is a σ-field containing $\mathcal{D}_f$, and hence contains $\mathcal{D} = \sigma(\mathcal{D}_f)$. Therefore, $C = \mathcal{D}$, and we have for any $A \in \mathcal{D}$ $A \otimes \Pi_{i=1}^{d-1} \in \sigma(B)$. Similarly, we have for $2 \leq j < d$ $(\Pi_{i=1}^{j-1} D) \otimes A \otimes (\Pi_{i=1}^{d-j} D)$ and $(\Pi_{i=1}^{d-1} D) \otimes A$ all contained in $\sigma(B)$, so it also contains all $\Pi_{i=1}^d A_i$ for each $A_i \in \mathcal{D}$. Therefore by (2.2) we have $\mathcal{T}_d \subset \sigma(B)$, and we have our result.

We see then that from Theorem 3.3 of [4] B is a separating class for probability measures on $(D_d, \mathcal{T}_d)$.

It is straightforward to verify that

$$B = \{\Pi_{i=1}^d A_i : A_i = \pi^{-1}_{t_1, \ldots, t_k} H_i, k = 1, 2, \ldots, 0 \leq t_1 < \cdots < t_k \leq 1, H_i \in \mathcal{R}^k\}. \tag{2.3}$$

Suppose now we have a probability space $(\Omega, \mathcal{F}, \mathsf{P})$ and a mapping X from Ω into D_d, for which each component x_i is a random element in D, that is, it is measurable $\mathcal{F}/\mathcal{D}$. Then for any $A_i \in \mathcal{D}$, $i = 1, \ldots, d$, we have

$$X^{-1}(\Pi_{i=1}^d A_i) = \bigcap_{i=1}^d \{\omega : x_i(\omega) \in A_i\} \in \mathcal{F}.$$

Therefore, from (2.2) and Theorem 13.1 of [4] we have that X is measurable $\mathcal{F}/\mathcal{D}_d$, that is, $(x_1, \dots, x_d)$ is a random element in D_d.

If $x_1, x_2, \dots, x$ are random elements from probability space $(\Omega, \mathcal{F}, \mathsf{P})$ to D (D_d), we write $x_n \Rightarrow x$ to mean the measures x_n induce on D (D_d) converge weakly to the measure on D (D_d) induced by x. Also we say $\{x_n\}$ is tight (on D or D_d) if the sequence of induced measures is tight.

We then have the following:

Lemma 2.3. Suppose $\{x_n^1, \dots, x_n^d\}$ is a sequence of random functions, each lying in D, defined on a common probability space $(\Omega, \mathcal{F}, \mathsf{P})$. Then, from above, for each n $\{x_n^1, \dots, x_n^d\}$ is a random element in D_d. Assume each $\{x_n^i\}$ is tight. Moreover, assume there exists a random element $(x^1, \dots, x^d)$ in D_d for which

$$(x_n^1(t_1), \dots, x_n^1(t_k), \dots, x_n^d(t_1), \dots, x_n^d(t_k)) \Rightarrow (x^1(t_1), \dots, x^1(t_k), \dots, x^d(t_1), \dots, x^d(t_k))$$

(weak convergence on $\mathbb{R}^{dk}$) for all k, $t_1, \dots, t_k$. Then $(x_n^1, \dots, x_n^d) \Rightarrow (x^1, \dots, x^d)$.

Proof. Let P_n^i, P^i denote the measures the x_n^i, x^i induce on D, and $\mathsf{P}_{n,d}$ the measure $\{x_n^1, \dots, x_n^d\}$ induces on D_d. Then each $\{P_n^i\}$ is tight. Therefore for any $\epsilon > 0$ there exists compact sets $A_\epsilon^i \in \mathcal{D}$ for which $\mathsf{P}_n^i(A_\epsilon^i) > 1 - \epsilon/d$. Then ([5], M6) we have $\Pi_{i=1}^d A_\epsilon^i$ compact, and

$$\mathsf{P}_n(\Pi_{i=1}^d A_\epsilon^i) = \mathsf{P}(\{\omega : x_n^i(\omega) \in A_\epsilon^i, i \le d\}) = \mathsf{P}(\cap\{\omega : x_n^i(\omega) \in A_\epsilon^i\})$$
$$1 - \mathsf{P}(\cup\{\omega : x_n^i \in A_\epsilon^i\}^c) \ge 1 - \sum \mathsf{P}_n^i({A_\epsilon^i}^c) \ge 1 - \epsilon.$$

Therefore $\{\mathsf{P}_{n,d}\}$ is tight. Since B is a separating class, and it can be expresssed as in (3), we must have $\{x_n^1, \dots, x_n^d\} \Rightarrow \{x^1, \dots, x^d\}$.

We proceed to show each of $X_n^{k,k}, \Re X_n^{k,k'}, \Im X_n^{k,k'}$ $k < k'$ converges weakly to independent copies of Brownian bridge.

The following lemma is needed throughout the remaining arguments.

Lemma 2.4. If random variables X_n, Y_n are such that $\{Y_n\}$ is tight and $X_n \xrightarrow{i.p.} 0$, then $X_n Y_n \xrightarrow{i.p.} 0$.

Proof: For $\epsilon > 0$ $M > 0$ we have

$$\mathsf{P}(|X_n|\,|Y_n| > \epsilon) = \mathsf{P}(|X_n|\,|Y_n| > \epsilon, |Y_n| > M) + \mathsf{P}(|X_n|\,|Y_n| > \epsilon, |Y_n| \le M)$$

$$\le \mathsf{P}(|Y_n| > M) + \mathsf{P}(|X_n| > \epsilon/M).$$

Therefore $\limsup_n \mathsf{P}(|X_n|\,|Y_n| > \epsilon) \le \limsup_n \mathsf{P}(|Y_n| > M)$ which can be made arbitrarily small. We get our result.

Let Z and U be as in Lemma 2.1. We can assume the first m columns of U are the orthonormal vectors $\mathbf{u}_{n,k}$ where in the following we suppress he dependence on n. We can also assume that Z and U are $n \times m$. Define $r_{jk} = \mathbf{u}_j^* \mathbf{z}_k$ for $j < k$, $r_{11} = \|\mathbf{z}_1\|$, and for $k \geq 2$, $r_{kk} = \|\mathbf{z}_k - \mathbf{p}_k\|$. We have then $r_{11}\mathbf{u}_1 = \mathbf{z}_1$, and for $k \geq 2$

$$r_{kk}\mathbf{u}_k = \mathbf{z}_k - \sum_{j=1}^{k-1} r_{jk}\mathbf{u}_j.$$

Letting R denote the $m \times m$ upper triangular matrix (r_{jk}) we obtain the QR factorization of Z: $Z = UR$. Letting $A = R^{-1}$ we have $U = ZA$. We have then for each k

$$\mathbf{u}_k = a_{kk}\mathbf{z}_k + \sum_{j=1}^{k-1} a_{jk}\mathbf{z}_j. \tag{2.4}$$

For $j < k$, $\mathbf{u}_j$ and $\mathbf{z}_k$ are independent. Therefore

$$E(r_{jk}) = 0 \quad \text{and} \quad E|r_{jk}|^2 = 1. \tag{2.5}$$

Therefore above the diagonal the entries of R are tight. By the weak law of large numbers

$$\|\mathbf{z}_k\|/\sqrt{n} \xrightarrow{i.p.} 1. \tag{2.6}$$

It is straightforward to verify

$$r_{kk}^2 = \|x_k\|^2 - \sum_{j=1}^{k-1} |r_{jk}|^2. \tag{2.7}$$

Therefore we have

$$\frac{r_{kk}^2}{\|\mathbf{z}_k\|^2} = 1 + O(1)/n, \tag{2.8}$$

where here and in the following $O(1)$ denotes a tight sequence of random variables. From (2.6) and (2.8) we get

$$r_{kk}/\sqrt{n} \xrightarrow{i.p.} 1. \tag{2.9}$$

We have $a_{kk} = 1/r_{kk}$ and for $j < k$ $a_{jk} = R_{kj}/\det(R)$, where R_{kj} is the kj cofactor of R:

$$R_{kj} = (-1)^{k+j}\det(M),$$

and $M = M_{kj}$ is the $(m-1) \times (m-1)$ matrix obtained by deleting the k^{th} row and j^{th} column of R. We have $\det(R) = \Pi_{i=1}^m r_{ii}$. For $\det(M)$ we use the Leibniz formula

$$\det(M) = \sum_{\sigma \in \mathcal{S}_{m-1}} \text{sgn}(\sigma)\Pi_{i=1}^{m-1} m_{i\sigma_i},$$

where $\mathcal{S}_{m-1}$ is the set of all permutations of $\{1, \dots, m-1\}$, the sum is over the collection of all permutations $\sigma \in \mathcal{S}_{m-1}$, and $\text{sgn}(\sigma)$, the signature of σ, is 1 if the reordering of $(1, \dots, m-1)$ given by σ can be brought back to $(1, \dots, m-1)$ by successively interchanging two entries an even number of times, -1 if an odd number of interchanges are needed.

We see then that a_{jk} can be written as a sum of $(m-1)!$ terms. The largest term in absolute value occurs for that σ where all r_{ii} $i \neq j, k$ are included. The remaining entry must be r_{jk}. Indeed, it will lie in row j of M, the only row of M not containing an r_{ii}, $i \neq j, k$, and column $k-1$ of M (column k of R) the only column of M not containing an r_{ii}, $i \neq j, k$. The σ creating this term is necessarily the top row of

$$\begin{matrix} \dots & k-1 & \dots & k-2 & \dots \\ \dots & j & \dots & k-1 & \dots \end{matrix}$$

except when $k = j+1$ in which case the top row is 1 2 ... $m-1$. Here the second row is 1 2 ... $m-1$. All other numbers in the top row are in increasing order. When $k > j+1$ it takes $k-j-1$ pairwise interchanges to bring $k-1$ to the right of $k-2$ (no interchanges when $k = j+1$). Therefore $\text{sgn}(\sigma) = (-1)^{k-j-1}$, and since $j + k + k - j - 1 = 2k - 1$ we have

$$a_{jk} = -r_{jk}/(r_{jj}r_{kk}) + O(1)/n^{3/2}.$$

We have

$$r_{jk}\left(\frac{1}{r_{jj}r_{kk}} - \frac{1}{n}\right) = \frac{r_{jk}}{n}\left(\frac{n}{r_{jj}r_{kk}} - 1\right),$$

so from (9)

$$a_{jk} = -r_{jk}/n + o(1)/n = O(1)/n, \tag{2.10}$$

where here and in the following $o(1)$ denotes a sequence of random variables converging in probability to zero. We have

$$r_{jk} = \left(\mathbf{z}_j^*\mathbf{z}_k - \sum_{i=1}^{j-1}\bar{r}_{ij}r_{ik}\right)/r_{jj} = \mathbf{z}_j^*\mathbf{z}_k/r_{jj} + O(1)/\sqrt{n}$$

$$= \mathbf{z}_j^*\mathbf{z}_k/\sqrt{n} + \frac{\mathbf{z}_j^*\mathbf{z}_k}{\sqrt{n}}\left(\frac{\sqrt{n}}{r_{jj}} - 1\right) + O(1)/\sqrt{n}.$$

By the Central Limit Theorem $\mathbf{z}_j^*\mathbf{z}_k/\sqrt{n}$ is tight. Therefore

$$a_{jk} = -\mathbf{z}_j^*\mathbf{z}_k/n^{3/2} + o(1)/n. \tag{2.11}$$

Let $\|\cdot\|$ represent the sup norm on functions. Write $\mathbf{z}_j = (z_j^1, \dots, z_j^n)^T$. Using (2.4) we have

$$X_n^{k,k}(t) = \sqrt{n}\left(\sum_{i=1}^{[nt]}|a_{kk}z_k^i + \sum_{j=1}^{k-1}a_{jk}z_j^i|^2 - \frac{[nt]}{n}\right)$$

$$= \sqrt{n}\left(a_{kk}^2\sum_{i=1}^{[nt]}|z_k^i|^2 + \sum_{i=1}^{[nt]}|\sum_{j=1}^{k-1}a_{jk}z_j^i|^2 + a_{kk}\sum_{i=1}^{[nt]}\sum_{j=1}^{k-1}\bar{a}_{jk}z_k^i\bar{z}_j^i \right.$$

$$\left. + a_{kk}\sum_{i=1}^{[nt]}\sum_{j=1}^{k-1}a_{jk}\bar{z}_k^i z_j^i - \frac{[nt]}{n}\right).$$

Using Cauchy-Schwarz, Lemma 2.4, the weak Law of Large Numbers, and (2.10) we have

$$\|\sqrt{n}\sum_{i=1}^{[nt]}|\sum_{j=1}^{k-1}a_{jk}z_j^i|^2\| \le n^{3/2}\sum_{i=1}^{k-1}|a_{jk}|^2\frac{1}{n}\sum_{j=1}^{k-1}\|\mathbf{z}_j\|^2 \xrightarrow{i.p.} 0 \tag{2.12}$$

We have using (2.9) and (2.10)

$$\|\sqrt{n}a_{kk}\sum_{i=1}^{[nt]}\sum_{j=1}^{k-1}\bar{a}_{jk}z_k^i\bar{z}_j^i\| \le (O(1)/\sqrt{n})\sum_{j=1}^{k-1}\frac{1}{\sqrt{n}}\|\sum_{i=1}^{[nt]}z_k^i\bar{z}_j^i\| \xrightarrow{i.p.} 0,$$

since $\|\cdot\|$ is continuous on $C[0,1]$, and the real and imaginary parts of $(\sqrt{2/n})\sum_{i=1}^{[nt]}z_k^i\bar{z}_j^i$, each satisfying the assumptions of Donsker's theorem ([3], Theorem 16.1), converge weakly to Wiener measure, which lies in $C[0,1]$, so that from Theorem 5.1 of [3] (with $h = \|\cdot\|$) $\|\frac{1}{\sqrt{n}}\sum_{i=1}^{[nt]}z_k^i\bar{z}_j^i\|$ is tight, and using Lemma 2.4 we get our result.

From (2.6), (2.7), and (2.9) we have

$$\|\sqrt{n}a_{kk}^2\sum_{i=1}^{[nt]}|z_k^i|^2 - \frac{\sqrt{n}}{\|\mathbf{z}_k\|^2}\sum_{i=1}^{[nt]}|z_k^i|^2\| = \|\mathbf{z}_k\|^2\sqrt{n}|a_{kk}^2 - 1/\|\mathbf{z}_k\|^2| = O(1)\frac{\sqrt{n}}{r_{kk}^2} \xrightarrow{i.p.} 0$$

Therefore

$$\|X_n^{k,k} - X_n^k\| \xrightarrow{i.p.} 0,$$

where

$$X_n^k(t) = \sqrt{n}\left(\frac{1}{\|\mathbf{z}_k\|^2}\sum_{i=1}^{[nt]}|z_k^i|^2 - \frac{[nt]}{n}\right).$$

We have

$$X_n^k(t) = \frac{\sqrt{n}}{\|\mathbf{z}_k\|^2}\left(\sum_{i=1}^{[nt]}(|z_k^i|^2 - 1) - \frac{[nt]}{n}(\|\mathbf{z}_k\|^2 - n)\right) = \frac{n}{\|\mathbf{z}_k\|^2}h_n(W_n^k(t)),$$

where

$$W_n^k = \frac{1}{\sqrt{n}}\sum_{i=1}^{[nt]}(|z_k^i|^2 - 1),$$

and $h_n : D \to D$ is defined as $h_n(X) = X(t) - ([nt]/n)X(1)$. Let $h(X) = X(t) - tX(1)$. We have for any $X \in D$ $\|h_n(X) - h(X)\| \leq \|X(1)|\,|t - [nt]/n| \leq |X(1)|/n \to 0$. If $X_n \to X$ in the Skorohod topology, then there exists $\{\lambda_n\}$, each increasing continous on [0,1] with $\lambda_n(0) = 0$, $\lambda_n(1) = 1$, such that $\|\lambda_n(t) - t\| \to 0$ and $\|X_n(t) - X(\lambda_n(t))\| \to 0$. Therefore

$$\|h_n(X_n(t)) - h(X(\lambda_n(t)))\| \leq \|h_n(X_n(t)) - h_n(X(\lambda_n(t)))\| + \|h_n(X(\lambda_n(t))) - h(X(\lambda_n(t)))\|$$

$$\leq \|X_n(t) - X(\lambda_n(t))\| + |X_n(1) - X(1)| + |X(1)|\,|([nt]/n) - t| \to 0.$$

Therefore the set E in Theorem 5.5 of [3] is empty, and by (9.13), Theorem 16.1 and Theorem 5.5 of [3] we have $h_n(W_n^k) \to_D h(W) = W^\circ$, W denoting Wiener measure.

We have $\|X_n^k - h_n(W_n^k)\| \leq |1 - n/\|\mathbf{z}_k\|^2| \max_t |h_n(W_n^k(t))|$. By (2.6) we have $|1 - n/\|\mathbf{z}_k\|^2| \xrightarrow{i.p.} 0$. Again, from Theorem 5.1 of [3] we have $\|h_n(W_n^k)\| \to_D \|W^\circ\|$. Therefore, by Lemma 2.4 we have

$$\|X_n^k - h_n(W_n^k)\| \xrightarrow{i.p.} 0.$$

Therefore, $X_n^{k,k} \to_D W^\circ$.

For $k < k'$

$$X_n^{k,k'}(t) = \sqrt{2n}\left(\sum_{i=1}^{[nt]}(a_{kk}\bar{z}_k^i + \sum_{j=1}^{k-1}\bar{a}_{jk}\bar{z}_j^i)(a_{k'k'}z_{k'}^i + \sum_{j'=1}^{k'-1}a_{j'k'}z_{j'}^i)\right)$$

$$= \sqrt{2n}\left(a_{kk}a_{k'k'}\sum_{i=1}^{[nt]}\bar{z}_k^i z_{k'}^i + a_{kk}\sum_{j'=1}^{k'-1}a_{j'k'}\sum_{i=1}^{[nt]}\bar{z}_k^i z_{j'}^i \right.$$
$$\left. + a_{k'k'}\sum_{j=1}^{k-1}\bar{a}_{jk}\sum_{i=1}^{[nt]}\bar{z}_j^i z_{k'}^i + \sum_{i=1}^{[nt]}(\sum_{j=1}^{k-1}\bar{a}_{jk}\bar{z}_j^i)(\sum_{j'=1}^{k'-1}a_{j'k'}z_{j'}^i)\right)$$

From Cauchy-Schwarz and (2.12) we have

$$\left\|\sqrt{n}\sum_{i=1}^{[nt]}(\sum_{j=1}^{k-1}\bar{a}_{jk}\bar{z}_j^i)(\sum_{j'=1}^{k'-1}a_{j'k'}z_{j'}^i)\right\|$$
$$\leq \left\|\left(\sqrt{n}\sum_{i=1}^{[nt]}\left|\sum_{j=1}^{k-1}\bar{a}_{jk}\bar{z}_j^i\right|^2\right)^{1/2}\left(\sqrt{n}\sum_{i=1}^{[nt]}\left|\sum_{j'=1}^{k'-1}a_{j'k'}z_{j'}^i\right|^2\right)^{1/2}\right\| \xrightarrow{i.p.} 0$$

Similar to what was done earlier we have for $j' \neq k$ and $j \neq k'$ we have both

$$\left\|\sqrt{n}a_{kk}a_{j'k'}\sum_{i=1}^{[nt]}\bar{z}_k^i z_{j'}^i\right\| \quad \text{and} \quad \left\|\sqrt{n}a_{k'k'}\bar{a}_{jk}\sum_{i=1}^{[nt]}\bar{z}_j^i z_{k'}^i\right\|$$

converging in probability to zero. Also

$$\left\| \sqrt{n} a_{kk} a_{k'k'} - \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} \bar{z}_k^i z_{k'}^i \right\| = \left| \frac{n}{r_{kk} r_{k'k'}} - 1 \right| \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} \bar{z}_k^i z_{k'}^i \right\| \xrightarrow{i.p.} 0.$$

We have using (2.11)

$$\left\| \frac{1}{\sqrt{n}} z_k^* z_{k'} \frac{[nt]}{n} + \sqrt{n} a_{kk} a_{kk'} \sum_{i=1}^{[nt]} |z_k^i|^2 \right\|$$

$$= \left\| \frac{1}{\sqrt{n}} z_k^* z_{k'} \frac{[nt]}{n} + \sqrt{n} a_{kk} a_{kk'} [nt] + n a_{kk} a_{kk'} \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} (|z_k^i|^2 - 1) \right\|$$

$$\leq \left\| \frac{1}{\sqrt{n}} z_k^* z_{k'} \frac{[nt]}{n} + \sqrt{n} a_{kk} [nt](-z_k^* z_{k'}/n^{3/2} + o(1)/n) \right\| + n a_{kk} |a_{kk'}| \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} (|z_k^i|^2 - 1) \right\|.$$

Since the function inside the norm of the second term converges weakly to Wiener measure, the second term converges in probability to zero. The first term is

$$\leq \left| \frac{1}{\sqrt{n}} \mathbf{z}_k^* \mathbf{z}_{k'} \right| |1 - \sqrt{n} a_{kk}| + o(1) \sqrt{n} a_{kk} \xrightarrow{i.p.} 0.$$

Therefore

$$\left\| X_n^{k,k'} - \sqrt{\frac{2}{n}} \left(\sum_{i=1}^{[nt]} \bar{z}_k^i z_{k'}^i - \frac{[nt]}{n} z_k^* z_{k'} \right) \right\| \xrightarrow{i.p.} 0.$$

We separate out the real and imaginary parts of the process $X_n^{k,k'}$ is approaching. Write $z_k = z_{kr} + i z_{ki}$, $z_{k'} = z_{k'r} + i z_{k'i}$. Then the real and imaginary parts of $X_n^{k,k'}$ are approaching, respectively

$$\sqrt{\frac{2}{n}} \left(\sum_{j=1}^{[nt]} (z_{kr}^j z_{k'r}^j + z_{ki}^j z_{k'i}^j) - \frac{[nt]}{n} \sum_{j=1}^{n} (z_{kr}^j z_{k'r}^j + z_{ki}^j z_{k'i}^j) \right) = h_n(W_n^{k,k',r}(t))$$

and

$$\sqrt{\frac{2}{n}} \left(\sum_{j=1}^{[nt]} (z_{kr}^j z_{k'i}^j - z_{ki}^j z_{k'r}^j) - \frac{[nt]}{n} \sum_{j=1}^{n} (z_{kr}^j z_{k'i}^j - z_{ki}^j z_{k'r}^j) \right) = h_n(W_n^{k,k',i}(t))$$

where

$$W_n^{k,k',r}(t) = \sqrt{\frac{2}{n}} \sum_{j=1}^{[nt]} (z_{kr}^j z_{k'r}^j + z_{ki}^j z_{k'i}^j) \quad \text{and} \quad W_n^{k,k',i}(t) = \sqrt{\frac{2}{n}} \sum_{j=1}^{[nt]} (z_{kr}^j z_{k'i}^j - z_{ki}^j z_{k'r}^j).$$

It is clear now that each of $X_n^{k,k}$, $\Re X_n^{k,k'}$, $\Im X_n^{k,k'}$ converges weakly to Brownian bridge. In order to show they converge weakly in D_{m^2} to independent copies of W°, we will show the weak convergence of the W_n^k, $W_n^{k,k',r}$, $W_n^{k,k',i}$ to W^k, $W^{k,k',r}$, $W^{k,k',i}$, independent copies of Wiener measure, using (9.13), Theorem 5.5 (on D_{m^2}), and Theorem 16.1 all in [3].

Let W_n denote the $m \times m$ matrix consisting of the W_n^k on the diagonal, the $W_n^{k,k',r}$ on the lower diagonal, and the $W_n^{k,k',i}$ on the upper diagonal. Let W denote an $m \times m$ matrix consisting of independent copies of Wiener measure.

We have each entry of W_n is tight, satisfying the first condition of Lemma 2.3. Choose k, $0 \le t_1 < \cdots < t_k \le 1$. To prove

$$(W_n(t_1), \ldots, W_n(t_k)) \to_D (W(t_1), \ldots, W(t_k)) \tag{2.13}$$

it is sufficient to show

$$\begin{aligned}(W_n(t_1), W_n(t_2) - W_n(t_1), \ldots, W_n(t_k) - W_n(t_{k-1})) \\ \to_D (W(t_1), W(t_2) - W(t_1), \ldots, W(t_k) - W(t_{k-1})).\end{aligned}$$

But the k matrices $W_n(t_\ell) - W_n(t_{\ell-1})$, where $t_0 \equiv 0$, are independent. By the natural extension to Theorem 3.2 in [3] it is sufficient to show each of these converges in distribution. We use the Cramér-Wold device (p. 48 of [3]). Thus we need to prove that linear combinations of the entries of $W_n(t_\ell) - W_n(t_{\ell-1})$ converge in distribution to the corresponding linear combinations of the entries of $W(t_\ell) - W(t_{\ell-1})$. Fix $A = (a_{ij}) \in \mathbb{R}^{m \times m}$. Let $\circ$ denote Hadamard product on $m \times m$ matrices and let $\mathbf{1}$ denote the m dimensional column vector consisting of 1's. Let

$$Y = \mathbf{1}^{\mathrm{T}}\big(A \circ \sqrt{n} W_n(1/n)\big)\mathbf{1}.$$

We have $\mathsf{E}Y = 0$ and $\mathsf{E}(Y^2) = \sum_{i,j} a_{ij}^2$. Therefore, from the central limit theorem

$$\mathbf{1}^{\mathrm{T}}\big(A \circ (W_n(t_\ell) - W_n(t_{\ell-1}))\big)\mathbf{1} \to_D N(0, (t_\ell - t_{\ell-1}) \sum_{i,j} a_{ij}^2),$$

the same distribution as $\mathbf{1}^{\mathrm{T}}\big(A \circ (W(t_\ell) - W(t_{\ell-1}))\big)\mathbf{1}$. Therefore, by Lemma 2.3, we are done.

It is clear that the analysis carries over to the real case, so that Theorem 1.2 is true. Indeed, when Z consists of i.i.d. standard Gaussian, we use in Lemma 2.1 the fact that for any $Q \in \mathcal{O}_n$ $QX \sim X$, and for the scaling of the X_n^k and Y_n^{jk} we have now the variance of a standard Gaussian is 1, while its fourth moment is 3.

3. Proof of Theorem 1.3. We let F_n denote the empirical distribution function of M_n with almost sure limiting distribution function F_y specified above. We will also use the fact [20] that, because $\mathsf{E}v_{11}^4 < \infty$, $\lambda_{\max}(M_n)$, the largest eigenvalue of M_n satisfies

$$\lambda_{\max}(M_n) \to (1+\sqrt{y})^2 \quad \text{a.s. as } n \to \infty. \tag{3.1}$$

We begin with two lemmas.

Lemma 3.1 Let S be a metric space with X_n, X random elements in S and $X_n \to_D X$. Suppose for each n, ℓ_n is a random positive integer, independent of $\{X_n\}$ such that for any positive integer j, $\mathsf{P}(\ell_n \le j) \to 0$ as $n \to \infty$. Then $X_{\ell_n} \to_D X$.

Proof: Let A be an X-continuity set. For any positive integer j we have

$$\begin{aligned}\mathsf{P}(X_{\ell_n} \in A | \ell_n = j) &= \mathsf{P}(X_{\ell_n} \in A, \ell_n = j)/\mathsf{P}(\ell_n = j) \\ &= \mathsf{P}(X_j \in A, \ell_n = j)/\mathsf{P}(\ell_n = j) = \mathsf{P}(X_j \in A).\end{aligned}$$

For $\epsilon > 0$ let positive integer M_1 be such that $|\mathsf{P}(X_j \in A) - \mathsf{P}(X \in A)| < \epsilon/2$ for all $j \ge M_1$. Let $M \ge M_1$ be such that $\mathsf{P}(\ell_n \le M_1) < \epsilon/4$ for all $n \ge M$. Then, using

$$\mathsf{P}(X_{\ell_n} \in A) = \sum_{j=1}^{\infty} \mathsf{P}(X_j \in A)\mathsf{P}(\ell_n = j)$$

we have for all $n \ge M$

$$\begin{aligned}|\mathsf{P}(X \in A) - \mathsf{P}(X_{\ell_n} \in A)| \le &\sum_{j=M_1+1}^{\infty} |\mathsf{P}(X \in A) - \mathsf{P}(X_j \in A)|\mathsf{P}(\ell_n = j) \\ &+ \sum_{j=1}^{M_1} |\mathsf{P}(X \in A) - \mathsf{P}(X_j \in A)|\mathsf{P}(\ell_n = j) < \epsilon/2 + \epsilon/2 = \epsilon.\end{aligned}$$

Therefore since ϵ was arbitrary we have $X_{\ell_n} \to_D X$.

Lemma 3.2 Let S' and S'' be separable metric spaces, with X', X'_n random elements of S', defined on probability space P', and X'', X''_n random elements of S'', defined on probability space P'' and let $\mathsf{P} = \mathsf{P}' \times \mathsf{P}$". Then $\{X'_n\}, X'$ and $\{X''_n\}, X''$ are independent on P. Suppose $X'_n \to_D X'$, $X''_n \to_D X''$ and for each n there exists a positive integer-valued function $\ell_n = \ell_n(X'_n)$ for which the ℓ_n satisfy the condition in Lemma 3.1. Then $(X'_n, X''_{\ell_n}) \to_D (X', X'')$ on P.

Proof: From Lemma 3.1 we have $X''_{\ell_n} \to_D X''$. Let A', A'' be respective X', X'' -continuity sets. Then for each n

$$\mathsf{P}(X'_n \in A', X''_{\ell_n} \in A'') = \sum_{j=1}^{\infty} \mathsf{P}(X'_n \in A', X''_{\ell_n} \in A'', \ell_n = j)$$
$$= \sum_{j=1}^{\infty} \mathsf{P}(X'_n \in A', X''_j \in A'', \ell_n = j) = \sum_{j=1}^{\infty} \mathsf{P}(X''_j \in A'')\mathsf{P}(X'_n \in A', \ell_n = j)$$
$$= \sum_{j=1}^{\infty} \mathsf{P}(X''_j \in A'')\mathsf{P}(X'_n \in A' | \ell_n = j)\mathsf{P}(\ell_n = j).$$

$$\mathsf{P}(X'_n \in A', X''_{\ell_n} \in A'') - \mathsf{P}(X' \in A')\mathsf{P}(X'' \in A'')$$
$$= \mathsf{P}(X'_n \in A', X''_{\ell_n} \in A'') - \mathsf{P}(X'_n \in A')\mathsf{P}(X'' \in A'')$$
$$+ \mathsf{P}(X'_n \in A')\mathsf{P}(X'' \in A'') - \mathsf{P}(X' \in A')\mathsf{P}(X'' \in A'')$$
$$= \sum_{j=1}^{\infty} (\mathsf{P}(X''_j \in A'') - \mathsf{P}(X'' \in A''))\mathsf{P}(X'_n \in A' | \ell_n = j)\mathsf{P}(\ell_n = j)$$
$$+ \mathsf{P}(X'_n \in A')\mathsf{P}(X'' \in A'') - \mathsf{P}(X' \in A')\mathsf{P}(X'' \in A'')$$

For $\epsilon > 0$ let M_1 be such that for all $j \geq M_1$

$$\max(|\mathsf{P}(X'_j \in A') - \mathsf{P}(X' \in A')|, |\mathsf{P}(X'_j \in A'') - \mathsf{P}(X'' \in A'')|) < \epsilon/3.$$

Let $M \geq M_1$ be such that for all $n \geq M$ $\mathsf{P}(\ell_n \leq M_1) < \epsilon/6$. Then for all $n \geq M$

$$|\mathsf{P}(X'_n \in A', X''_{\ell_n} \in A'') - \mathsf{P}(X' \in A')\mathsf{P}(X'' \in A'')|$$
$$\leq \sum_{j=M_1+1}^{\infty} |\mathsf{P}(X''_j \in A'') - \mathsf{P}(X'' \in A'')|\mathsf{P}(X'_n \in A' | \ell_n = j)\mathsf{P}(\ell_n = j)$$

$$+ \sum_{j=1}^{M_1} |\mathsf{P}(X''_j \in A'' - \mathsf{P}(X'' \in A'')|\mathsf{P}(X'_n \in A' | \ell_n = j)\mathsf{P}(\ell_n = j) + \epsilon/3 < \epsilon.$$

Since ϵ is arbitrary we have the result.

Recalling Y_n^{jk} in (1.5), let $Y_n = Y_n^{12}$. Much of the following are modifications to the results in [16], with X_n replaced by Y_n, with some being used exactly as stated in that paper. As in [16] some of the results make assumptions more general than what is needed to prove Theorem 1.2, in order to be able to use them in the future. Results in [15] will also be used and modified.

We proceed to prove Theorem 2.1 of [16] with X_n replaced by Y_n. We also assume that $X_n^i(F_n(\cdot)) \to_D W^0_{F_y(\cdot)}$ on $D[0,\infty)$ for $i=1,2$. Let ρ denote the sup metric in $C[0,1]$:

$$\rho(x,y) = \sup_{t\in[0,1]} |x(t)-y(t)| \quad \text{for } x,y \in D[0,1].$$

THEOREM 3.1. $Y_n(F_n(\cdot))$, $X_n^i(F_n(\cdot))$, $i=1,2$ all converging weakly to $W^0_{F_y(\cdot)}$, in $D[0,\infty)$, $F_n \to_D F_y$ i.p., and $\lambda_{\max} \equiv \lambda_{\max}(M_n) \to (1+\sqrt{y})^2$ i.p. $\Rightarrow Y_n \to_D W^0$.

Proof: The proof of Theorem 2.1 in [16] applied to Y_n remains unchanged up to the middle of p. 1179. For fixed M_n let $\lambda_{(1)} < \lambda_{(2)} < \cdots < \lambda_{(t)}$ be the t distinct eigenvalues of M_n with multiplicities $m_1, m_2, \ldots, m_t$. For fixed eigenvalue $\lambda_{(i)}$ the corresponding m_i columns of O_n are distributed as $O_{n,i}O_i$ where $O_{n,i}$ is $n\times m_i$ containing m_i orthonormal columns from the eigenspace of $\lambda_{(i)}$, and O_i is Haar distributed in the group of $m_i \times m_i$ orthogonal matrices, independent of M_n. The coordinates of $\mathbf{y}_1$ and $\mathbf{y}_2$ corresponding to $\lambda_{(i)}$ are respectively of the form

$$(O_{n,i}O_i)^T\mathbf{x}_{n,1} = a_{1,i}\mathbf{w}_{1,i} \quad \text{and } (O_{n,i}O_i)^T\mathbf{x}_{n,2} = a_{2,i}\mathbf{w}_{2,i},$$

where $a_{1,i} = \|O_{n,i}^T\mathbf{x}_{n,1}\|$, $a_{2,i} = \|O_{n,i}^T\mathbf{x}_{n,2}\|$, and $\mathbf{w}_{1,i} = (w_{1,i}^1, w_{1,i}^2, \ldots, w_{1,i}^{m_i})^T$, $\mathbf{w}_{2,i} = (w_{2,i}^1, w_{2,i}^2, \ldots, w_{2,i}^{m_i})^T$ are each uniformly distributed on the unit sphere in $\mathbb{R}^{m_i}$. Write

$$(O_{n,i}O_i)^T(\mathbf{x}_{n,1}+\mathbf{x}_{n,2}) = a_{1,2,i}\mathbf{w}_{1,2,i},$$

where $a_{1,2,i} = \|O_{n,i}^T(\mathbf{x}_{n,1}+\mathbf{x}_{n,2})\|$ and $\mathbf{w}_{1,2,i} = (w_{1,2,i}^1, w_{1,2,i}^2, \ldots, w_{1,2,i}^{m_i})^T$ is uniformly distributed on the unit sphere in $\mathbb{R}^{m_i}$. We have (2.4) in [16] holding for $a_i = a_{1,i}$ and $a_{2,i}$. Also as in (2.4) in [16] we have

$$\max_{1\le i\le t} \sqrt{n}|\mathbf{x}_{n,1}^T O_{n,i} O_{n,i}^T \mathbf{x}_{n,2}| \xrightarrow{i.p.} 0. \tag{3.1}$$

We have (2.3) in [16] for Y_n becomes

$$\rho(Y_n(\cdot), Y_n(F_n(F_n^{-1}(\cdot)))) = \max_{\substack{1\le i\le t\\ 1\le j\le m_i}} \sqrt{n}\left| a_{1,i}a_{2,i}\sum_{\ell=1}^{j} w_{1,i}^\ell w_{2,i}^\ell \right|. \tag{3.2}$$

For each $i \le t$ and $j \le m_i$

$$\sqrt{n}a_{1,i}a_{2,i}\sum_{\ell=1}^{j} w_{1,i}^\ell w_{2,i}^\ell = \frac{\sqrt{n}}{2}\left(a_{1,2,i}^2 \sum_{\ell=1}^{j}(w_{1,2,i}^\ell)^2 - a_{1,i}^2\sum_{\ell=1}^{j}(w_{1,i}^\ell)^2 - a_{2,i}^2\sum_{\ell=1}^{j}(w_{2,i}^\ell)^2 \right)$$

(a)

$$= \frac{\sqrt{n}}{2}\left(\left(a_{1,2,i}^2 - 2\frac{m_i}{n}\right)\sum_{\ell=1}^{j}(w_{1,2,i}^\ell)^2 - \left(a_{1,i}^2 - \frac{m_i}{n}\right)\sum_{\ell=1}^{j}(w_{1,i}^\ell)^2 - \left(a_{2,i}^2 - \frac{m_i}{n}\right)\sum_{\ell=1}^{j}(w_{2,i}^\ell)^2 \right)$$

(b)

$$+\frac{\sqrt{n}}{2}\left(2\frac{m_i}{n}\left(\sum_{\ell=1}^{j}(w_{1,2,i}^{\ell})^2-\frac{j}{m_i}\right)-\frac{m_i}{n}\left(\sum_{\ell=1}^{j}(w_{1,i}^{\ell})^2-\frac{j}{m_i}\right)-\frac{m_i}{n}\left(\sum_{\ell=1}^{j}(w_{2,i}^{\ell})^2-\frac{j}{m_i}\right)\right).$$

From (3.1) above and (2.4) in [16] we see the maximum of the absolute value of (a) over all $j \le m_i$, $1 \le i \le t$ converges in probability to zero. We see that the three sums in (b) are beta distributed the same as in (b) of[16] p. 1180. Therefore the same arguments leading to the convergence of (2.3) of [16] to zero in probability give us the convergence of (3.2) to zero i.p. Therefore for $y \le 1$ we have $Y_n \to_D W^0$.

For $y > 1$, the main difference is the appearance of $Y_n(t) = Y_n^{12}$ for $t < F_n(0) + 1/n$. Let $\underline{\mathbf{x}}_{n,1} = O_{n,1}^T \mathbf{x}_{n,1}$, $\underline{\mathbf{x}}_{n,2} = O_{n,1}^T \mathbf{x}_{n,2}$, and o_i denote the i^{th} column of O_1. Notice that $a_{i,1} = \|\underline{\mathbf{x}}_{n,i}\|$, $i = 1, 2$. We have

$$X_n^i(F_n(0)) = \sqrt{\frac{n}{2}}(a_{i,i}^2 - F_n(0)) \to_D W_{F_y(0)} \quad \text{as } n \to \infty$$

$i = 1, 2$. therefore, from Lemma 2.4

$$a_{i,1}^2 \xrightarrow{i.p.} F_y(0) = 1 - (1/y), \quad i = 1, 2. \tag{3.3}$$

Write

$$\underline{\mathbf{x}}_{n,1} = \frac{\underline{\mathbf{x}}_{n,1}^T \underline{\mathbf{x}}_{n,2}}{a_{2,1}^2}\underline{\mathbf{x}}_{n,2} + \underline{\mathbf{z}}.$$

We have $\underline{\mathbf{z}}^T \underline{\mathbf{x}}_{n,2} = 0$ and

$$\|\underline{\mathbf{z}}\| = \frac{\sqrt{a_{1,1}^2 a_{2,1}^2 - (\underline{\mathbf{x}}_{n,1}^T \underline{\mathbf{x}}_{n,2})^2}}{a_{2,1}}.$$

Notice that $\sqrt{n}\underline{\mathbf{x}}_{n,1}^T \underline{\mathbf{x}}_{n,2} = Y_n(F_n(0))$. Therefore from Lemma 2.4

$$\underline{\mathbf{x}}_{n,1}^T \underline{\mathbf{x}}_{n,2} \xrightarrow{i.p.} 0. \tag{3.4}$$

For $t < F_n(0) + 1/n$

$$Y_n(t) = \sqrt{n}\sum_{i=1}^{[nt]} \underline{\mathbf{x}}_{n,1}^T o_i o_i^T \underline{\mathbf{x}}_{n,2} = \sqrt{\frac{2}{F_n(0)}}\frac{Y_n(F_n(0))}{\sqrt{n}} A_n(t) + \frac{\sqrt{a_{1,1}^2 a_{2,1}^2 - (\underline{\mathbf{x}}_{n,1}^T \underline{\mathbf{x}}_{n,2})^2}}{\sqrt{F_n(0)}} B_n(t) + Y_n(F_n(0))\frac{[nt]}{nF_n(0)}$$

where

$$A_n(t) = \sqrt{\frac{nF_n(0)}{2}}\left(\sum_{i=1}^{[nt]} \frac{\underline{\mathbf{x}}_{n,2}^T}{a_{2,1}} o_i o_i^T \frac{\underline{\mathbf{x}}_{n,2}}{a_{2,1}} - \frac{[nt]}{nF_n(0))}\right)$$

and

$$B_n(t) = \sqrt{nF_n(0)} \sum_{i=1}^{[nt]} \frac{\underline{\mathbf{z}}^T}{\|\underline{\mathbf{z}}\|} o_i o_i^T \frac{\underline{\mathbf{x}}_{n,2}}{a_{2,1}}$$

Since O_1 is Haar distributed and independent of M_n, we see that A_n and B_n have the same distribution if $\underline{\mathbf{x}}_{n,2}/a_{2,1}$ and $\underline{\mathbf{z}}/\|\underline{\mathbf{z}}\|$ were nonrandom orthonormal vectors. $H_n(t)$ in [16] now becomes

$$H_n(t) = \sqrt{\frac{2}{F_n(0)}} \frac{Y_n(F_n(0))}{\sqrt{n}} A_n(F_n(0)\varphi_n(t)) + \frac{\sqrt{a_{1,1}^2 a_{2,1}^2 - (\underline{\mathbf{x}}_{n,1}^T \underline{\mathbf{x}}_{n,2})^2}}{\sqrt{F_n(0)}} B_n(F_n(0)\varphi_n(t))$$

$$+Y_n(F_n(0)) \left(\frac{[nF_n(0)\varphi_n(t)]}{nF_n(0)} - 1 \right) + Y_n(F_n(F_n^{-1}(t)),$$

where $\varphi_n(t) = \min(t/F_n(0), 1)$ for $t \in [0,1]$. Denote the sum of the last two terms by (a). Notice that for $s \in [0,1]$, from Theorem 1.2, both $A_n(F_n(0)s)$ and $B_n(F_n(0)s)$ converge weakly to independent Brownian bridges. We apply Lemma 3.2 where $X'_n = ((a), F_n(0))$, $\ell_n = nF_n(0)$, and $X''_{\ell_n} = (A_n(F_n(0)s), B_n(F_n(0)s)$. Since, from (3.3) and (3.4) the coefficient of A_n converges i.p. to zero and the coefficient of B_n converges i.p. to $\sqrt{F_y(0)} = \sqrt{1-(1/y)}$ we have H_n converging weakly to H appearing in [16] (notice the misprint on line 8, p. 1183 of [16]. The zero to the right of the arrow should be $\varphi(t)$). The final argument is exactly the same as in [16]. This completes the proof of the theorem

The next step is to extend Theorem 3.1 of [16] to random elements in $(D_d^b, \mathcal{T}_d^b)$. We denote the modulus of continuity of $x \in D[0,b]$ by $w(x,\cdot)$:

$$w(x,\delta) = \sup_{|s-t|<\delta} |x(s) - x(t)|, \quad \delta \in (0,b].$$

THEOREM 3.2. Let $\{(x_n^1, \dots, x_n^d)\}$ be a sequence of random elements of D_d^b, defined on a common probability space, each $\{x_n^i\}$ satisfy the assumptions of Theorem 15.5 of [3]: $\{x_n^i(0)\}$ is tight and for every positive ϵ and η, there exists a $\delta \in (0,b)$ and an integer n_0, such that, for all $n > n_0$, $\mathsf{P}(w(x_n^i, \delta) \geq \epsilon) \leq \eta$. If there exists a random element $(x^1, \dots, x^d)$ with $\mathsf{P}(x^i \in C[0,b]) = 1$ for each i, and such that

(3.5)

$$\left\{ \left(\int_0^b t^r x_n^1 dt, \dots, \int_0^b t^r x_n^d dt \right) \right\}_{r=0}^{\infty} \to_D \left\{ \left(\int_0^b t^r x^1 dt, \dots, \int_0^b t^r x^d dt \right) \right\}_{r=0}^{\infty} \quad \text{as } n \to \infty$$

(D denoting weak convergence on $\mathbb{R}^\infty$), then $(x_n^1, \dots, x_n^d) \Rightarrow (x^1, \dots, x^d)$.

Proof. From Theorems 5.1 and 15.5 of [3] and Lemma 2.3 weak convergence will follow from showing the distribution of

$$(x^1(t_1), \dots, x^1(t_k), \dots, x^d(t_1), \dots, x^d(t_k))$$

for all k, $t_1, \dots, t_k \in [0,1]$ is uniquely determined by the distribution of

$$\left\{ \left(\int_0^1 t^r x^1 dt, \dots, \int_0^1 t^r x^d dt \right) \right\}_{r=0}^{\infty}. \tag{3.6}$$

This is achieved by showing the distribution of

$$\sum_{i=1}^{d} \sum_{j=1}^{k} a_{ij} x^i(t_j)$$

is uniquely determined by the distribution of (3.6). By a simple extension of the proof of Theorem 3.1 in [16] this can be done.

Next we prove the analog of Theorem 4.2 in [16]. Write

$$Y_n(F_n(x)) = \sqrt{n}\mathbf{x}_{n,1}^T P^{M_n}([0,x])\mathbf{x}_{n,2},$$

$P^{M_n}(A)$ being the projection matrix on the subspace of $\mathbb{R}^n$ spanned by the eigenvectors of M_n having eigenvalues in A, a measurable subset of $\mathbb{R}^+$. Assuming v_{11} is symmetric, we have the following results from [16]:

Fact 3 in [16]: $P^{M_n}(A) \sim OP^{M_n}(A)O^t$ for any permutaion matrix O.

Lemma 4.1 in [16]: If one of the indices $i_1, j_1, \dots, i_4, j_4$ appears an odd number of times, then for Borel sets $A_1, \dots, A_4 \in \mathbb{R}^+$

$$\mathsf{E}\big(P_{i_1 j_1}^{M_n}(A_1)P_{i_2 j_2}^{M_n}(A_2)P_{i_3 j_3}^{M_n}(A_3)P_{i_4 j_4}^{M_n}(A_4)\big) = 0.$$

Assume also that each $\mathbf{x}_{n,j} = (\pm 1/\sqrt{n}, \dots \pm 1/\sqrt{n})^T$ and are orthogonal. Then necessarily n is even, say $n = 2p$, and exactly p entries of $\mathbf{x}_{n,2}$ are of opposite sign with the corresponding entries of $\mathbf{x}_{n,1}$. Moreover, Fact 3 in [16] is true for O diagonal with ± 1's on its diagonal, using exactly the same argument. If O is diagonal of this type with signs matching those of $\mathbf{x}_{n,1}$ coordinatewise, then

$$Y_n(F_n(x)) = \sqrt{n}(O\mathbf{x}_{n,1})^T OP^{M_n}([0,x])O^T O\mathbf{x}_{n,2} \sim \sqrt{n}(O\mathbf{x}_{n,1})^T P^{M_n}([0,x])O\mathbf{x}_{n,2}. \tag{3.7}$$

Therefore we can assume the sign of all the entries of $\mathbf{x}_{n,1}$ are positive. Let now O be a permutation matrix which moves all the positive entries of the new $\mathbf{x}_{n,2}$ to the first p positions. Then using (3.7) again we conclude that we can assume that all the entries of $\mathbf{x}_{n,1}$ and the first p entries of $\mathbf{x}_{n,2}$ are positive, and that the remaining entries of $\mathbf{x}_{n,2}$ are negative.

Theorem 3.3. Assume v_{11} is symmetrically distributed about 0, $\mathbf{x}_{n,j} = (\pm 1/\sqrt{n}, \dots, \pm 1/\sqrt{n})^T$, $j = 1, 2$, and are orthogonal. Then

$$\mathsf{E}\big(Y_n(F_n(0))\big)^4 \le \mathsf{E}(27P_{11}^{M_n}(\{0\}))^2 \tag{3.8}$$

and for $0 \le x_1 \le x_2$

$$\mathsf{E}\big(Y_n(F_n(x_2)) - Y_n(F_n(x_1))\big)^4 \le \mathsf{E}(27P_{11}^{M_n}((x_1,x_2]))^2 \tag{3.9}$$

Proof: With $A = \{0\}$ or $(x_1, x_2]$ (corresponding to (3.8), (3.9) respectively, we have

$$\mathsf{E}\big(Y_n(F_n(0))\big)^4 = \frac{1}{n^2}\mathsf{E}\left(\sum_{i\le n; j\le p} P_{ij}^{M_n}(A) - \sum_{p+1\le i,j\le n} P_{ij}^{M_n}(A)\right)^4 \tag{3.10}$$

$$= \frac{1}{n^2}\mathsf{E}\left(\sum_{i\le p} P_{ii}^{M_n}(A) - \sum_{p+1\le i\le n} P_{ii}^{M_n}(A) + 2\sum_{i<j\le p} P_{ij}^{M_n}(A) - 2\sum_{p+1\le i<j\le n} P_{ij}^{M_n}(A)\right)^4$$

$\le$ (using for nonnegative a, b, c $(a+b+c)^4 \le 27(a^4+b^4+c^4)$)

$$\text{(a)} \qquad \frac{27}{n^2}\mathsf{E}\left(\sum_{i\le p} P_{ii}^{M_n}(A) - P_{i+p\ i+p}^{M_n}(A)\right)^4$$

$$+$$

$$\text{(b)} \qquad \frac{54}{n^2}\mathsf{E}\left(\sum_{\substack{i,j\le p\\ i\ne j}} P_{ij}^{M_n}(A)\right)^4,$$

where in (b) we used Fact 3 of [16], which says that P^{M_n} is distributed the same as $OP^{M_n}O^T$ for permutation matrices O, on the $P_{ij}^{M_n}$'s with $i \ne j$ and both larger than p. Suppressing the dependence on M_n and A, we have from Fact 3 and Lemma 4.1 in [16]

$$\text{(b)} = \frac{216p(p-1)}{n^2}\big(12(p-2)\mathsf{E}(P_{12}^2P_{13}^2) + 3(p-2)(p-3)\mathsf{E}(P_{12}^2P_{34}^2) + 12(p-2)(p-3)\mathsf{E}(P_{12}P_{23}P_{34}P_{14}) + 2\mathsf{E}(P_{12}^4)\big).$$

Bounds involving $\mathsf{E}(P_{12}P_{23}P_{34}P_{14})$ and $\mathsf{E}(P_{12}^2P_{34}^2)$ were derived in [16], from which we get

$$(n-2)(n-3)\mathsf{E}(P_{12}P_{23}P_{34}P_{14}) \le \mathsf{E}(P_{11}P_{22})$$

and

$$(n-2)(n-3)\mathsf{E}(P_{12}^2P_{34}^2) \le \mathsf{E}(P_{11}P_{22}).$$

A bound on $\mathsf{E}(P_{12}^2P_{13}^2)$ is also needed. Starting from the fact that $P^2 = P$, we take the expected value of both sides of

$$P_{12}^2\left(\sum_{j\geq 3} P_{1j}^2 + P_{11}^2 + P_{12}^2\right) = P_{12}^2 P_{11}$$

and use Fact 3 in [16] to get

$$(n-2)\mathsf{E}(P_{12}^2P_{13}^2) \leq \mathsf{E}(P_{12}^2P_{11}).$$

Therefore for $p \geq 2$

$$\text{(b)} \leq \frac{216p(p-1)}{n^2}\left(12\frac{p-2}{n-2}\mathsf{E}(P_{12}^2P_{11}) + 15\frac{(p-2)(p-3)}{(n-2)(n-3)}\mathsf{E}(P_{11}P_{22}) + 2\mathsf{E}(P_{12}^4)\right).$$

Thus, using Fact 3 in [16] and the facts that $P_{11} \in [0,1]$, $P_{12}^2 \leq P_{11}P_{22}$ since P is nonnegative definite, and $ab \leq \frac{1}{2}(a^2+b^2)$, we get

$$\text{(b)} < 648\mathsf{E}(P_{11}^2).$$

In (a) we expand the fourth power of the sum. Using Fact 3 in [16] we see that any term involving an odd number of $P_{ii} - P_{i+p\ i+p}$ is zero. Therefore

$$(a) = \frac{27p}{n^2}(\mathsf{E}(P_{11}-P_{22})^4 + 3(p-1)\mathsf{E}(P_{11}-P_{22})^2(P_{33}-P_{44})^2 \leq 27\mathsf{E}(P_{11}-P_{22})^2$$
$$\leq 54\mathsf{E}P_{11}^2.$$

Therefore, the expression in (3.10) is bounded by $\mathsf{E}(27P_{11})^2$, and the proof is complete.

Notice that for unit $\mathbf{x}_n \in \mathbb{R}^n$ $\mathbf{x}_n^T P_{11}^{M_n}(\cdot)\mathbf{x}_n$ is a (random) probability measure with mass at the eigenvalues of M_n. In [15] it is proven that

(3.11)
$$\{\sqrt{n/2}(\mathbf{x}_n^T M_n^r \mathbf{x}_n - (1/n)\mathsf{tr}\,(M_n^r))\}_{r=1}^{\infty} \to_D \left\{\int_{(1-\sqrt{y})^2}^{(1+\sqrt{y})^2} x^r dW_{F_y(x)}^0\right\}_{r=1}^{\infty} \quad as \quad n \to \infty$$

(D denoting weak convergence on $\mathbb{R}^\infty$) for every sequence $\{\mathbf{x}_n\}$, $\mathbf{x}_n \in \mathbb{R}^n$, $\|\mathbf{x}_n\| = 1$ if and only if $\mathsf{E}v_{11} = 0$, $\mathsf{E}v_{11}^2 = 1$, and $\mathsf{E}v_{11}^4 = 3$. It is proven by showing the mixed moments of the left side of (3.11) depends on the first, second and fourth moment of v_{11} after two sets of truncations and centralizations. After the final truncation and centralization the mixed moments are shown to be bounded regardless of the value of the fourth moment as long as it is finite. Thus after removing the $\sqrt{n}$ on the left side of (3.11) we find that the difference of the moments of the distribution $\mathbf{x}_n^T P^{M_n}(\cdot)\mathbf{x}_n$ and that of F_n, the empirical distribution of the eigenvalues of M_n, approach each other i.p. as $n \to \infty$. Since it is known that $F_n \to_D F_y$ a.s. from the method of moments we conclude that

$$\mathbf{x}_n^T P^{M_n}(\cdot)\mathbf{x}_n \to_D F_y \quad \text{i.p.}$$

With $\mathbf{x}_n = (1, 0, \dots, 0)^T$ we conclude that

$$P_{11}^{M_n}(\cdot) \to_D F_y \quad \text{i.p.} \tag{3.12}$$

The next results extends (3.11) to several different $\mathbf{x}_n$'s simultaneously.

THEOREM 3.4. Assume $\mathsf{E}v_{11} = 0$ and $\mathsf{E}v_{11}^2 = 1$. Fix d a positive integer. Let for every n $\mathbf{x}_n^1, \dots, \mathbf{x}_n^d$, $\mathbf{x}_n^j = (x_1^j, \dots, x_n^j)^T$, be d unit vectors in $\mathbb{R}^n$. Then the limiting distributional behavior of

$$\{\sqrt{n/2}(\mathbf{x}_n^{1\,T} M_n^r \mathbf{x}_n^1 - (1/n)\mathsf{tr}\,(M_n^r)), \dots, \sqrt{n/2}(\mathbf{x}_n^{d\,T} M_n^r \mathbf{x}_n^d - (1/n)\mathsf{tr}\,(M_n^r))\}_{r=1}^{\infty} \tag{3.13}$$

is the same as that when v_{11} is $N(0,1)$ if either:

a) $\mathsf{E}v_{11}^4 = 3$ or

b) for each $j \le d$

$$\sum_{i=1}^n (x_i^j)^4 \to 0, \qquad \text{as } n \to \infty.$$

Proof of a). By [15], through a series of truncations and centralizations, it is sufficient to assume that $v_{ij} = v_{ij,n}$ iid with $|v_{11}| \le 2n^{1/4}$, $\mathsf{E}v_{11} = 0$, $\mathsf{E}v_{11}^2 \to 1$, $\mathsf{E}v_{11}^4 \to 3$ as $n \to \infty$, and $(1/n)\mathsf{tr}\,M_n^r$ can be replaced by $\mathsf{E}\mathbf{x}_n^{i\,T} M_n^r \mathbf{x}_n^i$. We will use the method of moments. We will show for positive integers $m_1, \dots, m_d$, r_j^i, $i \le d$, $j \le m_i$, with $m = \sum_{i=1}^d m_i$, the limiting behavior of

$$\begin{aligned} n^{m/2}\mathsf{E}[(\mathbf{x}_n^{1\,T} M_n^{r_1^1}\mathbf{x}_n^1 - \mathsf{E}\mathbf{x}_n^{1\,T} M_n^{r_1^1}\mathbf{x}_n^1) \cdots (\mathbf{x}_n^{1\,T} M_n^{r_{m_1}^1}\mathbf{x}_n^1 - \mathsf{E}\mathbf{x}_n^{1\,T} M_n^{r_{m_1}^1}\mathbf{x}_n^1) \\ \cdots (\mathbf{x}_n^{d\,T} M^{r_1^d}\mathbf{x}_n^d - \mathsf{E}\mathbf{x}_n^{d\,T} M^{r_1^d}\mathbf{x}_n^d) \cdots (\mathbf{x}_n^{d\,T} M^{r_{m_d}^d}\mathbf{x}_n^d - \mathsf{E}\mathbf{x}_n^{d\,T} M^{r_{m_d}^d}\mathbf{x}_n^d)] \end{aligned} \tag{3.14}$$

depends only on $\mathsf{E}v_{11}^2$ and $\mathsf{E}v_{11}^4$ and therefore is the same when the original v_{ij}'s are $N(0,1)$.

Let $r = \sum_{i=1}^d \sum_{j=1}^{m_i} r_j^i$. We have

$(s^r/n^{m/2}) \times (3.14)$

$$
(3.15) \quad = \sum_{\substack{i^{11}, j^{11}, i_2^{11}, \dots, i^{11}_{r^1_1}, k_1^{11}, \dots, k^{11}_{r^1_1} \\ \vdots \\ i^{1m_1}, j^{1m_1}, i_2^{1m_1}, \dots, i^{1m_1}_{r^1_{m_1}}, k_1^{1m_1}, \dots, k^{1m_1}_{r^1_{m_1}} \\ \vdots \\ i^{d1}, j^{d1}, i_2^{d1}, \dots, i^{d1}_{r^d_1}, k_1^{d1}, \dots, k^{1}_{r^d_1} \\ \vdots \\ i^{dm_d}, j^{dm_d}, i_2^{dm_d}, \dots, i^{dm_d}_{r^d_{m_d}}, k_1^{dm_d}, \dots, k^{d\ m_d}_{r^d_{m_d}}}} x^1_{i^{11}} x^1_{j^{11}} \cdots x^1_{i^{1m_1}} x^1_{j^{1m_1}} \cdots x^d_{i^{d1}} x^d_{j^{d1}} \cdots x^d_{i^{dm_d}} x^d_{j^{dm_d}}
$$

$$
\mathsf{E}\big[\Pi_{\ell=1}^d \Pi_{\ell'=1}^{m_\ell} (v_{i^\ell k_1^{\ell\ell'}} v_{i_2^{\ell\ell'} k_1^{\ell\ell'}} \cdots v_{j^{\ell\ell'} k^{\ell\ell'}_{r^\ell_{\ell'}}} - \mathsf{E}(v_{i^{\ell\ell'} k_1^{\ell\ell'}} v_{i_2^{\ell\ell'} k_1^{\ell\ell'}} \cdots v_{j^{\ell\ell'} k^{\ell\ell'}_{r^\ell_{\ell'}}}))\big].
$$

Now the only difference between (3.14) here and (3.15) of [15] is that (3.15) in [15] involves only one unit vector whereas (3.14) here involves d unit vectors. The value $m = \sum m_i$ here, which is the total number of moments considered in (3.14), can be identified with the m in [15], the number of moments considered in (3.15) of [15]. The expected value in (3.15) here is essentially the same as the expected value in (3.16) in [15]. The dependence of the unit vector $\mathbf{x}_n$ in the argument presented in [15] is that the absolute value of the sum of its entries is bounded by $n^{1/2}$, its entries are bounded by 1 in absolute value, and its length is bounded. The argument here is identical to the one in [15] using the additional fact that $|\sum_{i=1}^n x_i^j x_i^k| \le 1$ for $j, k \in \{1, \dots, d\}$. We have then a).

Proof of b). The proof follows exactly the same as in the proof of Theorem 4.1 in [16] using the additional fact that for $j_1, \dots, j_4 \in \{1, \dots, d\}$

$$
\sum_{i=1}^n x_i^{j_1} x_i^{j_2} x_i^{j_3} x_i^{j_4} \le \max_{k \le 4} \sum_{i=1}^n {x_i^{j_k}}^4.
$$

This completes the proof of Theorem 3.4

Notice that

(3.16)

$$
\sqrt{n/2}({\mathbf{x}_{n,k}}^T M_n^r \mathbf{x}_{n,k} - (1/n)\operatorname{tr} M_n^r) = \int_0^\infty x^r dX_n^k(F_n(x)) = -\int_0^\infty r x^{r-1} X_n^k(F_n(x)) dx
$$

for $k \le m$, and for $j < k$

$$
(3.17) \qquad \sqrt{n}{\mathbf{x}_{n,j}}^T M_n^r \mathbf{x}_{n,k} = -\int_0^\infty r x^{r-1} Y_n^{jk}(F_n(x)) dx.
$$

When v_{11} is $N(0,1)$ we have from Theorem 1.2 the conclusion of Theorem 1.3. Therefore, from Theorem 5.1 of [3] the quantities in (3.16) and (3.17) converge weakly, together with the quantites

$$
(3.18) \qquad \sqrt{n/2}\left(\frac{(\mathbf{x}_{n,j} + \mathbf{x}_{n,k})^T}{\sqrt{2}} M_n^r \frac{(\mathbf{x}_{n,j} + \mathbf{x}_{n,k})}{\sqrt{2}} - (1/n)\operatorname{tr} M_n^r\right).
$$

since

$$(3.18) = \frac{1}{2}\sqrt{n/2}(\mathbf{x}_{n,j}^T M_n^r \mathbf{x}_{n,j} - (1/n)\mathrm{tr}\, M_n^r) + \frac{1}{2}\sqrt{n/2}(\mathbf{x}_{n,k}^T M_n^r \mathbf{x}_{n,k} - (1/n)\mathrm{tr}\, M_n^r) + \sqrt{n/2}\mathbf{x}_{n,j}^T M_n^r \mathbf{x}_{n,j}.$$

Therefore, when the $m(m+1)/2$ vectors $\mathbf{x}_{n,k}$ and $\frac{(\mathbf{x}_{n,j}+\mathbf{x}_{n,k})}{\sqrt{2}}$ are considered in Theorem 3.4 and either a) or b) hold then the quantities in (3.16) and (3.18) converge weakly to random variables having the same distribution as when v_{11} is $N(0,1)$. Since the quantity in (3.17) can be written as a linear combination of quantities in (3.16) and (3.18) we conclude that when a) or b) hold the quantities

$$\int_0^\infty x^r X_n^k(F_n(x))dx \quad k \le m, \qquad \int_0^\infty x^r Y_n^{jk}(F_n(x))dx \quad j < k$$

converge weakly to random variables, the same distribution as when v_{11} is N(0,1). Using (3.1) we have, when $b > (1+\sqrt{y})^2$

$$\int_0^b x^r X_n^k(F_n(x))dx \quad k \le m, \qquad \int_0^b x^r Y_n^{jk}(F_n(x))dx \quad j < k$$

converging weakly to variables with the same distribution as when v_{11} is $N(0,1)$. Therefore, we have (3.5) of Theorem 3.2. Under the assumptions of Theorem 3.3 we have (3.8), (3.9) , and (3.12), which can be used as in the last paragraph of [16] to show that the $Y_n^{jk}(F_n(\cdot))$ also satisfy the assumptions of Theorem 15.5 of [3]. Therefore, under the assumptions of Theorem 1.3, from Theorem 1.2 and Theorem 3.2, for each $b > (1+\sqrt{y})^2$ we have the $X_n^k(F_n(\cdot))$, $Y_n^{jk}(F_n(\cdot))$, $j < k$ all converging weakly in D_d^b to independent copies of Brownian bridge, composed with F_y, and hence the convergence is also on $D[0,\infty)$ for each of the processes . From Theorem 2.1 in [16] and Theorem 3.1 in this paper, we have the $X_n^k(\cdot)$, $Y_n^{jk}(\cdot)$ each converging weakly to Brownian bridge. The proof of Theorem 1.3 will follow once it is shown there is joint convergence to independent copies.

Notice that each of the limits $X_n^k(\cdot)$, $Y_n^{jk}(\cdot)$ reside in $C[0,1]$ and the limits $X_n^k(F_n(\cdot))$, $Y_n^{jk}(F_n(\cdot))$ in $C[0,\infty)$, where the topology in the latter is obtained from uniform convergence on $[0,b]$ for every $b > 0$. In fact the latter limits reside in the closed set

$$C' \equiv \{x \in C[0,\infty) : x(t) = x_0 \text{ for } t \in [0,(1-\sqrt{y})^2] \text{ and for some } x_0, 0 \text{ for } t \in [(1+\sqrt{y})^2,\infty)\}.$$

Consider first $y \le 1$. Then we can assume that there is one x_0 in C', namely 0. Let $\mathcal{C}^0$ denote the class of Borel sets in $C[0,1]$ and $\mathcal{C}'$ the class of Borel sets in C'. Define F_y^{-1} to be $(1-\sqrt{y})^2$ for $t=0$, $(1+\sqrt{y})^2$ for $t=1$, and $F_y^{-1}(t)$ for $t \in (0,1)$. It is straightforward to verify that the map $X(\cdot) \to X(F_y^{-1}(\cdot))$ from C' to $C[0,1]$ is continuous and is the inverse of $X(\cdot) \to X(F_y(\cdot))$ from $C[0,1]$ to C'. Let $\{W^{0k}_{F_y(\cdot)}, W^{0jk}_{F_y(\cdot)}, j < k \le m\}$ denote the weak limit of $\{X^k_n(F_n(\cdot)),\ Y^{jk}_n(F_n(\cdot)), j < k \le m\}$, where the entries of $\{W^{0k}_{(\cdot)}, W^{0jk}_{(\cdot)}, j < k \le m\} = \{W^{0k}_{F_y(F_y^{-1}(\cdot))}, W^{0jk}_{F_y(F_y^{-1}(\cdot))}, j < k \le m\}$ are independent copies of Brownian bridge. Let for $A \in \mathcal{C}'$ $F_y^{-1}(A) = \{X \in D[0,1] : X(F_y(\cdot)) \in A\}$ be the inverse image of A under F_y^{-1}. Then $F_y^{-1}(A) \in \mathcal{C}^0$. Suppose $A_k, A_{jk} \in \mathcal{C}'$ for $j < k \le m$. Then

$$
\begin{aligned}
(3.19)\quad & \mathsf{P}(W^{0k}_{F_y(\cdot)} \in A_k, W^{0jk}_{F_y(\cdot)} \in A_{jk}, j < k \le m) \\
&= \mathsf{P}(W^{0k}_{(\cdot)} \in F_y^{-1}(A_k), W^{0jk}_{(\cdot)} \in F_y^{-1}(A_{jk}), j < k \le m) \\
&= \Pi_k P(W^{0k}_{(\cdot)} \in F_y^{-1}(A_k)) \times \Pi_{j<k} \mathsf{P}(W^{0jk}_{(\cdot)} \in F_y^{-1}(A_{jk})) \\
&= \Pi_k \mathsf{P}(W^{0k}_{F_y(\cdot)} \in A_k) \times \Pi_{j<k} \mathsf{P}(W^{0jk}_{F_y(\cdot)} \in A_{jk}).
\end{aligned}
$$

Therefore the entries of $\{W^{0k}_{F_y(\cdot)}, W^{0jk}_{F_y(\cdot)}, j < k \le m\}$ are independent.

Using the same argument used in Lemma 2.3, the sequence $\{X^k_n, Y^{jk}_n, j < k\}_{n=1}^\infty$ is tight. Suppose on some subsequence $\{X^k_n, Y^{jk}_n, j < k \le m\}$ converges weakly to the random element $\{W^{0k}, W^{0jk}, j < k \le m\}$ in D^1_d. Then each entry is Brownian bridge and the entries of $\{W^{0k}_{F_y(\cdot)}, W^{0jk}_{F_y(\cdot)}, j < k \le m\}$ are independent. We invoke Theorem 8.3.7 of [6]: Let X and Y be Polish spaces (separable and can be metrized with a complete metric), let A be a Borel subset of X, and let $f : A \to Y$ be Borel measurable and injective (1-to-1). Then $f(A)$ is a Borel subset of Y.

Therefore, with $F_y(A)$ denoting the image of A under F_y, for sets $A_k, A_{jk} \in \mathcal{C}^0$ we have $F_y(A_k), F_y(A_{jk}) \in \mathcal{C}'$ and

$$
\begin{aligned}
(3.20)\quad & \mathsf{P}(W^{0k} \in A_k, W^{0jk} \in A_{jk}) = \mathsf{P}(W^{0k}_{F_y(\cdot)} \in F_y(A_k), W^{0jk}_{F_y(\cdot)} \in F_y(A_{jk})) \\
&= \Pi_k \mathsf{P}(W^{0k}_{F_y(\cdot)} \in F_y(A_k)) \times \Pi_{j<k} \mathsf{P}(W^{0jk}_{F_y(\cdot)} \in F_y(A_{jk})) \\
&= \Pi_k \mathsf{P}(W^{0k} \in A_k) \times \Pi_{j<k} \mathsf{P}(W^{0jk} \in A_{jk}).
\end{aligned}
$$

Therefore the W^{0k}, W^{0jk} are independent and we have Theorem 1.3 in this case.

For $y > 1$ we express the processes in the form of a matrix. Let W_n denote the $m \times m$ matrix with $W_{nkk} = X^k_n$, and for $j < k$ $W_{njk} = W_{nkj} = Y^{jk}_n$. Let $O_{n,1}$ and O_1 as in Theorem 3.1. Let $\varphi_n(t)$ be as in Theorem 3.1 with $\varphi(t) = \min(t/(1 - 1/y)), 1)$ as its a.s. limit. Let $\psi_n(t) = \max(t, F_n(0))$ with $\psi(t) \equiv \max(t, 1 - 1/y)$ as its a.s. limit. Let B_m be the $m \times m$ matrix consisting of $1/\sqrt{2}$'s on its diagonal and 1's on its off-diagonal elements. Let $\underline{X}_m$ be the $m_1 \times m$ matrix with i-th column $O^T_{n,1}\mathbf{x}_{n,i}$, let $I_{m_1,s}$ be the $m_1 \times m_1$ diagonal matrix consisting of 1's on its first s diagonal entries, 0 on the remaining diagonal entries, and let I_{m_1} be the $m_1 \times m_1$ identity matrix. Notice that $m_1 = nF_n(0)$. Denote "$\circ$" as the Hadamard product. Then we have

$$
W_n(t) = \sqrt{n} B_m \circ \left(\underline{X}^T_m O_1 I_{m_1,[m_1\varphi_n(t)]} O^T_1 \underline{X}_m - \frac{[m_1\varphi_n(t)]}{n} I_{m_1} \right) - W_n(F_n(0)) + W_n(\psi_n(t)).
$$

Let W' be the weak limit of W_n on a subsequence. Then on this subsequence $W_n(\psi_n(\cdot)) \to_D W'(\psi(\cdot))$, and $W_n(\psi_n(F_n(\cdot))) = W_n(F_n(\cdot)) \to_D W'_{F_{y(\cdot)}}$, where the entries of $W'_{F_{y(\cdot)}}$ on and above the diagonal are independent copies of Brownian bridge, composed with F_y. Confining to the interval $[1-1/y, 1]$ these entries will also be independent copies on $C[1-1/y, 1]$. If we define F_y^{-1} just on $[1-1/y, 1]$ we have for $X \in C'$ $X(F_y^{-1}(F_y)) = X$. Therefore from (3.19) we see that the entries on and above the diagonal of $W'_{F_y(\cdot)}$ are independent. For $X, Y \in C[1-1/y, 1]$ $X \neq Y$ we have $X(F_y(\cdot)) \neq Y(F_y(\cdot))$ so that the 1-1 condition of Theorem 8.3.7 of [6] is satisfied. We also have $X(F_y(F_y^{-1})) = X$. Therefore we have from (3.20) with the entries of W' confined to $[1-1/y, 1]$ and the sets Borel subsets of $C[1-1/y, 1]$, the entries of W' on $[1-1/y, 1]$ on and above the diagonal are independent. This uniquely determines the limiting distribution, so we see that $W_n(\psi_n(\cdot)) \to_D W^0(\psi(\cdot))$, where W^0 is Brownian bridge, with entries on and above the diagonal independent.

Let $\underline{X}_m = U_m R_m$ be the QR factorization of $\underline{X}_m$, where the columns of U_m are orthonormal, and R_m is $m \times m$ upper triangular, with nonnegative diagonal entries. Extending (3.3) and (3.4) to all columns of $\underline{X}_m$ we have

$$R_m^T R_m = \underline{X}_m^T \underline{X}_m \xrightarrow{i.p.} (1-(1/y))I_m.$$

From this it is straightforward to prove

$$R_m \xrightarrow{i.p.} \sqrt{1-(1/y)}I_m. \tag{3.21}$$

Write

$$\sqrt{n}B_m \circ \left(\underline{X}_m^T O_1 I_{m_1,[m_1\varphi_n(t)]} O_1^T \underline{X}_m - \frac{[m_1\varphi_n(t)]}{n} I_{m_1}\right) - W_n(F_n(0)) \tag{3.22}$$

$$= \frac{1}{\sqrt{F_n(0)}} B_m \circ R_m^T \sqrt{m_1}\left(U_m^T O_1 I_{m_1,[m_1\varphi_n(t)]} O_1^T U_m - \frac{[m_1\varphi_n(t)]}{m_1} I_{m_1}\right) R_m$$
$$+ W_n(F_n(0))\left(\frac{[m_1\varphi_n(t)]}{m_1} - 1\right).$$

As in Theorem 2.1 of [16], we use Theorem 5.1 of [3] applied to

$$\left(W_n, \sqrt{m_1}\left(U_m^T O_1 I_{m_1,[m_1 s]} O_1^T U_m - \frac{[m_1 s]}{m_1} I_{m_1}\right), R_m, F_n(0), \varphi_n, \psi_n\right).$$

We also apply Lemma 3.2 where $X'_n = (W_n, F_n(0))$, $\ell_n = m_1$, and X''_{ℓ_n} is the second component of the above six-tuple. Therefore, from Theorem 1.2, (3.21), and (3.22) we have

$$W_n \to_D \sqrt{1-(1/y)}\hat{W}^0_\varphi + W^0_{1-(1/y)}(\varphi - 1) + W^0_\psi,$$

where $\hat{W}^0$ is an independent copy of W^0. Since this limit is the same when v_{11} is $N(0,1)$ we have this limit having independent elements on and above the diagonal. This completes the proof of Theorem 1.3.

4. Proof of Theorem 1.4 We first need the following:

Lemma 5 (Lemma 2.7 in Bai and Silverstein (1998). For $X = (X_1, \ldots, X_n)^T$ i.i.d. standardized entries, and C, an $n \times n$ matrix, we have, for any $p \geq 2$

$$\mathsf{E}|X^*CX - \operatorname{tr} C|^p \leq K_p((\mathsf{E}|X_1|^4 \operatorname{tr} CC^*)^{p/2} + \mathsf{E}|X_1|^{2p} \operatorname{tr}(CC^*)^{p/2}).$$

Suppose C, $n \times n$, is bounded in spectral norm and X contains i.i.d. complex Gaussian entries. Then for any $p \geq 2$

$$\mathsf{E}|X^*CX - \operatorname{tr} C|^p \leq K_p \|C\|^p ((\mathsf{E}|X_1|^4)^{p/2} n^{p/2} + \mathsf{E}|X_1|^{2p} n) \leq K_p n^{p/2}. \tag{4.1}$$

Recalling $S_n = U_n \Lambda_n U_n^*$ in its spectral decomposition with eigenvalues arranged in nondecreasing order, for any real x let $\Lambda_n(x)$ denote the diagonal matrix containing $nF_n(x)$ one's on the upper part of its diagonal. Therefore $F_n(x) = (1/n)\operatorname{tr} \Lambda_n(x)$. Notice that $G_n(x) = \mathbf{v}_n^* U_n \Lambda_n(x) U_n^* \mathbf{v}_n = \sum_{\lambda_k \leq x} |\mathbf{u}_k^* \mathbf{v}_n|^2$ where $U_n = (\mathbf{u}_1, \ldots, \mathbf{u}_n)$, is the distribution function of a random variable which takes values $\lambda_1, \ldots, \lambda_n$ (eigenvalues of S_n) with probabilities $|\mathbf{u}_1^* \mathbf{v}_n|^2, \ldots, |\mathbf{u}_n^* \mathbf{v}_n|^2$. Now, since $U_n^* \mathbf{v}_n$ is uniformly distributed on the n dimensional unit sphere in $\mathbb{C}^n$ it has the distribution of a normalized vector $\mathbf{z}_n$ of n i.i.d. complex Gaussian entries: $U^* \mathbf{v}_n \sim (1/\|\mathbf{z}_n\|)\mathbf{z}_n$. By (4.1) we have

$$\mathsf{E}|(1/n)\mathbf{z}_n^* \Lambda_n(x) \mathbf{z}_n - F_n(x)|^4 \leq K n^{-2}.$$

Moreover

$$|G_n(x) - (1/n)\mathbf{z}_n^* \Lambda_n(x) \mathbf{z}_n| = (1/n)\mathbf{z}_n^* \Lambda_n(x) \mathbf{z}_n |n/\|\mathbf{z}_n\|^2 - 1| \xrightarrow{a.s.} 0$$

by the strong law of large numbers. Therefore we have with probability one, G_n converges in distribution to F, and the largest value in the support of G_n, namely the largest eigenvalue of S_n, converges with probability one to $\lambda_{\max}$. Therefore for any $\lambda > \lambda_{\max}$ with probability one, for all n large $(\mathbf{v}_n^*(\lambda I - S_n)^{-1}\mathbf{v}_n, \mathbf{v}_n^*(\lambda I - S_n)^{-2}\mathbf{v}_n)$ exists and converges to $(\int(\lambda - x)^{-1} dF(x), (\int(\lambda - x)^{-2} dF(x))$.

Suppose that for all $\lambda > \lambda_{\max}$ $\int(\lambda - x)^{-1} dF(x) \leq 1/\theta$. Then necessarily $\lim_{\lambda \to \lambda_{\max}^+} \int(\lambda - x)^{-1} dF(x) \leq 1/\theta$, which means for all $\epsilon > 0$ $\int(\lambda_{\max} + \epsilon - x)^{-1} dF(x) < 1/\theta$ Since almost surely $\mathbf{v}_n^*((\lambda_{\max} + \epsilon)I - S_n)^{-1}\mathbf{v}_n \to \int(\lambda_{\max} + \epsilon - x)^{-1} dF(x)$, we must have with probability one, for all n large $\lambda_n^1 < \lambda_{\max} + \epsilon$. Since ϵ is arbitrary we must have almost surely $\lambda_n^1 \to \lambda_{\max}$.

Suppose now there exists $\lambda > \lambda_{\max}$ such that $\int(\lambda - x)^{-1}dF(x) > 1/\theta$ Then let $\lambda_1 > \lambda_{\max}$ be the unique value such that $\int(\lambda_1 - x)^{-1}dF(x) = 1/\theta$. For small $\epsilon > 0$

$$\int(\lambda_1 - \epsilon - x)^{-1}dF(x) > 1/\theta \quad \text{and} \quad \int(\lambda_1 + \epsilon - x)^{-1}dF(x) < 1/\theta.$$

Since almost surely

$$\mathbf{v}_n^*((\lambda_1 - \epsilon)I - S_n)^{-1}\mathbf{v}_n \to \int(\lambda_1 - \epsilon - x)^{-1}dF(x)$$
$$\text{and} \quad \mathbf{v}_n^*((\lambda_1 + \epsilon)I - S_n)^{-1}\mathbf{v}_n \to \int(\lambda_1 + \epsilon - x)^{-1}dF(x)$$

we have almost surely for all n large $\lambda_1 - \epsilon < \lambda_n^1 < \lambda_1 + \epsilon$. Since ϵ is arbitrary we must have $\lambda_n^1 \xrightarrow{a.s.} \lambda_1$.

For small $\epsilon > 0$ we have with probability one, for all n large

$$\mathbf{v}_n^*((\lambda_1 + \epsilon)I - S_n)^{-2}\mathbf{v}_n \le \mathbf{v}_n^*(\lambda_n^1 I - S_n)^{-2}\mathbf{v}_n \le \mathbf{v}_n^*((\lambda_1 - \epsilon)I - S_n)^{-2}\mathbf{v}_n.$$

where the extremes approach almost surely $\int(\lambda_1 + \epsilon - x)^{-2}dF(x)$, $\int(\lambda_1 i\epsilon - x)^{-2}dF(x)$, respectively. Since ϵ we have

$$\mathbf{v}_n^*(\lambda_n^1 I - S_n)^{-2}\mathbf{v}_n \xrightarrow{a.s.} \int(\lambda_1 - x)^{-2}dF(x),$$

which gives us (1.12).

Let $b \in (\lambda_{\max}, \lambda_1)$ and $a = (\lambda_{\max} + b)/2$. Select $d > \lambda_1$. Define for $t \in [b, d]$ $\Phi_n(t) \equiv b$, if $\lambda_n^1 \notin [b, d]$ and $\equiv \lambda_n^1$ if $\lambda_n^1 \in [b, d]$. Then Φ_n is a random element in $D_0[b, d]$, those elements of $D[b, d]$ whose range is also in $[b, d]$ and nondecreasing (pp. 144-145 of [3]). Then with probability one, for all n large, $\Phi_n \equiv \lambda_n^1$ and converges to λ_1.

Identify $\mathbf{v}_n$ with $\mathbf{x}_{n,k'}$ in (1.4). Define $X_n^k(x) = X_n^{k,k'}(F_n(x))$. We have X_n^k a random element in $D[0, \infty)$, the set of all functions on $[0, \infty)$ having discontinuities of the first kind ([9]). It is straightforward to extend the material in [3] pp. 144-145 and Theorem 4.4 to bounded nondecreasing functions in $D[0, \infty)$ to conclude that $X_n^k(x)$ converges weakly to

$$W_{k,r}^0(F(x)) + iW_{k,i}^0(F(x)) \tag{4.2}$$

on $D_2[0, \infty)$ (two copies of $D[0, \infty)$) (Note: this is the only place where we need the limiting distribution function F to be continuous).

Let for $x \in [0, a]$

$$Y_n^k(x) = I_{\{\lambda_{\max}(S_n) \le a\}} X_n^k(x),$$

where I_A is the indicator function on the set A. Then from Theorem 4.1 of [3] Y_n^k converges weakly to (4.2) on $D_2[0,a]$ (two copies of $D[0,a]$).

Define the mapping f from $D_2[0,a]$ to $C_2[b,d]$ (two copies of $C[b,d]$, the space of continuous functions on $[b,d]$) by

$$f(X) = -\int_0^a (t-x)^{-2} X(x)dx \quad t \in [b,d].$$

Then

$$f(Y_n^k) = -I_{\{\lambda_{\max}(S_n)\le a\}} \int_0^a (t-x)^{-2} X_n^k(x)dx$$
$$I_{\{\lambda_{\max}(S_n)\le a\}} \int_0^a (t-x)^{-1} dX_n^k(x) = I_{\{\lambda_{\max}(S_n)\le a\}} \sqrt{2n}\mathbf{x}_{n,k}^*(tI - S_n)^{-1}\mathbf{v}_n.$$

We claim that f is a continuous mapping. Suppose $X_n \to X$ in $D_2[0,a]$ in the Skorohod topology. Then $X_n(s) \to X(s)$ for continuity points s of X, and because X lies in $D_2[0,a]$, this set is outside a set of Lebesgue measure 0. Using the fact that convergence in the Skorohod topology renders the X_n and X uniformly bounded we have by the dominated convergence theorem

$$|f(X_n) - f(X)| \le ((b-\lambda_{\max})/2)^2 \int_0^a |X_n(x) - X(x)|dx \to 0,$$

uniformly for $t \in [b,d]$. Therefore f is continuous.

Therefore from Theorem 5.1 of [3] we have

$$I_{\{\lambda_{\max}(S_n)\le a\}} \sqrt{2n}\mathbf{x}_{n,k}^*(tI - S_n)^{-1}\mathbf{v}_n$$
$$\to_D \int (t-x)^{-1} dW_{k,r}^0(F(x)) + i\int (t-x)^{-1} dW_{k,i}^0(F(x))$$

on $D_2[b,d]$. From the material on pp.144-145 of [3] we have

$$I_{\{\lambda_{\max}(S_n)\le a\}} \sqrt{2n}\mathbf{x}_{n,k}^*(\Phi_n I - S_n)^{-1}\mathbf{v}_n$$
$$\to_D \int (\lambda_1 - x)^{-1} dW_{k,r}^0(F(x)) + i\int (\lambda_1 - x)^{-1} dW_{k,i}^0(F(x)).$$

Using again Theorem 4.1 of [3] we get (1.10).

We get the same result for G_n in the real Gaussian case. For the matrix M_n it is proven in Section 3 that $G_n \to_D F_y$ i.p. For the former the steps above follow identically, resulting in (1.13). For the latter, since the finite result is distributional in nature we may as well assume $G_n \to_D F_y$ a.s. (since this is true on an appropriate subsequence of an arbitrary subsequence of natural numbers). Thus we get (1.13) with $F = F_y$.

REFERENCES

[1] Baik, J., Ben Arous, G., and Péché. (2005) Phase transition of the largest eigenvalue for non-null complex sample covariance matrices. *Ann. Probab.* **33** 1643-1697.

[2] Baik, J, and Silverstein, J.W. (2006) Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal.* **97** 1382-1408

[3] Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

[4] Billingsley, P. (1995) *Probability and Measure* Third Edition. Wiley, New York.

[5] Billingsley, P. (1999). *Convergence of Probability Measures* Second Edition. Wiley, New York.

[6] Cohn, D.L. (1980) *Measure Theory* Birkhauser Boston.

[7] Grenander, U. and Silverstein, J.W. (1977). Spectral analysis of networks with random topologies. *SIAM J. Appl. Math.* **37** 499-519.

[8] Jonsson, D. (1982). Some limit theorems for the eigenvalues of a sample covariance matrix. *J. Multivariate Anal.* **12** 1-38.

[9] Lindvall, T. (1973). Weak convergence of probability measures and random functions in the function space $D[0,\infty))$. *J. Appl. Probab.* **10** 109-121.

[10] Marčenko, V.A. and Pastur, L.A. (1967). Distribution of eigenvalues for some sets of random matrices. *Math. USSR-Sb.* **1**, 457-483.

[11] Rao, R. and Silverstein, J.W. (2010) Fundamental limit of sample generalized eigenvalue based detection of signals in noise using relatively few signal-bearing and noise-only samples *IEEE Journal of Selected Topics in Signal Processing* **3** 468-480.

[12] Silverstein, J.W. (1979). On the randomnes of eigenvectors generated from networks with random topologies. *SIAM J. Appl. Math.* **37** 235-245.

[13] Silverstein, J.W. (1981). Describing the behavior of random matrices using sequences of measures on orthogonal groups. *SIAM J. Math. Anal.* **12** 274-281.

[14] Silverstein, J.W. (1984). Some limit theorems on the eigenvectors of large dimensional sample covariance matrices. *J. Multivariate Anal.* **15** 295-324.

[15] Silverstein, J.W. (1989). On the eigenvectors of large dimensional sample covariance matrices. *J. Multivariate Anal.* **30** 1-16.

[16] Silverstein, J.W. (1990). Weak convergence of random functions defined by the eigenvectors of sample covariance matrices. *Ann. Probab.* **18** 1174-1194.

[17] Silverstein, J.W. (1995). Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices *J. Multivariate Anal.* **55** 331-339.

[18] Wachter, K.W. (1978). The strong limits of random matrix spectra for sample matrices of independent elements. *Ann. Probab.* **6** 1-18.

[19] Yin, Y.Q. (1986). Limiting spectral distribution for a class of random matrices. *J. Multivariate Anal.* **20** 50-68.

[20] Yin, Y.Q., Bai, Z.D., and Krishnaiah, P.R. (1988). On limit of the largest eigenvalue of the large dimensional sample covariance matrix. *Probab. Th. Rel. Fields* **78** 509-521.