Weak Convergence of Random Functions Defined by the Eigenvectors of Sample Covariance Matrices

by

Jack W. Silverstein*
North Carolina State University

Summary

Let $\{v_{ij}\}$, $i,j = 1,2,\ldots$, be i.i.d. symmetric random variables with $\mathsf{E}(v_{11}^4) < \infty$, 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 $n \to \infty$. Denote by $O_n\Lambda_nO_n^T$ the spectral decomposition of M_n. Define $X \in D[0,1]$ by $X_n(t) = \sqrt{\frac{n}{2}}\sum_{i=1}^{[nt]}(y_i^2 - \frac{1}{n})$, where $(y_1, y_2, \ldots, y_n)^T = O^T(\pm\frac{1}{\sqrt{n}}, \pm\frac{1}{\sqrt{n}}, \ldots, \pm\frac{1}{\sqrt{n}})^T$. It is shown that $X_n \xrightarrow{\mathcal{D}} W^\circ$ as $n \to \infty$, where W° is Brownian bridge. This result sheds some light on the problem of describing the behavior of the eigenvectors of M_n for n large and for general v_{11}.

* Supported by the National Science Foundation under grants MCS-8101703-A01 and DMS-8603966

AMS 1980 subject classifications. Primary 60F05; Secondary 62H99.

Key Words and Phrases. Weak convergence on $D[0,1]$ and $D[0,\infty)$, eigenvectors of sample covariance matrix, Brownian bridge, Haar measure, signed measures.

1. Introduction. Let $\{v_{ij}\}$, $i,j=1,2,\ldots$, be i.i.d. random variables with $\mathsf{E}(v_{11})=0$, 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 $n\to\infty$. The symmetric, nonnegative definite matrix M_n can be viewed as the sample covariance matrix of s samples of an n dimensional random vector having i.i.d. components distributed the same as v_{11} (assuming knowledge of the common mean being zero). The spectral behavior of M_n for n large is important to areas of multivariate analysis (including principal component analysis, regression, and signal processing) where n and s are the same order of magnitude, so that standard asymptotic analysis cannot be applied. Although eigenvalue results will be discussed, this paper is chiefly concerned with the behavior of the eigenvectors of M_n. It continues the analysis begun in [8].

Throughout the following, $O_n\Lambda_nO_n^T$ will denote the spectral decomposition of M_n, where the eigenvalues of M_n are arranged along the diagonal of Λ_n in nondecreasing order. The orthogonal matrix O_n, the columns being eigenvectors of M_n, is not uniquely determined, owing to the multiplicities of the eigenvalues and the direction any eigenvector can take on. This problem will be addressed later on. For now it is sufficient to mention that, by appropriately enlarging the sample space where the v_{ij}'s are defined, it is possible to construct O_n measurable in a natural manner from the eigenspaces associated with M_n.

Results previously obtained suggest similarity of behavior of the eigenvectors of M_n for large n to the eigenvectors of matrices of Wishart type, that is, when v_{11} is normally distributed ([8],[9],[10],[11]). In this case it is well-known that O_n induces the Haar (uniform) measure on $\mathcal{O}_n$, the $n\times n$ orthogonal group. In [9] it is conjectured that, for general v_{11} and for n large, O_n is somehow close to being Haar distributed. The attempt to make the notion of closeness more precise has led in [9] to an investigation into the behavior of random elements, X_n, of $D[0,1]$ (the space of *r.c.l.l.* functions on [0,1]) defined by the eigenvectors. They are constructed as follows:

For each n let $\vec{x}_n\in R^n$, $\|x_n\|=1$, be nonrandom, and let $\vec{y}_n=(y_1,y_2,\ldots,y_n)^T=O_n^T\vec{x}_n$. Then, for $t\in[0,1]$

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

The importance of X_n to understanding the behavior of the eigenvectors of M_n stems from three facts. First, the behavior of X_n for all $\vec{x}_n$ reflects to some degree the uniformity or nonuniformity of O_n. If O_n were Haar distributed, then $\vec{y}_n$ would be uniformly distributed over the unit sphere in R^n, rendering an identifiable distribution for X_n, invariant across unit $\vec{x}_n\in R^n$. On the other hand, significant departure from Haar measure would be suspected if the distribution of X_n depended strongly on $\vec{x}_n$. Second, we are able to compare some aspects of the distribution of O_n for all n on a common space, namely $D[0,1]$. Third, the limiting behavior of X_n is known when O_n is Haar distributed. Indeed, in this case the distribution of $\vec{y}_n$, being uniform on the unit sphere in R^n, is the same as that of a normalized vector of i.i.d. mean-zero Gaussian components. Upon applying standard results on weak convergence of measures, it is straightforward to show

$$X_n \xrightarrow{\mathcal{D}} W^\circ \quad \text{as } n \to \infty \tag{1.1}$$

($\mathcal{D}$ denoting weak convergence in $D[0,1]$) where W° is Brownian bridge ([9]). Thus, for arbitrary v_{11}, (1.1) holding for all $\{\vec{x}_n\}$, $\|x_n\| = 1$, can be viewed as evidence supporting the conjecture that O_n is close to being uniformly distributed in $\mathcal{O}_n$ for n large.

Verifying (1.1) more generally is difficult since there is very little useful direct information available on the variables $y_1^2, \ldots, y_n^2$. However, under the assumption $\mathsf{E}(v_{11}^4) < \infty$, the results in [9],[10],[11] reduce the problem to verifying tightness of $\{X_n\}$. For the following we may, without loss of generality, assume $\mathsf{E}(v_{11}^2) = 1$. The results are limit theorems on random variables defined by $\{M_n\}$. From one of the theorems ([11], to be given below) it follows that for $\mathsf{E}(v_{11}^4) = 3$, any weakly convergent subsequence of $\{X_n\}$ converges to W° for any sequence $\{\vec{x}_n\}$ of unit vectors, while if $\mathsf{E}(v_{11}^4) \neq 3$, there exists sequences $\{\vec{x}_n\}$ for which $\{X_n\}$ fails to converge weakly. This suggests some further similarity of v_{11} to $N(0,1)$ may be necessary. It is remarked here that the theorem also implies, for finite $\mathsf{E}(v_{11}^4)$, the necessity of $\mathsf{E}(v_{11}) = 0$ in order for (1.1) to hold for $\vec{x}_n = (1, 0, \ldots, 0)^T$.

The main purpose of this paper is to establish the following partial solution to the problem:

THEOREM 1.1. *Assume v_{11} is symmetric (that is, symmetrically distributed about 0), and $\mathsf{E}(v_{11}^4) < \infty$. Then (1.1) holds for $\vec{x}_n = (\pm\frac{1}{\sqrt{n}}, \pm\frac{1}{\sqrt{n}}, \ldots, \pm\frac{1}{\sqrt{n}})^T$.*

From the theorem one can easily argue other choices of $\vec{x}_n$ for which (1.1) holds, namely vectors close enough to those in the theorem so that the resulting X_n approaches in the Skorohod metric random functions satisfying (1.1). It will become apparent that the techniques used in the proof of Theorem 1.1 cannot easily be extended to $\vec{x}_n$ having more variability in the magnitude of its components, while the symmetry requirement may be weakened with a deeper analysis. At present the possibility exists that only for v_{11} mean-zero Gaussian will (1.1) be satisfied for all $\{\vec{x}_n\}$.

However, from Theorem 1.1 and the previously mentioned results emerges the possibility of classifying the distribution of O_n into varying degrees of closeness to Haar measure. The eigenvectors of M_n with v_{11} symmetric and fourth moment finite display a certain amount of uniform behavior, and O_n can possibly be even more closely related to Haar measure if $\mathsf{E}(v_{11}^4)/[\mathsf{E}(v_{11}^2)]^2 = 3$. As will be seen below when it is formally stated, the limit theorem in [11] itself can be viewed as demonstrating varying degrees of similarity to Haar measure.

The proof of Theorem 1.1 relies on two results on the eigenvalues of M_n and on a modification of the limit theorem in [11]. Let F_n denote the empirical distribution function of the eigenvalues of M_n (that is, $F_n(x) = (1/n) \times$ (number of eigenvalues of $M_n \leq x$), where we may as well assume $x \geq 0$). If $\mathsf{Var}(v_{11}) = 1$ (no other assumption on the moments), then it is known ([4],[5],[12],[13]) that, for every $x \geq 0$,

$$F_n(x) \xrightarrow{a.s.} F_y(x) \quad \text{as } n \to \infty, \tag{1.2}$$

where F_y is a continuous, nonrandom probability distribution function depending only on y, having a density with support on $[(1-\sqrt{y})^2, (1+\sqrt{y})^2]$, and for $y > 1$, F_y places mass $1 - 1/y$ at 0. Moreover, if $\mathsf{E}(v_{11}) = 0$, $\mathsf{E}(v_{11}^2) = 1$, then $\lambda_{\max}(M_n)$, the largest eigenvalue of M_n, satisfies

$$\lambda_{\max}(M_n) \xrightarrow{a.s.} (1+\sqrt{y})^2 \quad \text{as } n \to \infty \tag{1.3}$$

if and only if $\mathsf{E}(v_{11}^4) < \infty$ ([1],[3],[14]).

For the theorem in [11], we first make the following observation. It is straightforward to show that (1.1),(1.2),(1.3) imply

$$X_n(F_n(x)) \xrightarrow{\mathcal{D}} W_x^y \equiv W^\circ(F_y(x)) \tag{1.4}$$

on $D[0,\infty)$. The proof of Theorem 1.1 essentially verifies the truth of the implication in the other direction and then the truth of (1.4). An extension of the theorem in [11] is needed for the latter to establish the uniqueness of any weakly convergent subsequence. The theorem states:

$$\left\{ \sqrt{\tfrac{n}{2}}(\vec{x}_n^T M_n^r \vec{x}_n - \tfrac{1}{n}\mathrm{tr}(M_n^r)) \right\}_{r=1}^{\infty} = \left\{ \int_0^\infty x^r \, dX_n(F_n(x)) \right\}_{r=1}^{\infty} \tag{1.5}$$

$$\left\{ -\int_0^\infty r x^{r-1} \, X_n(F_n(x)) dx \right\}_{r=1}^{\infty} \xrightarrow{\mathcal{D}} \left\{ \int_{(1-\sqrt{y})^2}^{(1+\sqrt{y})^2} x^r \, dW_x^y \right\}_{r=1}^{\infty} \quad \text{as } n \to \infty$$

($\mathcal{D}$ denoting weak convergence on R^∞) for every sequence $\{\vec{x}_n\}$ of unit vectors if and only if $\mathsf{E}(v_{11}) = 0$, $\mathsf{E}(v_{11}^2) = 1$, and $\mathsf{E}(v_{11}^4) = 3$ (we remark here that the limiting random variables in (1.5) are well defined stochastic integrals, being jointly normal each with mean 0). The proof of this theorem will be modified to show (1.5) still holds under the assumptions of Theorem 1.1 (without a condition on the fourth moment of v_{11} other than it being finite).

The proof will be carried out in the next three sections. Section 2 presents a formal description of O_n to account for the ambiguities mentioned earlier, followed by a result which converts the problem to one of showing weak convergence of $X_n(F_n(\cdot))$ on $D[0,\infty)$. Section 3 contains results on random elements in $D[0,b]$ for any $b>0$, which are extensions of certain criteria for weak convergence given in [2]. In section 4 the proof is completed by showing the conditions in section 3 are met. Some of the results will be stated more generally than presently needed to render them applicable for future use.

2. Converting to $D[0,\infty)$. Let us first give a more detailed description of the distribution of O_n which will lead us to a concrete construction of $\vec{y}_n$. For an eigenvalue λ with multiplicity r we assume the corresponding r columns of O_n to be generated uniformly, that is, its distribution is the same as $O_{n,r}O_r$ where $O_{n,r}$ is $n\times r$ containing r orthonormal columns from the eigenspace of λ, and $O_r\in\mathcal{O}_r$ is Haar distributed, independent of M_n. The O_r's corresponding to distinct eigenvalues are also assumed to be independent. The coordinates of $\vec{y}_n$ corresponding to λ are then of the form

$$(O_{n,r}O_r)^T\vec{x}_n = O_r^T O_{n,r}^T\vec{x}_n = \|O_{n,r}^T\vec{x}_n\|\vec{w}_r$$

where $\vec{w}_r$ is uniformly distributed on the unit sphere in R^r. We will use the fact that the distribution of $\vec{w}_r$ is the same as that of a normalized vector of i.i.d. mean zero Gaussian components. Notice that $\|O_{n,r}^T\vec{x}_n\|$ is the length of the projection of x_n on the eigenspace of λ.

Thus, $\vec{y}_n$ can be represented as follows:

Enlarge the sample space defining M_n to allow the construction of $z_1,z_2,\ldots,z_n$, i.i.d. N(0,1) random variables independent of M_n. For a given M_n let $\lambda_{(1)}<\lambda_{(2)}<\cdots<\lambda_{(t)}$ be the t distinct eigenvalues with multiplicities $m_1,m_2,\ldots,m_t$. For $i=1,2,\ldots,t$ let a_i be the length of the projection of $\vec{x}_n$ on the eigenspace of $\lambda_{(i)}$. Define $m_0=0$. Then, for each i we define the coordinates

$$(y_{m_1+\cdots+m_{i-1}+1}, y_{m_1+\cdots+m_{i-1}+2},\cdots,y_{m_1+\cdots+m_i})$$

of $\vec{y}_n$ to be the respective coordinates of

$$a_i(z_{m_1+\cdots+m_{i-1}+1}, z_{m_1+\cdots+m_{i-1}+2},\cdots,z_{m_1+\cdots+m_i})\Big/\sqrt{\sum_{k=1}^{m_i} z^2_{m_1+\cdots+m_{i-1}+k}}\,. \tag{2.1}$$

We are now in a position to prove

THEOREM 2.1. *$X_n(F_n(\cdot))\xrightarrow{\mathcal{D}} W^\circ_{F_y(\cdot)}$ in $D[0,\infty)$, $F_n(x)\xrightarrow{i.p.}F_y(x)$, and $\lambda_{\max}\xrightarrow{i.p.}(1+\sqrt{y})^2\implies X_n\xrightarrow{\mathcal{D}}W^\circ$.*

PROOF. We assume the reader is familiar with the basic results in [2] of showing weak convergence of random elements of a metric space (most notably Theorems 4.1, 4.4, and

Corollary 1 to Theorem 5.1), in particular, the results on the function spaces $D[0,1]$ and $C[0,1]$. For the topology and conditions of weak convergence in $D[0,\infty)$ see [6]. For our purposes, the only information needed regarding $D[0,\infty)$ beyond that of [2] is the fact that weak convergence of a sequence of random functions on $D[0,\infty)$ is equivalent to the following: for every $B>0$ there exists a $b>B$ such that the sequence on $D[0,b]$ (under the natural projection) converges weakly. Let ρ denote the sup metric used on $C[0,1]$ and $D[0,1]$ (used only in the latter when limiting distributions lie in $C[0,1]$ with probability 1), that is, for $x,y\in D[0,1]$

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

We need one further general result on weak convergence, which is an extension of the material on pp. 144-145 in [2] concerning random changes of time. Let

$$\underline{D}[0,1]=\{x\in D[0,1]:\, x \text{ is nonnegative and nondecreasing}\}$$

Since it is a closed subset of $D[0,1]$ we take the topology of $\underline{D}[0,1]$ to be the Skorohod topology of $D[0,1]$ relativized to it. The mapping

$$h: D[0,\infty)\times \underline{D}[0,1]\longrightarrow D[0,1]$$

defined by $h(x,\varphi)=x\circ\varphi$ is measurable (same argument as in [2], p. 232, except i in (39) now ranges on all natural numbers). It is a simple matter to show that h is continuous for each

$$(x,\varphi)\in C[0,\infty)\times C[0,1]\cap \underline{D}[0,1].$$

Therefore, we have (by Corollary 1 to Theorem 5.1 of [2])

$$(2.2)\qquad (Y_n,\Phi_n)\xrightarrow{\mathcal{D}}(Y,\Phi) \text{ in } D[0,\infty)\times\underline{D}[0,1],\ \mathsf{P}\big(Y\in C[0,\infty)\big)=\mathsf{P}\big(\Phi\in C[0,1]\big)=1$$

$$\Longrightarrow Y_n\circ\Phi_n\xrightarrow{\mathcal{D}}Y\circ\Phi \quad \text{ in } D[0,1].$$

We can now proceed with the proof of the theorem. Since we are only concerned with distributional results we may as well assume that for all $x\geq 0$, $F_n(x)\xrightarrow{a.s.}F_y(x)$ and $\lambda_{\max}\xrightarrow{a.s.}(1+\sqrt{y})^2$ (for this is true on an appropriate subsequence of an arbitrary subsequence of the natural numbers). For $t\in[0,1]$ let $F_n^{-1}(t)=$ largest λ_j such that $F_n(\lambda_j)\leq t$ (0 for $t<F_n(0)$). We have $X_n\big(F_n(F_n^{-1}(t))\big)=X_n(t)$ except on intervals $[\frac{m}{n},\frac{m+1}{n})$ where $\lambda_m=\lambda_{m+1}$. Let $F_y^{-1}(t)$ be the inverse of $F_y(x)$ for $x\in\big((1-\sqrt{y})^2,(1+\sqrt{y})^2\big]$

We consider first the case $y\leq 1$. Let $F_y^{-1}(0)=(1-\sqrt{y})^2$. It is straightforward to show for all $t\in(0,1]$, $F_n^{-1}(t)\xrightarrow{a.s.}F_y^{-1}(t)$. Let $\tilde{F}_n^{-1}(t)=\max\big((1-\sqrt{y})^2,F_n^{-1}(t)\big)$. Then, for all

$t \in [0,1]$, $\tilde{F}_n^{-1}(t) \xrightarrow{a.s.} F_y^{-1}(t)$, and since $\lambda_n \xrightarrow{a.s.} (1+\sqrt{y})^2$ we have $\rho\big(\tilde{F}_n^{-1}, F_y^{-1}\big) \xrightarrow{a.s.} 0$. Therefore, from (2.2) (and Theorem 4.4 of [2]) we have

$$X_n\big(F_n(\tilde{F}_n^{-1}(\cdot))\big) \xrightarrow{\mathcal{D}} W^\circ_{F_y(F_y^{-1}(\cdot))} = W^\circ \quad \text{in } D[0,1].$$

Since $F_y(x) = 0$ for $x \in [0,(1-\sqrt{y})^2]$ we have $X_n(F_n(\cdot)) \xrightarrow{\mathcal{D}} 0$ in $D[0,(1-\sqrt{y})^2]$, which implies $X_n(F_n(\cdot)) \xrightarrow{i.p.} 0$ in $D[0,(1-\sqrt{y})^2]$, and since the zero function lies in $C[0,(1-\sqrt{y})^2]$ we conclude that

$$\sup_{t\in[0,(1-\sqrt{y})^2]} |X_n(F_n(x))| \xrightarrow{i.p.} 0.$$

We have then

$$\rho\Big(X_n\big(F_n(F_n^{-1}(\cdot))\big), X_n\big(F_n(\tilde{F}_n^{-1}(\cdot))\big)\Big) \le 2 \times \sup_{t\in[0,(1-\sqrt{y})^2]} |X_n(F_n(x))| \xrightarrow{i.p.} 0.$$

Therefore, we have (by Theorem 4.1 of [2])

$$X_n\big(F_n(F_n^{-1}(\cdot))\big) \xrightarrow{\mathcal{D}} W^\circ \quad \text{in } D[0,1].$$

Notice if v_{11} has a density then we would be done with this case of the proof since for $n \le s$ the eigenvalues would be distinct with probability 1, so that $X_n\big(F_n(F_n^{-1}(\cdot))\big) = X_n(\cdot)$ almost surely. However, for more general v_{11}, the multiplicities of the eigenvalues need to be accounted for.

For each M_n let $\lambda_{(1)} < \lambda_{(2)} < \cdots < \lambda_{(t)}$, $(m_1, m_2, \ldots, m_t)$, and $(a_1, a_2, \ldots, a_t)$ be defined as above. We have from (2.1)

$$\rho\big(X_n(\cdot), X_n(F_n(F_n^{-1}(\cdot)))\big) = \max_{\substack{1\le i\le t\\ 1\le j\le m_i}} \sqrt{\frac{n}{2}} \left| a_i^2 \frac{\sum_{\ell=1}^{j} z^2_{m_1+\cdots+m_{i-1}+\ell}}{\sum_{k=1}^{m_i} z^2_{m_1+\cdots+m_{i-1}+k}} - \frac{j}{n} \right|. \tag{2.3}$$

The measurable function h on $D[0,1]$ defined by

$$h(x) = \rho(x(t), x(t-0))$$

is continuous on $C[0,1]$ (note that $h(x) = \lim_{\delta\downarrow 0} w(x,\delta)$ where $w(x,\delta)$ is the modulus of continuity of x) and is identically zero on $C[0,1]$. Therefore (using Corollary 1 to Theorem 5.1 of [2]) $h\big(X_n(F_n(F_n^{-1}(\cdot)))\big) \xrightarrow{\mathcal{D}} 0$, which is equivalent to

$$\max_{1\le i\le t} \sqrt{\frac{n}{2}} \left| a_i^2 - \frac{m_i}{n} \right| \xrightarrow{i.p.} 0. \tag{2.4}$$

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

$$\sqrt{\frac{n}{2}}\left(a_i^2\frac{\sum_{\ell=1}^{j} z^2_{m_1+\cdots+m_{i-1}+\ell}}{\sum_{k=1}^{m_i} z^2_{m_1+\cdots+m_{i-1}+k}} - \frac{j}{n}\right) =$$

$$(a) \qquad \sqrt{\frac{n}{2}}(a_i^2 - \frac{m_i}{n})\frac{\sum_{\ell=1}^{j} z^2_{m_1+\cdots+m_{i-1}+\ell}}{\sum_{k=1}^{m_i} z^2_{m_1+\cdots+m_{i-1}+k}} +$$

$$(b) \qquad \sqrt{\frac{n}{2}}\frac{m_i}{n}\left(a_i^2\frac{\sum_{\ell=1}^{j} z^2_{m_1+\cdots+m_{i-1}+\ell}}{\sum_{k=1}^{m_i} z^2_{m_1+\cdots+m_{i-1}+k}} - \frac{j}{m_i}\right).$$

From (2.4) we have the maximum of the absolute value of (a) over $1 \le i \le t$ converges in probability to zero. For the maximum of (b) we see that the ratio of chi-square random variables is beta distributed with parameters $p = j/2$, $q = (m_i - j)/2$. Such a random variable with $p = r/2$, $q = (m - r)/2$ has mean r/m and fourth central moment bounded by Cr^2/m^4 where C does not depend on r and m. Let $b_{m_i,j}$ represent the expression in parentheses in (b). Let $\epsilon > 0$ be arbitrary. We use Theorem 12.2 of [2] after making the following associations: $S_j = \sqrt{m_i}b_{m_i,j}$, $m = m_i$, $u_\ell = \sqrt{C}/m_i$, $\gamma = 4$, $\alpha = 2$, and $\lambda = \epsilon\sqrt{2n/m_i}$. We then have

$$\mathsf{P}\left(\max_{1\le j\le m_i}\left|\sqrt{\frac{n}{2}}\frac{m_i}{n}b_{m_i,j}\right| > \epsilon \;\middle|\; M_n\right) \le \frac{C' m_i^2}{4n^2\epsilon^4}.$$

By Boole's inequality we have

$$\mathsf{P}\left(\max_{\substack{1\le i\le t\\ 1\le j\le m_i}}\left|\sqrt{\frac{n}{2}}\frac{m_i}{n}b_{m_i,j}\right| > \epsilon \;\middle|\; M_n\right) \le \frac{C'}{4\epsilon^4}\max_{1\le i\le t}\frac{m_i}{n}.$$

Therefore

$$(2.5) \qquad \mathsf{P}\left(\max_{\substack{1\le i\le t\\ 1\le j\le m_i}}\left|\sqrt{\frac{n}{2}}\frac{m_i}{n}b_{m_i,j}\right| > \epsilon\right) \le \frac{C'}{4\epsilon^4}\mathsf{E}\left(\max_{1\le i\le t}\frac{m_i}{n}\right).$$

We have $F_n(x) \xrightarrow{a.s.} F_y(x) \implies \sup_{x\in[0,\infty)} |F_n(x)-F_y(x)| \xrightarrow{a.s.} 0 \implies$ (since F_y is continuous on $(-\infty,\infty)$) $\sup_{x\in[0,\infty)} |F_n(x)-F_n(x-0)| \xrightarrow{a.s.} 0$, which is equivalent to $\max_{1\le i\le t} m_i/n \xrightarrow{a.s.} 0$. Therefore, by the dominated convergence theorem, we have the left hand side of (2.5) $\longrightarrow 0$. We therefore have (2.3) $\xrightarrow{i.p.} 0$ and we conclude (again from Theorem 4.1 of [2]) that $X_n \xrightarrow{\mathcal{D}} W^\circ$ in $D[0,1]$.

For $y > 1$ we assume n is sufficiently large so that $n/s > 1$. Then $F_n(0) = m_1/n \geq 1-(s/n) > 0$. For $t \in [0, 1-(1/y)]$ define $F_y^{-1}(t) = (1-\sqrt{y})^2$. For $t \in (1-(1/y), 1]$ we have $F_n^{-1}(t) \xrightarrow{a.s.} F_y^{-1}(t)$. Define as before $\tilde{F}_n^{-1}(t) = \max\big((1-\sqrt{y})^2, F_n^{-1}(t)\big)$. Again, $\rho(\tilde{F}_n^{-1}, F_y^{-1}) \xrightarrow{a.s.} 0$, and from (2.2) (and Theorem 4.4 of [2]) we have

$$X_n\big(F_n(\tilde{F}_n^{-1}(t))\big) \xrightarrow{\mathcal{D}} W^\circ_{F_y(F_y^{-1}(t))} = \begin{cases} W^\circ_{1-(1/y)} & \text{for } t \in [0, 1-(1/y)], \\ W^\circ_t & \text{for } t \in [1-(1/y), 1]. \end{cases}$$

We have

$$\begin{aligned} &\rho\big(X_n\big(F_n(F_n^{-1}(\cdot))\big), X_n\big(F_n(\tilde{F}_n^{-1}(\cdot))\big)\big) \\ &= \sup_{x\in[0,(1-\sqrt{y})^2]} \big|X_n\big(F_n(x)\big) - X_n\big(F_n((1-\sqrt{y})^2)\big)\big| \\ &\xrightarrow{\mathcal{D}} \sup_{x\in[0,(1-\sqrt{y})^2]} |W^\circ_{F_y(x)} - W^\circ_{F_y((1-\sqrt{y})^2)}| = 0 \end{aligned}$$

which implies

$$\rho\big(X_n\big(F_n(F_n^{-1}(\cdot))\big), X_n\big(F_n(\tilde{F}_n^{-1}(\cdot))\big)\big) \xrightarrow{i.p.} 0$$

Therefore (by Theorem 4.1 of [2])

$$X_n\big(F_n(F_n^{-1}(\cdot))\big) \xrightarrow{\mathcal{D}} W^\circ_{F_y(F_y^{-1}(\cdot))}.$$

For $t < F_n(0) + \frac{1}{n}$

$$X_n(t) = \sqrt{\frac{n}{2}} \left(a_1^2 \; \frac{\sum_{i=1}^{[nt]} z_i^2}{\sum_{\ell=1}^{nF_n(0)} z_\ell^2} - \frac{[nt]}{n} \right)$$

$$= \frac{a_1^2}{\sqrt{F_n(0)}} \sqrt{\frac{nF_n(0)}{2}} \left(\frac{\sum_{i=1}^{[nt]} z_i^2}{\sum_{\ell=1}^{nF_n(0)} z_\ell^2} - \frac{[nt]}{nF_n(0)} \right) + \frac{[nt]}{nF_n(0)} \sqrt{\frac{n}{2}} (a_1^2 - F_n(0))$$

Notice that $\sqrt{\frac{n}{2}}(a_1^2 - F_n(0)) = X_n(F_n(0))$.

For $t \in [0,1]$ let $\varphi_n(t) = \min(\frac{t}{F_n(0)}, 1)$, $\varphi(t) = \min(\frac{t}{1-(1/y)}, 1)$, and

$$Y_n(t) \quad = \quad \sqrt{\frac{n}{2}} \left(\frac{\sum_{i=1}^{[nt]} z_i^2}{\sum_{\ell=1}^{n} z_\ell^2} \quad - \quad \frac{[nt]}{n} \right).$$

Then $\varphi_n \xrightarrow{i.p.} \varphi$ in $D_0 \equiv \{x \in \underline{D}[0,1] : x(1) \leq 1\}$ (see [2], p.144), and for $t < F_n(0) + \frac{1}{n}$

$$Y_{nF_n(0)}(\varphi_n(t)) \quad = \quad \sqrt{\frac{nF_n(0)}{2}} \left(\frac{\sum_{i=1}^{[nt]} z_i^2}{\sum_{\ell=1}^{nF_n(0)} z_\ell^2} \quad - \quad \frac{[nt]}{nF_n(0)} \right).$$

For all $t \in [0,1]$ let

$$H_n(t) = \frac{a_1^2}{\sqrt{F_n(0)}} Y_{nF_n(0)}(\varphi_n(t)) + X_n(F_n(0)) \left(\frac{[nF_n(0)\varphi_n(t)]}{nF_n(0)} - 1 \right) + X_n\big(F_n(F_n^{-1}(t))\big).$$

Then $H_n(t) = X_n(t)$ except on intervals $[\frac{m}{n}, \frac{m+1}{n})$ where $0 < \lambda_m = \lambda_{m+1}$. We will show $H_n \xrightarrow{\mathcal{D}} W^\circ$ in $D[0,1]$.

Let $\psi_n(t) = F_n(0)t$, $\psi(t) = (1 - (1/y))t$, and

$$V_n(t) = \frac{1}{\sqrt{2n}} \sum_{i=1}^{[nt]} (z_i^2 - 1).$$

Then $\psi_n \xrightarrow{i.p.} \psi$ in D_0 and

$$Y_n(t) = \frac{V_n(t) - \frac{[nt]}{n} V_n(1)}{1 + \sqrt{\frac{2}{n}} V_n(1)}. \tag{2.6}$$

Since $X_n\big(F_n(F_n^{-1}(\cdot))\big)$ and V_n are independent we have (using Theorems 4.4, 16.1 of [2])

$$\left(X_n\big(F_n(F_n^{-1}(\cdot))\big), V_n, \varphi_n, \psi_n\right) \xrightarrow{\mathcal{D}} \left(W^\circ_{F_y(F_y^{-1}(\cdot))}, \overline{W}, \varphi, \psi, \right)$$

where $\overline{W}$ is a Weiner process, independent of W°. We immediately get ([2], p.145)

$$\left(X_n\big(F_n(F_n^{-1}(\cdot))\big), V_n \circ \psi_n, \varphi_n\right) \xrightarrow{\mathcal{D}} \left(W^\circ_{F_y(F_y^{-1}(\cdot))}, \overline{W} \circ \psi, \varphi\right).$$

Since $V_n(\psi_n(t)) = \sqrt{F_n(0)}V_{nF_n(0)}(t)$, we have

$$\rho(V_n \circ \psi_n, \sqrt{1-(1/y)}V_{nF_n(0)}) = |\sqrt{F_n(0)} - \sqrt{1-(1/y)}| \sup_{t\in[0,1]} |V_{nF_n(0)}(t)| \xrightarrow{i.p.} 0.$$

Therefore

$$\text{(2.7)} \qquad \big(X_n\big(F_n(F_n^{-1}(\cdot))\big), V_{nF_n(0)}, \varphi_n\big) \xrightarrow{\mathcal{D}} \left(W^\circ_{F_y(F_y^{-1}(\cdot))}, \frac{1}{\sqrt{1-(1/y)}}\overline{W} \circ \psi, \varphi\right).$$

Notice that $\frac{1}{\sqrt{1-(1/y)}}\overline{W} \circ \psi$ is again a Weiner process, independent of W°.

From (2.6) we have

$$Y_n(t) - (V_n(t) - tV_n(1)) = V_n(t)\frac{t - \frac{[nt]}{n} + \sqrt{\frac{2}{n}}(tV_n(1) - V_n(t))}{1 + \sqrt{\frac{2}{n}}V_n(1)}.$$

Therefore

$$\text{(2.8)} \qquad \rho\big(Y_{nF_n(0)}(t), V_{nF_n(0)}(t) - tV_{nF_n(0)}(1)\big) \xrightarrow{i.p.} 0.$$

From (2.7), (2.8), and the fact that $W_t - tW_1$ is Brownian bridge it follows that

$$\big(X_n\big(F_n(F_n^{-1}(\cdot))\big), Y_{nF_n(0)}, \varphi_n\big) \xrightarrow{\mathcal{D}} \left(W^\circ_{F_y(F_y^{-1}(\cdot))}, \widehat{W}^\circ, \varphi\right)$$

where $\widehat{W}^\circ$ is another Brownian bridge, independent of W°.

The mapping $h : D[0,1] \times D[0,1] \times D_0 \longrightarrow D[0,1]$ defined by

$$h(x_1, x_2, z) = \sqrt{1-(1/y)}x_2 \circ z + x_1(0)(z-1) + x_1$$

is measurable, and is continuous on $C[0,1] \times C[0,1] \times D \cap C[0,1]$. Also, from (2.4) we have $a_1^2 \xrightarrow{i.p.} 1-(1/y)$. Finally, it is easy to verify

$$\frac{[nF_n(0)\varphi_n(t)]}{nF_n(0)} \xrightarrow{i.p.} 0$$

Therefore, we can conclude (using Theorem 4.1 and Corollary 1 of Theorem 5.1 of [2])

$$H_n \xrightarrow{\mathcal{D}} \sqrt{1-(1/y)}\widehat{W}^\circ \circ \varphi + W^\circ_{1-(1/y)}(\varphi - 1) + W^\circ_{F_y(F_y^{-1}(\cdot))} \equiv H.$$

It is immediately clear that H is a mean 0 Gaussian process lying in $C[0,1]$. It is a routine matter to verify for $0 \le s \le t \le 1$

$$\mathsf{E}(H_sH_t) = s(1-t).$$

Therefore, H is Brownian bridge.

We see that $\rho(X_n, H_n)$ is the same as the right hand side of (2.3) except $i = 1$ is excluded. The arguments leading to (2.4) and (2.5) ($2 \le i \le t$) are exactly the same as before. The fact that $\max_{2\le i\le t} m_i/n \xrightarrow{i.p.} 0$ follows from the case $y \le 1$ since the non-zero eigenvalues (including multiplicities) of AA^T and A^TA are identical for any rectangular A. Thus

$$\rho(X_n, H_n) \xrightarrow{i.p.} 0$$

and we have X_n converging weakly to Brownian bridge. □

3. A new condition for weak convergence. In this section we establish two results on random elements of $D[0, b]$ needed for the proof of the Theorem 1.1. In the following, we denote the modulus of continuity of $x \in D[0, b]$ by $w(x, \cdot)$:

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

To simplify the analysis we assume, for now, $b = 1$.

THEOREM 3.1. *Let $\{X_n\}$ be a sequence of random elements of $D[0, 1]$ whose probability measures satisfy the assumptions of Theorem 15.5 of* [2], *that is, $\{X_n(0)\}$ is tight, and for every positive ϵ and η, there exists a $\delta \in (0, 1)$ and an integer n_0, such that, for all $n > n_0$, $\mathsf{P}(w(X_n, \delta) \ge \epsilon) \le \eta$. If there exists a random element X with $\mathsf{P}(X \in C[0, 1]) = 1$ and such that*

$$\left\{\int_0^1 t^r X_n(t)dt\right\}_{r=0}^{\infty} \xrightarrow{\mathcal{D}} \left\{\int_0^1 t^r X(t)dt\right\}_{r=0}^{\infty} \quad \text{as } n \to \infty \tag{3.1}$$

(($\mathcal{D}$) in (3.1) denoting weak convergence on R^∞), then $X_n \xrightarrow{\mathcal{D}} X$.

PROOF. Note that the mappings

$$x \longrightarrow \int_0^1 t^r x(t)dt$$

are continuous in $D[0, 1]$. Therefore, by Theorems 5.1 and 15.5, $X_n \xrightarrow{\mathcal{D}} X$ will follow if we can show the distribution of X is uniquely determined by the distribution of

$$\left\{\int_0^1 t^r X(t)dt\right\}_{r=0}^{\infty}. \tag{3.2}$$

Since the finite dimensional distributions of X uniquely determine the distribution of X, it suffices to show for any integer m and numbers a_i, t_i, $i = 0, 1, \ldots, m$ with $0 = t_0 < t_1 < \cdots < t_m = 1$, the distribution of

$$\sum_{i=0}^{m} a_i X(t_i) \tag{3.3}$$

is uniquely determined by the distribution of (3.2).

Let $\{f_n\}, f$ be uniformly bounded measurable functions on [0,1] such that $f_n \to f$ pointwise as $n \to \infty$. Using the dominated convergence theorem we have

$$\int_0^1 f_n(t)X(t)dt \to \int_0^1 f(t)X(t)dt \quad \text{as } n \to \infty. \tag{3.4}$$

Let $\epsilon > 0$ be any number less than half the minimum distance between the t_i's. Notice for the indicator function $I_{[a,b]}$ we have the sequence of continuous "ramp" functions $\{R_n(t)\}$ with

$$R_n(t) = \begin{cases} 1 & t \in [a,b], \\ 0 & t \in [a - \frac{1}{n}, b + \frac{1}{n}]^c, \end{cases}$$

and linear on each of the sets $[a - \frac{1}{n}, a]$, $[b, b + \frac{1}{n}]$, satisfying $R_n \downarrow I_{[a,b]}$ as $n \to \infty$. Notice also that we can approximate any ramp function uniformly on [0,1] by polynomials. Therefore, using (3.4) for polynomials appropriately chosen, we find that the distribution of

$$\sum_{i=0}^{m-1} a_i \int_{t_i}^{t_i+\epsilon} X(t)dt + a_m \int_{1-\epsilon}^{1} X(t)dt \tag{3.5}$$

is uniquely determined by the distribution of (3.2).

Dividing (3.5) by ϵ and letting $\epsilon \to 0$ we get a.s. convergence to (3.3), (since $X \in C[0,1]$ with probability one), and we are done. □

Theorem 3.2. *Let X be a random element of $D[0,1]$. Suppose there exists constants $B > 0$, $\gamma \geq 0$, $\alpha > 1$, and a random nondecreasing, right-continuous function $F : [0,1] \to [0,B]$ such that, for all $0 \leq t_1 \leq t_2 \leq 1$ and $\lambda > 0$*

$$\mathsf{P}(|X(t_2) - X(t_1)| \geq \lambda) \leq \frac{1}{\lambda^\gamma}\mathsf{E}[(F(t_2) - F(t_1))^\alpha]. \tag{3.6}$$

Then for every $\epsilon > 0$ and δ, an inverse of a positive integer, we have

$$\mathsf{P}(w(X,\delta) \geq 3\epsilon) \leq \frac{KB}{\epsilon^\gamma}\mathsf{E}\left[\max_{j<\delta^{-1}} (F((j+1)\delta) - F(j\delta))^{\alpha-1}\right], \tag{3.7}$$

where j ranges on positive integers, and K depends only on γ and α.

This theorem is proven by modifying the proofs of the first three theorems in section 12 of [2]. It is essentially an extension of part of a result contained in Theorem 12.3 of [2]. The original arguments, for the most part, remain unchanged. We will indicate only the specific changes and refer the reader to [2] for details. The extensions of two of the theorems in [2] will be given below as lemmas. However, some definitions must first be given.

Let $\xi_1, \dots, \xi_m$ be random variables, and $S_k = \xi_1 + \cdots + \xi_k$ ($S_0 = 0$). Let

$$M_m = \max_{0 \le k \le m} |S_k| \quad \text{and}$$

$$M'_m = \max_{0 \le k \le m} \min(|S_k|, |S_m - S_k|).$$

LEMMA 3.1 (extends Theorem 12.1 of [2]). *Suppose $u_1, \dots, u_m$ are non-negative random variables such that*

$$\mathsf{P}(|S_j - S_i| \ge \lambda, |S_k - S_j| \ge \lambda) \le \frac{1}{\lambda^{2\gamma}} \mathsf{E}\left[\left(\sum_{i<\ell\le k} u_\ell\right)^{2\alpha}\right] < \infty, \qquad 0 \le i \le j \le k \le m$$

for some $\alpha > \frac{1}{2}$, $\gamma \ge 0$, and for all $\lambda > 0$. Then, for all $\lambda > 0$

$$\mathsf{P}(M'_m \ge \lambda) \le \frac{K}{\lambda^{2\gamma}} \mathsf{E}[(u_1 + \cdots + u_m)^{2\alpha}], \tag{3.8}$$

where $K = K_{\gamma,\alpha}$ depends only on γ and α.

PROOF. We follow [2], p. 91. The constant K is chosen the same way and the proof proceeds by induction on m. The arguments for $m = 1$ and 2 are the same, except, for the latter, $(u_1 + u_2)^{2\alpha}$ is replaced by $\mathsf{E}(u_1 + u_2)^{2\alpha}$. Assuming (3.8) is true for all integers less than m, we find an integer h, $1 \le h \le m$ such that

$$\frac{\mathsf{E}[(u_1 + \cdots + u_{h-1})^{2\alpha}]}{\mathsf{E}[(u_1 + \cdots + u_m)^{2\alpha}]} \le \frac{1}{2} \le \frac{\mathsf{E}[(u_1 + \cdots + u_h)^{2\alpha}]}{\mathsf{E}[(u_1 + \cdots + u_m)^{2\alpha}]},$$

the sum on the left hand side being 0 if $h = 1$.

Since $2\alpha > 1$, we have for all nonnegative x and y

$$x^{2\alpha} + y^{2\alpha} \le (x + y)^{2\alpha}.$$

We have then

$$\begin{aligned}\mathsf{E}[(u_{h+1} + \cdots + u_m)^{2\alpha}] &\le \mathsf{E}[(u_1 + \cdots + u_m)^{2\alpha}] - \mathsf{E}[(u_1 + \cdots + u_h)^{2\alpha}] \\ &\le \mathsf{E}[(u_1 + \cdots + u_m)^{2\alpha}](1 - \tfrac{1}{2}) = \tfrac{1}{2}\mathsf{E}[(u_1 + \cdots + u_m)^{2\alpha}].\end{aligned}$$

Therefore, defining U_1, U_2, D_1, D_2 as in [2], we get the same inequalities as in (12.30)-(12.33) ([2], p. 92) with $u^{2\alpha}$ replaced by $\mathsf{E}[(u_1+\cdots+u_m)^{2\alpha}]$. The rest of the proof follows exactly. □

Lemma 3.2 (extends Theorem 12.2 of [2]). *If, for random nonnegative u_ℓ, there exists $\alpha > 1$ and $\gamma \geq 0$ such that, for all $\lambda > 0$*

$$\mathsf{P}(|S_j - S_i| \geq \lambda) \leq \frac{1}{\lambda^\gamma}\mathsf{E}\left[\left(\sum_{i<\ell\leq j} u_\ell\right)^{2\alpha}\right] < \infty, \quad 0 \leq i \leq j \leq m$$

then

$$\mathsf{P}(M_n \geq \lambda) \leq \frac{K'_{\gamma,\alpha}}{\lambda^\gamma}\mathsf{E}[(u_1+\cdots+u_m)^{2\alpha}], \qquad K'_{\gamma,\alpha} = 2^\gamma(1+K_{\frac{1}{2}\gamma,\frac{1}{2}\alpha}).$$

Proof. Following [2] we have for $0 \leq i \leq j \leq k \leq m$

$$\mathsf{P}(|S_j - S_i| \geq \lambda, |S_k - S_j| \geq \lambda) \leq \mathsf{P}^{\frac{1}{2}}(|S_j - S_i| \geq \lambda)\mathsf{P}^{\frac{1}{2}}(|S_k - S_j| \geq \lambda) \leq \frac{1}{\lambda\gamma}\mathsf{E}\left[\left(\sum_{i<\ell\leq k} u_\ell\right)^{2\alpha}\right],$$

so Lemma 3.1 is satisfied with constants $\frac{1}{2}\gamma, \frac{1}{2}\alpha$. The rest follows exactly as in [2], p. 94, with $(u_1+\cdots+u_m)^\alpha$ in (12.46), (12.47) replaced by the expected value of the same quantity. □

We can now proceed with the proof of Theorem 3.2. Following the proof of Theorem 12.3 of [2] we fix positive integers $j < \delta^{-1}$ and m and define

$$\xi_i = X\left(j\delta + \frac{i}{m}\delta\right) - X\left(j\delta + \frac{i-1}{m}\delta\right), \qquad i = 1, 2, \ldots, m.$$

The partial sums of the ξ_i's satisfy Lemma 3.2 with

$$u_i = F\left(j\delta + \frac{i}{m}\delta\right) - F\left(j\delta + \frac{i-1}{m}\delta\right).$$

Therefore

$$\mathsf{P}\left(\max_{1\leq i\leq m}\left|X\left(j\delta + \frac{i}{m}\delta\right) - X(j\delta)\right| \geq \epsilon\right) \leq \frac{K}{\epsilon^\gamma}\mathsf{E}[(F((j+1)\delta) - F(j\delta))^\alpha] \qquad K = K'_{\gamma,\alpha}.$$

Since $X \in D[0,1]$ we have

$$\begin{aligned}\mathsf{P}&\left(\sup_{j\delta\leq s\leq(j+1)\delta}|X(s) - X(j\delta)| > \epsilon\right)\\ &= \mathsf{P}\left(\max_{1\leq i\leq m}\left|X\left(j\delta + \frac{i}{m}\delta\right) - X(j\delta)\right| > \epsilon \text{ for all } m \text{ sufficiently large}\right)\\ &\leq \liminf_m \mathsf{P}\left(\max_{1\leq i\leq m}\left|X\left(j\delta + \frac{i}{m}\delta\right) - X(j\delta)\right| \geq \epsilon\right) \leq \frac{K}{\epsilon^\gamma}\mathsf{E}[(F((j+1)\delta) - F(j\delta))^\alpha].\end{aligned}$$

By considering a sequence of numbers approaching ϵ from below we get from the continuity theorem

$$\mathsf{P}\left(\sup_{j\delta\le s\le (j+1)\delta} |X(s)-X(j\delta)| \ge \epsilon\right) \le \frac{K}{\epsilon^\gamma}\mathsf{E}[(F((j+1)\delta)-F(j\delta))^\alpha]. \tag{3.9}$$

Summing both sides of (3.9) over all $j<\delta^{-1}$ and using the corollary to Theorem 8.3 of [2] we get

$$\begin{aligned}\mathsf{P}(w(X,\delta)\ge 3\epsilon) &\le \frac{K}{\epsilon^\gamma}\mathsf{E}\Bigg[\sum_{j<\delta^{-1}} (F((j+1)\delta)-F(j\delta))^\alpha\Bigg] \\ &\le \frac{K}{\epsilon^\gamma}\mathsf{E}\Big[\max_{j<\delta^{-1}} (F((j+1)\delta)-F(j\delta))^{\alpha-1}(F(1)-F(0))\Big] \\ &\le \frac{KB}{\epsilon^\gamma}\mathsf{E}\Big[\max_{j<\delta^{-1}} (F((j+1)\delta)-F(j\delta))^{\alpha-1}\Big],\end{aligned}$$

and we are done. □

For general $D[0,b]$ we simply replace (3.1) by

$$\left\{\int_0^b t^r X_n(t)dt\right\}_{r=0}^\infty \xrightarrow{\mathcal{D}} \left\{\int_0^b t^r X(t)dt\right\}_{r=0}^\infty \quad \text{as } n\to\infty \tag{3.10}$$

and (3.7) by

$$\mathsf{P}(w(X,b\delta)\ge 3\epsilon) \le \frac{KB}{\epsilon^\gamma}\mathsf{E}\left[\max_{j<\delta^{-1}} (F(b(j+1)\delta)-F(bj\delta))^{\alpha-1}\right], \tag{3.11}$$

j and δ^{-1} still positive integers.

4. Completing the proof. We finish up by verifying the conditions of Theorem 3.1.

THEOREM 4.1. *Let* $\mathsf{E}(v_{11})=0$, $\mathsf{E}(v_{11}^2)=1$, *and* $\mathsf{E}(v_{11}^4)<\infty$. *Suppose the sequence of vectors* $\{\vec{x}_n\}$, $\vec{x}_n=(x_{n1},x_{n2},\dots,x_{nn})^T$, $\|\vec{x}_n\|=1$ *satisfies*

$$\sum_{i=1}^n x_{ni}^4 \longrightarrow 0 \quad \text{as } n\to\infty. \tag{4.1}$$

Then (1.5) holds.

PROOF. Let $\bar{v}_{ij}=\bar{v}_{ij}(n)=v_{ij}I_{(|v_{ij}|\le n^{1/4})}-\mathsf{E}(v_{ij}I_{(|v_{ij}|\le n^{1/4})})$, and let $\bar{M}_n=\frac{1}{s}\bar{V}_n\bar{V}_n^T$, where $\bar{V}_n=(\bar{v}_{ij})$. We have $\mathsf{E}(\bar{v}_{11})=0$, $\mathsf{E}(\bar{v}_{11}^2)\to 1$, and $\mathsf{E}(\bar{v}_{11}^4)\to\mathsf{E}(v_{11}^4)$ as $n\to\infty$.

The main part of the proof in [11] establishing (1.5) (under the additional assumption $\mathsf{E}(v_{11}^4)=3$) relies on a multidimensional version of the method of moments, together with the fact that (1.5) holds in the Wishart case. It is shown that for any integer $m \geq 2$ and positive integers $r_1, r_2, \ldots, r_m$, the asymptotic behavior of

$$n^{m/2}\mathsf{E}\left[\left(\vec{x}_n^T \bar{M}_n^{r_1}\vec{x}_n - \mathsf{E}(\vec{x}_n^T \bar{M}_n^{r_1}\vec{x}_n)\right)\left(\vec{x}_n^T \bar{M}_n^{r_2}\vec{x}_n - \mathsf{E}(\vec{x}_n^T \bar{M}_n^{r_2}\vec{x}_n)\right) \cdots \left(\vec{x}_n^T \bar{M}_n^{r_m}\vec{x}_n - \mathsf{E}(\vec{x}_n^T \bar{M}_n^{r_m}\vec{x}_n)\right)\right] \tag{4.2}$$

depends only on $\mathsf{E}(\bar{v}_{11}^2)$ and $\mathsf{E}(\bar{v}_{11}^4)$. Using the fact that (4.2) converges to the appropriate limit when v_{11} is N(0,1), (1.5) follows when $\mathsf{E}(v_{11}^4) = 3$.

This is the only place in the proof in [11] that refers to the value of $\mathsf{E}(v_{11}^4)$, the remaining arguments depending on this value only to the extent of it being finite, and thus apply to the present case. Therefore, we will be done if we can show that (4.1) implies the asymptotic behavior of (4.2) depends only on $\mathsf{E}(\bar{v}_{11}^2)$. Although it will be necessary to repeat some of the discussion in the original proof, we refer the reader to [11] for specific details.

We have (dropping the dependency of n on the components of $\vec{x}_n$)

$$\left(\frac{s^{r_1+\cdots+r_m}}{n^{m/2}}\right) \times (4.2) \ =$$

$$\sum x_{i^1}x_{j^1}\cdots x_{i^m}x_{j^m}\mathsf{E}\left[(\bar{v}_{i^1k_1^1}\bar{v}_{i_2^1k_1^1}\cdots\bar{v}_{i_{r_1}^1k_{r_1}^1}\bar{v}_{j^1k_{r_1}^1} - \mathsf{E}(\bar{v}_{i^1k_1^1}\bar{v}_{i_2^1k_1^1}\cdots\bar{v}_{i_{r_1}^1k_{r_1}^1}\bar{v}_{j^1k_{r_1}^1})) \cdots (\bar{v}_{i^mk_1^m}\bar{v}_{i_2^mk_1^m}\cdots\bar{v}_{i_{r_m}^mk_{r_m}^m}\bar{v}_{j^mk_{r_m}^m} - \mathsf{E}(\bar{v}_{i^mk_1^m}\bar{v}_{i_2^mk_1^m}\cdots\bar{v}_{i_{r_m}^mk_{r_m}^m}\bar{v}_{j^mk_{r_m}^m}))\right], \tag{4.3}$$

where the sum is over $i^1, j^1, i_2^1, \ldots, i_{r_1}^1, k_1^1, \ldots, k_{r_1}^1, \ldots, i^m, j^m, i_2^m, \ldots, i_{r_m}^m, k_1^m, \ldots, k_{r_m}^m$.

We consider one of the ways the two set of indices
$I \equiv \{i^1, j^1, i_2^1, \ldots, i_{r_1}^1, \ldots, i^m, j^m, i_2^m, \ldots, i_{r_m}^m\}$, and $K \equiv \{k_1^1, \ldots, k_{r_1}^1, \ldots, k_1^m, \ldots, k_{r_m}^m\}$ can each be partitioned. Associated with the two partitions are the terms in (4.3) (for n large) where indicies are equal in value if and only if they belong to the same class. We consider only those partitions corresponding to terms in (4.3) that contribute a non-negligible amount to (4.2) in the limit. Let ℓ be the number of classes of I indices containing only one element from $J \equiv \{i^1, j^1, i^2, j^2, \ldots, i^m, j^m\}$. Let d denote the number of classes of I indices containing no elements from J, plus the number of classes of K indices. Then the contribution to (4.3) of those terms associated with the two partitions is bounded in absolute value by

$$Cn^{(\ell/2)+d}\mathsf{E}\left(|\bar{v}_{i^1k_1^1}, \ldots, \bar{v}_{j^1k_{r_1}^1}, \ldots, \bar{v}_{i^mk_1^m}, \ldots, \bar{v}_{j^mk_{r_m}^m}|\right) \tag{4.4}$$

the expected value being one of those associated with the two partitions. We can write this expected value in the form

$$A_{a_1b_1}^1 A_{a_2b_2}^2 \cdots A_{a_{r'}b_{r'}}^{r'}$$

where $A^j_{a_j b_j}$ corresponds to $\bar{v}_{a_j b_j}$ appearing in (4.4), so that if $\bar{v}_{a_j b_j}$ appears t times, then $A^j_{a_j b_j} = \mathsf{E}(|\bar{v}^t_{a_j b_j}|)$. There are r' distinct elements of $\bar{V}_n$ involved in (4.4). For each ordered pair (a_j, b_j) either a_j or b_j will be repeated in at least one other ordered pair (see [11]). We say that a_j or b_j is free if it does not appear in any other ordered pair.

From [11] it was argued that $d = r_1 + \cdots + r_m - (m/2) - (\ell/2)$, for each j $A^j_{a_j b_j} = \mathsf{E}(\bar{v}^2_{11})$ or $\mathsf{E}(\bar{v}^4_{11})$, and, for our purposes, we can assume without loss of generality that each I class containing no element from J has at least two elements, and any $A^j_{a_j b_j}$ for which b_j is free involves $\bar{v}_{p^t q^t}$'s for at least two different t's. We continue the argument from this point.

We immediately conclude that the number of I classes containing no elements from J, and the number of K classes are, respectively, bounded by $\frac{1}{2}(r_1 + \cdots + r_m - m)$ and $\frac{1}{2}(r_1 + \cdots + r_m)$. We can also conclude that $\ell = 0$, since any I class containing only one element from J implies an element of $\bar{V}_n$ apearing in (4.4) an odd number of times, which is impossible. Therefore, we can conclude that there are $\frac{1}{2}(r_1 + \cdots + r_m)$ K classes, and $\frac{1}{2}(r_1 + \cdots + r_m - m)$ I classes containing no elements from J, which further implies no I class exists containing an element from J and an element from $(I - J)$. Therefore, the I classes split up into separate J and $I - J$ classes, and each K class and $I - J$ class consists of two elements.

Now, if $A^j_{a_j b_j} = \mathsf{E}(\bar{v}^4_{11})$ for some j, then either a_j or b_j must be associated with a class containing at least three elements, forcing a_j to be associated with a J class consisting of at least four elements. Therefore, the sum of the terms in (4.3) associated with the two partitions, divided by $(s^{r_1 + \cdots + r_m})/n^{m/2}$, will be bounded by

$$C \times \sum_{i=1}^{n} x_i^4$$

and because of (4.1) these terms will not contribute anything to (4.2) in the limit. We conclude that the asymptotic behavior of (4.2) depends only on $\mathsf{E}(\bar{v}^2_{11})$, and we are done. □

Let R_+ denote the nonnegative reals and $\mathcal{B}_+$, $\mathcal{B}^4_+$ denote the Borel σ-fields on, respectively, R_+ and R^4_+. For any $n \times n$ symmetric, nonnegative definite matrix B and any $A \in \mathcal{B}_+$, let $P^B(A)$ denote the projection matrix on the subspace of R^n spanned by the eigenvectors of B having eigenvalues in A (the collection of projections $\{P^B((-\infty, a]) : a \in R\}$ is usually referred to as the spectral family of B). We have $\mathrm{tr} P^B(A)$ equal to the number of eigenvalues of B contained in A. If B is random, then it is straightforward to verify the following facts:

a) For every $\vec{x}_n \in R^n$, $\|\vec{x}_n\| = 1$, $\vec{x}_n^T P^B(\cdot)\vec{x}_n$ is a random probability measure on R_+ placing mass on the eigenvalues of B.

b) For any four elements $P^B_{i_1j_1}(\cdot)$, $P^B_{i_2j_2}(\cdot)$, $P^B_{i_3j_3}(\cdot)$, $P^B_{i_4j_4}(\cdot)$ of $P^B(\cdot)$, the function defined on rectangles $A_1 \times A_2 \times A_3 \times A_4 \in \mathcal{B}^4_+$ by

$$\mathsf{E}\big(P^B_{i_1j_1}(A_1)P^B_{i_2j_2}(A_2)P^B_{i_3j_3}(A_3)P^B_{i_4j_4}(A_4)\big) \tag{4.5}$$

generates a signed measure $m^B_n = m_n^{B,(i_1,j_1,\ldots,i_4,j_4)}$ on $(R^4_+, \mathcal{B}^4_+)$ such that $|m^B_n(A)| \le 1$ for every $A \in \mathcal{B}^4_+$.

When $B = M_n$ we also have

c) For any $A \in \mathcal{B}_+$ the distribution of $P^{M_n}(A)$ is invariant under permutation transformations, that is, $P^{M_n}(A) \sim OP^{M_n}(A)O^T$ for any permutation matrix O (use the fact that $P^B(\cdot)$ is uniquely determined by $\{B^r\}_{r=1}^\infty$ along with $OP^B(\cdot)O^T = P^{OBO^T}(\cdot)$ and $\{M_n^r\}_{r=1}^\infty \sim \{(OM_nO^T)^r\}_{r=1}^\infty$).

d) For $0 \le x_1 \le x_2$

$$\tfrac{1}{n}P^{M_n}([0,x_1]) = F_n(x_1),$$

$$X_n(F_n(x_1)) = \sqrt{\tfrac{n}{2}}\big(\vec{x}_n^T P^{M_n}([0,x_1])\vec{x}_n - \tfrac{1}{n}\mathrm{tr}(P^{M_n}([0,x_1]))\big), \quad \text{and}$$

$$X_n(F_n(x_2)) - X_n(F_n(x_1)) = \sqrt{\tfrac{n}{2}}\big(\vec{x}_n^T P^{M_n}((x_1,x_2])\vec{x}_n - \tfrac{1}{n}\mathrm{tr}(P^{M_n}((x_1,x_2]))\big).$$

LEMMA 4.1. *Assume v_{11} is symmetric. If one of the indices $i_1, j_1, \ldots, i_4, j_4$ appears an odd number of times, then $m^{M_n}_n \equiv 0$.*

PROOF. Assume first that v_{11} is bounded. Then $\lambda_{\max}$ is bounded which implies $m^{M_n}_n$ has bounded support. Therefore, $m^{M_n}_n$ is uniquely determined by its mixed moments ([7], pp. 97-102). It is straightforward to show these moments can be expressed as

$$\mathsf{E}\big(({M_n}^{r_1})_{i_1j_1}({M_n}^{r_2})_{i_2j_2}({M_n}^{r_3})_{i_3j_3}({M_n}^{r_4})_{i_4j_4}\big) \tag{4.6}$$

for arbitrary nonnegative integers $r_1, \ldots, r_4$. If, say $r_1 = 0$ and $i_1 \ne j_1$, then obviously (4.6) is zero. We can assume then a positive power r_ℓ for which $i_\ell \ne j_\ell$. Upon expanding (4.6) as a sum of expected values of products of entries of V_n we find each product contains a v_{ij} appearing an odd number of times. Therefore, all mixed moments of $m^{M_n}_n$ are zero, implying $m^{M_n}_n \equiv 0$.

For arbitrary symmetric v_{11} we truncate. For any $c > 0$ let $v^c_{ij} = v_{ij}I_{(|v_{ij}|\le c)}$, and $M^c_n = \frac{1}{s}V^c_nV^{cT}_n$, where $V^c_n = (v^c_{ij})$. For any realization of the v_{ij}'s and $A_1 \times A_2 \times A_3 \times A_4 \in \mathcal{B}^4_+$

$$P^{M^c_n}_{i_1j_1}(A_1)P^{M^c_n}_{i_2j_2}(A_2)P^{M^c_n}_{i_3j_3}(A_3)P^{M^c_n}_{i_4j_4}(A_4) \quad = \quad P^{M_n}_{i_1j_1}(A_1)P^{M_n}_{i_2j_2}(A_2)P^{M_n}_{i_3j_3}(A_3)P^{M_n}_{i_4j_4}(A_4)$$

for c large enough. Therefore, by the dominated convergence theorem we have (4.5) (with $B = M_n$) equal to zero for all rectangles in $\mathcal{B}_+^4$, implying $m_n^{M_n} \equiv 0$. □

THEOREM 4.2. *Assume v_{11} is symmetric and $\vec{x}_n = (\pm\frac{1}{\sqrt{n}}, \pm\frac{1}{\sqrt{n}}, \ldots, \pm\frac{1}{\sqrt{n}})^T$. Let $G_n(x) = 4F_n(x)$. Then*

$$\mathsf{E}\big((X_n(F_n(0)))^4\big) \le \mathsf{E}\big((G_n(0))^2\big), \tag{4.7}$$

and for any $0 \le x_1 \le x_2$

$$\mathsf{E}\big((X_n(F_n(x_2)) - X_n(F_n(x_1)))^4\big) \le \mathsf{E}\big((G_n(x_2) - G_n(x_1))^2\big). \tag{4.8}$$

PROOF. With $A = \{0\}$ in (4.7), $A = (x_1, x_2]$ in (4.8) we use d) to find the left hand sides of (4.7) and (4.8) equal to

$$\frac{1}{4n^2}\mathsf{E}\left(\sum_{i\neq j}\gamma_{ij}P_{ij}^{M_n}(A)\right)^4 \tag{4.9}$$

where $\gamma_{ij} = \mathrm{sgn}(\vec{x}_n)_i(\vec{x}_n)_j$. For the remainder of the argument we simplify the notation by supressing the dependence of the projection matrix on M_n and A. Upon expanding (4.9) we use c) to combine identically distributed factors, and Lemma 4.1 to arrive at

$$\begin{aligned}(4.9) \;=\; \tfrac{(n-1)}{n}\big(&12(n-2)\mathsf{E}(P_{12}^2P_{13}^2) + 3(n-2)(n-3)\mathsf{E}(P_{12}^2P_{34}^2)\\ &+12(n-2)(n-3)\mathsf{E}(P_{12}P_{23}P_{34}P_{14}) + 2\mathsf{E}(P_{12}^4)\big).\end{aligned} \tag{4.10}$$

We can write the second and third expected values in (4.10) in terms of the first expected value and expected values involving P_{11}, P_{22}, and P_{12} by making further use of c) and the fact that P is a projection matrix (i.e., $P^2 = P$). For example, we take the expected value of both sides of the identity

$$P_{12}P_{23}\left(\sum_{j\ge4}P_{3j}P_{1j} + P_{31}P_{11} + P_{32}P_{12} + P_{33}P_{13}\right) = P_{12}P_{23}P_{13}$$

and get

$$(n-3)\mathsf{E}(P_{12}P_{23}P_{34}P_{14}) + 2\mathsf{E}(P_{11}P_{12}P_{23}P_{31}) + \mathsf{E}(P_{12}^2P_{13}^2) = \mathsf{E}(P_{12}P_{23}P_{13}).$$

Proceeding in the same way we find

$$(n-2)\mathsf{E}(P_{11}P_{12}P_{23}P_{31}) + \mathsf{E}(P_{11}^2P_{12}^2) + \mathsf{E}(P_{11}P_{22}P_{12}^2) = \mathsf{E}(P_{11}P_{12}^2) \quad \text{and}$$

$$(n-2)\mathsf{E}(P_{12}P_{23}P_{13}) + 2\mathsf{E}(P_{11}P_{12}^2) = \mathsf{E}(P_{12}^2).$$

Therefore

$$(n-2)(n-3)\mathsf{E}(P_{12}P_{23}P_{34}P_{14}) = \mathsf{E}(P_{12}^2) + 2\mathsf{E}(P_{11}P_{22}P_{12}^2)$$
$$+2\mathsf{E}(P_{11}^2P_{22}^2) - (n-2)\mathsf{E}(P_{12}^2P_{13}^2) - 4\mathsf{E}(P_{11}P_{12}^2).$$

Since $P_{11} \geq \max(P_{11}P_{22}, P_{11}^2)$ and $P_{12}^2 \leq P_{11}P_{22}$ (since P is nonnegative definite) we have

$$(n-2)(n-3)\mathsf{E}(P_{12}P_{23}P_{34}P_{14}) \leq \mathsf{E}(P_{11}P_{22}) - (n-2)\mathsf{E}(P_{12}^2P_{13}^2).$$

Similar arguments will yield

$$(n-3)\mathsf{E}(P_{12}^2P_{34}^2) + 2\mathsf{E}(P_{12}^2P_{13}^2) + \mathsf{E}(P_{12}^2P_{33}^2) = \mathsf{E}(P_{12}^2P_{33}) \quad \text{and}$$

$$(n-2)\mathsf{E}(P_{12}^2P_{13}^2) + \mathsf{E}(^4_{12}) + \mathsf{E}(P_{11}^2P_{12}^2) = \mathsf{E}(P_{11}P_{12}^2).$$

After multiplying the first equation by $n-2$ and adding it to the second, we get

$$(n-2)(n-3)\mathsf{E}(P_{12}^2P_{34}^2) + 3(n-2)\mathsf{E}(P_{12}^2P_{13}^2)$$
$$= (n-2)\mathsf{E}(P_{12}^2P_{33}) - (n-2)\mathsf{E}(P_{12}^2P_{33}^2) + \mathsf{E}(P_{11}P_{12}^2) - \mathsf{E}(P_{11}^2P_{12}^2) - \mathsf{E}(P_{12}^4)$$
$$= \mathsf{E}(P_{11}P_{22}) + \mathsf{E}(P_{11}^2P_{22}^2) - 2\mathsf{E}(P_{11}P_{22}^2) - \mathsf{E}(P_{12}^4) \leq \mathsf{E}(P_{11}P_{22}) - \mathsf{E}(P_{12}^4).$$

Combining the above expressions we obtain

$$(4.9) \;\leq\; 15\frac{(n-1)}{n}\mathsf{E}(P_{11}P_{22}).$$

Therefore, using c) and d), we get

$$(4.9) \leq \tfrac{15}{n^2}\mathsf{E}(\textstyle\sum_{i\neq j}P_{ii}P_{jj}) \leq \mathsf{E}((4\tfrac{1}{n}\mathrm{tr}P)^2) = \begin{cases} \mathsf{E}\big((G_n(0))^2\big) & \text{for } A = \{0\}, \\ \mathsf{E}\big((G_n(x_2) - G_n(x_1))^2\big) & \text{for } A = (x_1, x_2], \end{cases}$$

and we are done. □

We can now complete the proof of Theorem 1.1. We may assume $\mathsf{E}(v_{11}^2) = 1$. Choose any $b > (1+\sqrt{y})^2$. We have (1.3) and, by Theorem 4.1, (1.5), which imply

$$\left\{\int_0^b x^r X_n(F_n(x))dx\right\}_{r=0}^{\infty} \xrightarrow{\mathcal{D}} \left\{\int_0^b x^r W_x^y dx\right\}_{r=0}^{\infty} \qquad \text{as } n \to \infty,$$

so that (3.10) is satisfied. By Theorems 3.2 and 4.2 we have, for any $n \geq 16$, (3.11) with $X = X_n(F_n(\cdot))$, $F = 4F_n$, $B = 4$, $\gamma = 4$, and $\alpha = 2$. From (1.2) and Theorem 5.1 of [2] we have for every $\delta \in (0, b]$

$$w(F_n, \delta) \xrightarrow{i.p.} w(F_y, \delta) \qquad \text{as } n \to \infty.$$

Since F_y is continuous on $[0, \infty)$, we apply the dominated convergence theorem to the right hand side of (3.11) and find that, for every $\epsilon > 0$, $\mathsf{P}(w(X_n(F_n(\cdot)), \delta) \geq \epsilon)$ can be made arbitrarily small for all n sufficiently large by choosing δ appropriately. Therefore, by Theorem 3.1, $X_n(F_n(\cdot)) \xrightarrow{\mathcal{D}} W^\circ_{F_y(\cdot)}$ in $D[0, b]$, which implies $X_n(F_n(\cdot)) \xrightarrow{\mathcal{D}} W^\circ_{F_y(\cdot)}$ in $D[0, \infty)$, and by Theorem 2.1 we conclude that $X_n \xrightarrow{\mathcal{D}} W^\circ$. □

Acknowledgment. I would like to thank the Associate Editor and the referee for their helpful suggestions.

REFERENCES

[1] Bai, Z.D., Silverstein, J.W., and Yin, Y.Q. (1988). A note on the largest eigenvalue of a large dimensional sample covariance matrix. *J. Multivariate Anal.* **26** 166-168.

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

[3] Geman, S. (1980). A limit theorem for the norm of random matrices. *Ann. Probab.* **8** 252-261.

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

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

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

[7] Shohat, J.A. and Tamarkin, J.D. (1950). *The Problem of Moments*. Mathematical Surveys, No. 1, American Mathematical Society, Providence.

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

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

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

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

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

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

[14] 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.

Department of Mathematics
Box 8205
North Carolina State University
Raleigh, North Carolina 27695-8205