SIAM J. APPL. MATH.
Vol. 37, No. 2, October 1979

0036-1399/79/3702-0001 $01.00/0

ON THE RANDOMNESS OF EIGENVECTORS GENERATED FROM NETWORKS WITH RANDOM TOPOLOGIES*

JACK W. SILVERSTEIN†

Abstract. A model for the generation of neural connections at birth led to the study of W, a random, symmetric, nonnegative definite linear operator defined on a finite, but very large, dimensional Euclidean space [1]. A limit law, as the dimension increases, on the eigenvalue spectrum of W was proven, implying that realizations of W (being identified with organisms in a species) appear totally different on the microscopic level and yet have almost identical spectral densities.

The present paper considers the eigenvectors of W. Evidence is given to support the conjecture that, contrary to the deterministic aspect of the eigenvalues, the eigenvectors behave in a completely chaotic manner, which is described in terms of the normalized uniform (Haar) measure on the group of orthogonal transformations on a finite dimensional space. The validity of the conjecture would imply a *tabula rasa* property on the ensemble ("species") of all realizations of W.

1. The conjecture. In [1] we presented neural networks constructed at birth. The connections between neurons are generated according to a controlled probability model, the assumptions being consistent with known limitations on genetic coding. This led to the study of W, a random, symmetric, nonnegative definite linear operator defined on n-dimensional space, where n is very large. Each realization of W can be viewed as describing the neural state of a particular organism from some population or species. A limit law ($n \to \infty$) on the eigenvalue spectrum was proven, which implies that the spectrum of W is close to a fixed one. Thus, realizations of W look totally different on the microscopic level, yet have almost identical spectral densities.

Let us briefly mention how W is constructed. For a given positive integer d, an $n \times dn$ matrix $P = (P_{ij})$ of probabilities which is formed under rather general conditions, is used to construct the random $n \times dn$ matrix $V = (v_{ij})$. All elements of V are independent, and each v_{ij} is either 1 or -1 with equal probability, or zero, and Prob $(v_{ij}^2 = 1) = P_{ij}$. Each row of P is a rotation of the first row. Let C be the sum of the first row of P. Then

$$W = \frac{1}{C} VV^T. \tag{1}$$

The matrix V contains the connection characteristics between a group of n neurons synapsing onto another group of dn neurons. Thus, the synaptic strength between neuron i in the first group and neuron j in the second group is given by V_{ij}. The structure of P allows the model to have a homogeneous behavior in the establishment of connections. A spatial consideration can also be imposed on the network by choosing the first row of P accordingly. However, the only condition needed for the limit law is to have $C \to \infty$ as $n \to \infty$.

In the theory developed for neural networks [2], [3] patterns of the external world are represented as vectors, and become input to the network. The complete spectral decomposition of W reveals the initial response characteristics of the network on its environment. For example, at birth the network will be predisposed toward the eigenvectors of W with large eigenvalues. Since W is randomly generated, the assignment of eigenvectors to eigenvalues will vary from one realization of W to another. The manner in which the eigenvectors vary is the main topic of the paper.

* Received by the editors December 12, 1977, and in revised form October 2, 1978.
† Department of Mathematics, North Carolina State University, Raleigh, North Carolina 27650. This research was supported by the National Science Foundation under Grant MCS75-15153-A01.

We will provide evidence supporting the conjecture that the behavior of the eigenvectors is completely chaotic. An attempt at formalizing this conjecture is the following: Let $O(=O(n))$ be a random $n \times n$ orthogonal matrix distributed according to the normalized uniform (Haar) measure on the orthogonal group $\mathcal{O}_n$. Let D_n be a fixed $n \times n$ diagonal matrix with nonnegative diagonal elements arranged in nondecreasing order and such that the spectrum of D_n approaches the limiting spectrum of W as $n \to \infty$. The conjecture is that for n large the distribution of $W' \equiv OD_nO^T$ is close (in some sense) to the distribution of W.

In connection with neural networks it is shown in [1] that W can be made as close to the identity matrix as desired by choosing first d sufficiently large and then n sufficiently large. This implies a *tabula rasa* property on each realization of W: no input from the external world is preferred at birth. When d is not very large, any realization of W is genetically biased toward certain inputs and against others. However, the validity of the conjecture would imply that no genetic bias to the external world is preferred by the *ensemble* of all realizations of W. Thus, in biological terms, a *tabula rasa* property would still exist, but now on the "species" of organisms rather than for an individual organism.

In relation to real biological systems, the reader is referred to the remarks made in § 3 of [1]. The present conclusion is not intended to support *tabula rasa*, but rather to demonstrate what properties an ensemble of networks can possess under the particular assumptions making up the controlled probability model. It is felt, however, that this mathematical approach to modeling the generation of neural connections can be of some help in understanding the development of several substructures in the nervous system.

The evidence is based on properties of real valued functions defined on the spectral family $\{P_a^M\}$ of $M = W$ or W', where for each $a \in [0, \infty)$, P_a^M is the projection operator of the space spanned by the eigenvectors of M with eigenvalues $\leqq a$. The results, stated and proven in the next section, are limit theorems, the convergence being in probability (i.p.) as $n \to \infty$, and they show the limits to be the same for W and W'. Theorems 1 and 3 are concerned with the limiting value of $x^TP_a^Mx$, where for each n, $x \in R^n$ is an arbitrary but fixed, unit vector. They assert that for $M = W$ or W', $x^TP_a^Mx \xrightarrow{\text{i.p.}} F(a)$ as $n \to \infty$, where F is the limiting spectral distribution function of W.

This property of W is probably not enough to prove the conjecture but it does help in strengthening our belief. It enables us to rule out possibilities. In particular, we can eliminate the case that at least one eigenvector is fixed but might be assigned to different eigenvalues from one realization of W to another. This is so since if x is one of these eigenvectors $x^TP_a^Wx$ is either 0 or 1. The mean and second moment of this quantity are then equal so that the only limiting values that can be reached in probability would be 0 or 1. (Since $x^TP_a^Wx$ is bounded between 0 and 1, convergence in probability to one value implies the variance approaches 0.) The distribution function F is continuous so that a continuum of values must be taken on.

We also considered the quantity $H(M_1, M_2) \equiv (1/n) \operatorname{tr} (M_1 - M_2)^2$, where tr is the trace function, originally as a measure of differences between eigenspaces of two realizations W_1, W_2 of W. Small values of $H(P_a^{W_1}, P_a^{W_2})$ would indicate that the eigenspace corresponding to eigenvalues of W in $[0, a]$ does not vary greatly, whereas not-so-small values would suggest the opposite. At the same time we used H as a possible means of supporting the conjecture. Theorem 2 shows that for two independent realizations W_1', W_2' of W', $H(P_a^{W_1'}, P_a^{W_2'}) \xrightarrow{\text{i.p.}} h(a) \equiv 2F(a)(1-F(a))$. The graph of h for $d = 1$ is displayed in Fig. 1. Two computer simulations were carried out to calculate values of $H(P_a^{W_1}, P_a^{W_2})$.

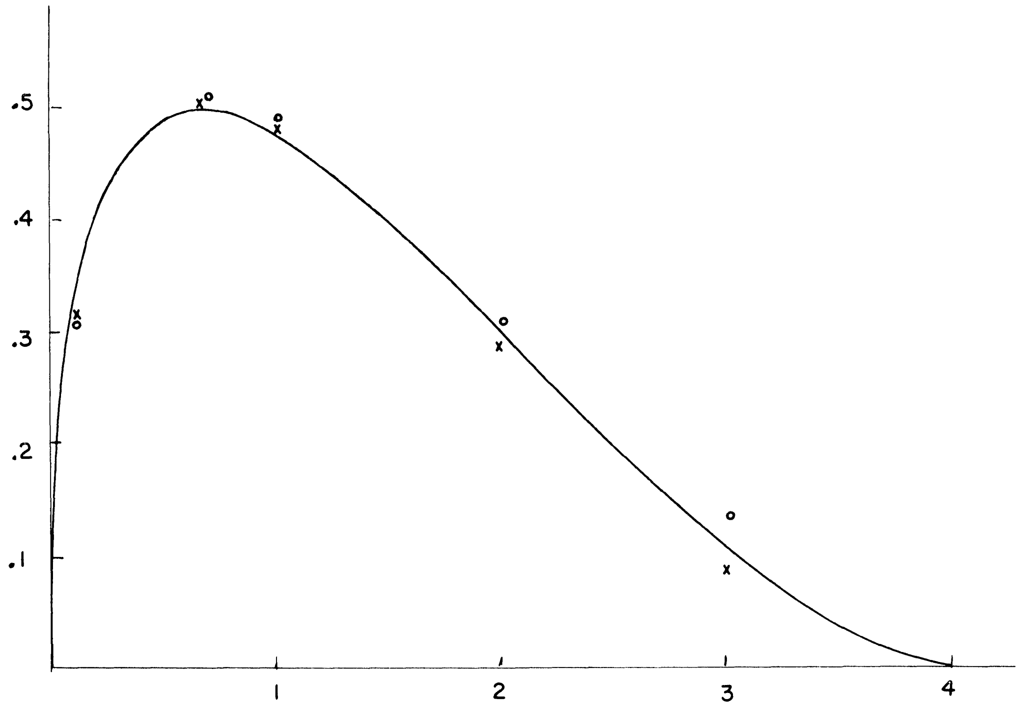

FIG. 1. *Solid curve is $h(a)$ ($d = 1$). $\times$ and $\circ$ denote two separate simulations of $H(P_a^{W_1'}, P_a^{W_2'})$ ($d = 1$, $n = 100$).*

For each simulation ($n = 100$) $h(a)$ was determined for five values of a, and these points are plotted in Fig. 1. The agreement is quite good.

We then proceeded to establish a proof that $H(P_a^{W_1}, P_a^{W_2}) \xrightarrow{\text{i.p.}} h(a)$ as $n \to \infty$. This result is Theorem 4 in the next section.

Since the complete spectral decomposition of an operator in finite dimensional space can be gotten from its spectral family, we feel that the results characterize completely the spectral family of W, so that the conjecture is likely to be true. We are currently trying to prove this. Since the probability space of $n \times n$ matrices is changing for increasing n, it appears to be difficult to establish the result in terms of a limit theorem. However, there are ways of formalizing the conjecture without the need for an underlying space. Let μ_n and μ_n' be the Borel measures generating W and W' respectively. Let B_n be some class of Borel sets of $n \times n$ matrices whose description does not depend upon n, and which generates all Borel sets. For example, B_n could be the compact sets, or perhaps the complete set of Borel sets. Then one could try to prove a statement such as the following: for any $\varepsilon > 0$, there exists an N such that for all $n \geqq N$, $|\mu_n(A) - \mu_n'(A)| < \varepsilon$ for all $A \in B_n$. In view of the current state of probability theory this is a strong statement concerning two measures which is usually difficult to prove. Nonetheless, in our case we remain optimistic.

2. The results. Theorems 3 and 4 rely heavily on the arguments used and the results established in [1]. We will refer back to this paper when necessary. Therefore it will be assumed that the reader is familiar with [1]. Theorems 1 and 2 use the following properties of the uniform (Haar) measure on $\mathcal{O}_n$:

1. If $O \in \mathcal{O}_n$ is uniformly distributed, then O^T is uniformly distributed.

2. For fixed $A \in \mathcal{O}_n$ and $O \in \mathcal{O}_n$ uniformly distributed, $A0$ and $0A$ are both uniformly distributed.

3. If $O_1 \in \mathcal{O}_n$ and $O_2 \in \mathcal{O}_n$ are independent and uniformly distributed then O_1O_2 is uniformly distributed.

4. If $x \in R^n$ is a fixed unit vector, and $O \in \mathcal{O}_n$ is uniformly distributed, then the distribution of Ox on S_{n-1}, the surface of the unit sphere in R^n, is uniform, i.e. generated from the surface element on S_{n-1}.

The first three properties follow almost directly from the invariant property of the Haar measure on $\mathcal{O}_n$. The fourth one seems just as obvious but is not as easy to prove. We will give a proof using spherical harmonics and the associated addition theorem (a good reference for this topic is [4]). A spherical harmonic is the restriction to S_{n-1} of a homogeneous polynomial defined on R^n satisfying Laplace's equation. A basis for $L^2(S_{n-1}, \omega)$, where ω denotes uniform measure, can be made consisting entirely of spherical harmonics. Let $S = \{S_{m,j}: j = 1, \cdots N(m); m = 0, 1, 2, \cdots\}$ denote such a basis. Here, $S_{m,j}$ is a spherical harmonic of degree m and $N(m)$ is the number of linearly independent spherical harmonics of degree m. The addition theorem implies that for each m and $y, z \in S_{n-1}$

$$\sum_{j=1}^{N(m)} S_{m,j}(y)S_{m,j}(z) \tag{2.1}$$

depends only on the value of $\langle y, z\rangle$. Let (2.1) be equal to $\phi(\langle y, z\rangle)$.

Another important property of S is that the span of S is dense in $C(S_{n-1})$, the space of continuous functions on S_{n-1}, with respect to the supremum norm. Therefore, the span of S is dense in $L^1(S_{n-1}, \mu)$, where μ is an arbitrary Borel measure on S_{n-1} (S_{n-1} being a compact Hausdorff space implies $C(S_{n-1})$ is dense in $L^1(S_{n-1}, \mu)$), so that μ is uniquely determined by the values

$$C_{m,j} \equiv \int_{S_{n-1}} S_{m,j}(y)\, d\mu(y), \qquad j = 1, 2, \cdots, N(m), \quad m = 0, 1, 2, \cdots. \tag{2.2}$$

Let μ now be the measure on S_{n-1} generated by Ox. Since this mapping from $\mathcal{O}_n$ onto S_{n-1} is continuous, μ is a Borel measure. Also, μ is rotationally invariant, i.e. for any Borel set $A \subset S_{n-1}$ and any fixed $O \in \mathcal{O}_n$, $\mu(0A) = \mu(A)$. We will show that these two properties uniquely characterize μ.

For $z_1, z_2 \in S_{n-1}$ let $O \in \mathcal{O}_n$ be such that $Oz_2 = z_1$. Then

$$\begin{aligned}\int_{S_{n-1}} \phi(\langle y, z_1\rangle)\, d\mu(y) &= \int_{S_{n-1}} \phi(\langle y, Oz_2\rangle)\, d\mu(y) = \int_{S_{n-1}} \phi(\langle O^T y, z_2\rangle)\, d\mu(y) \\ &= \int_{S_{n-1}} \phi(\langle y, z_2\rangle)\, d\mu(y).\end{aligned} \tag{2.3}$$

The last equality is due to the fact that μ is rotationally invariant. Therefore integrating (2.1) with respect to y will result in a linear combination of orthonormal spherical harmonics equalling a constant, which is also a spherical harmonic. Therefore for all $b \geqq 1$, $C_{m,j} = 0$, thus establishing the uniqueness of a Borel, rotationally invariant probability measure on S_{n-1}.

THEOREM 1. *For each n let $x \in R^n$ be a fixed unit vector. Then for any $a \in [0, \infty)$*

$$x^T P_a^{W'} x \xrightarrow{\text{i.p.}} F(a) \quad \text{as } n \to \infty \tag{2.4}$$

Proof. Let m_n be the number of eigenvalues of W' less than or equal to a. Let $y = O^T x$. Then

$$x^T P_a^{W'} x = \sum_{i=1}^{m_n} y_i^2. \tag{2.5}$$

By property 4, y is uniformly distributed on the unit sphere. This random vector can be generated by normalizing an $N(0, I)$ distributing vector (I being the $n \times n$ identity matrix). Therefore $x^T P_a^{W'} x$ is beta distributed with parameters $p = m_n/2$, $q = (n - m_n)/2$, so that

$$E(x^T P_a^{W'} x) = \frac{p}{p+q} = \frac{m_n}{n} \to F(a) \quad \text{as } n \to \infty \tag{2.6}$$

and

$$\mathrm{Var}\,(x^T P_a^{W'} x) = \frac{pq}{(p+q)^2(p+q+1)} = \frac{2m_n}{n^2}\left(\frac{n-m_n}{n+2}\right) \to 0 \quad \text{as } n \to \infty \tag{2.7}$$

which implies (2.4).

THEOREM 2. *If W_1' and W_2' are two independent generations of W', then for every pair of nonnegative a_1, a_2*

$$\frac{1}{n} \mathrm{tr}\,(P_{a_1}^{W_1'} - P_{a_2}^{W_2'})^2 \xrightarrow{\text{i.p.}} F(a_1) + F(a_2) - 2F(a_1)F(a_2). \tag{2.8}$$

Proof. Let m_n^i = number of eigenvalues of $W_i' \leqq a_i$, $i = 1, 2$. By Property 3 and the invariant property of the trace under similarity transformations we have

$$\frac{1}{n} \mathrm{tr}\,(P_{a_1}^{W_1'} - P_{a_2}^{W_2'})^2 = \frac{1}{n} \mathrm{tr}\,(D - P_{a_2}^{W_3'})^2 \tag{2.9}$$

where D is the diagonal matrix having its first m_n^1 diagonal entries equal to one and the rest 0, and W_3' is another W' generation. Simplifying further we have

$$\begin{aligned} \frac{1}{n} \mathrm{tr}\,(D - P_{a_2}^{W_3'})^2 &= \frac{m_n^1}{n} + \frac{m_n^2}{n} - \frac{1}{n} \mathrm{tr}\, D P_{a_2}^{W_3'} - \frac{1}{n} \mathrm{tr}\, P_{a_2}^{W_3'} D \\ &= \frac{m_n^1}{n} + \frac{m_n^2}{n} - \frac{2}{n} \sum_{i=1}^{m_n^1} (P_{a_2}^{W_3'})_{ii}. \end{aligned} \tag{2.10}$$

Let $O = (O_{ij})$ be the orthogonalizing matrix of W_3', and let $O_{\cdot j}$ denote the jth column of O. Then

$$P_{a_2}^{W_3'} = \sum_{i=1}^{m_n^2} O_{\cdot j} O_{\cdot j}^T \tag{2.11}$$

and

$$\frac{1}{n} \sum_{i=1}^{m_n^1} (P_{a_2}^{W_3'})_{ii} = \frac{1}{n} \sum_{i=1}^{m_n^1} \sum_{j=1}^{m_n^2} O_{ij}^2. \tag{2.12}$$

Because of properties 1 and 4 each row and column of O is uniformly distributed on S_{n-1}. Using the formulas in (2.6) and (2.7) for the mean and variance of a beta distribution we get

$$E\left(\frac{1}{n} \sum_{i=1}^{m_n^1} (P_{a_2}^{W_3'})_{ii}\right) = \frac{m_n^1}{n} \frac{m_n^2}{n} \to F(a_1)F(a_2) \quad \text{as } n \to \infty \tag{2.13}$$

and

$$E\left(\left(\sum_{j=1}^{m_n^2} O_{ij}^2\right)^2\right)=\frac{m_n^2}{n}\left(\frac{m_n^2+2}{n+2}\right) \tag{2.14}$$

and

$$E(O_{ij}^2 O_{i'j'}^2)=\frac{1}{n(n+2)} \quad \text{for } i=i', j\neq j'; \quad \text{or } i\neq i', j=j'. \tag{2.15}$$

By property 2 all pairs $(O_{ij}, O_{i'j'})$ for $i\neq i'$, $j\neq j'$, are identically distributed since by elementary transformations each can be brought to the (1, 1) and (2, 2) positions. The expected value of $O_{11}^2 O_{22}^2$ can be derived from the identity

$$1=\left(\sum_{i=1}^{n} O_{i1}^2\right)\left(\sum_{j=1}^{n} O_{j2}^2\right)=E\left(\left(\sum_{i=1}^{n} O_{i1}^2\right)\left(\sum_{j=1}^{n} O_{j2}^2\right)\right). \tag{2.16}$$

Using (2.15) we have

$$1=n\times\frac{1}{n(n+2)}+n(n-1)E(O_{11}^2 O_{22}^2) \tag{2.17}$$

so that

$$E(O_{11}^2 O_{22}^2)=\frac{n+1}{n(n-1)(n+2)}. \tag{2.18}$$

Therefore, using (2.15) and (2.18) we have

$$\begin{aligned} E\left(\left(\sum_{j=1}^{m_n^2} O_{1j}^2\right)\left(\sum_{j'=1}^{m_n^2} O_{2j}^2\right)\right)&=\frac{m_n^2}{n(n+2)}+m_n^2(m_n^2-1)\frac{(n+1)}{n(n-1)(n+2)}\\ &=\frac{m_n^2}{n(n+2)}\left(1+(m_n^2-1)\frac{(n+1)}{(n-1)}\right) \end{aligned} \tag{2.19}$$

and finally, using (2.12), (2.14), and (2.19)

$$\begin{aligned} E\left(\left(\frac{1}{n}\sum_{i=1}^{m_n^1}(P_{a_2}^{W_3'})_{ii}\right)^2\right)&=\frac{1}{n^2}\left(\frac{m_n^1 m_n^2}{n}\left(\frac{m_n^2+2}{n+2}\right)+\frac{m_n^1(m_n^1-1)m_n^2}{n(n+2)}\left(1+(m_n^2-1)\frac{(n+1)}{(n-1)}\right)\right)\\ &\to (F(a_1)F(a_2))^2 \qquad \text{as } n\to\infty \end{aligned} \tag{2.20}$$

Therefore the variance of (2.12) goes to zero in the limit which implies the convergence of (2.12) to $F(a_1)F(a_2)$ in probability. This together with (2.10) gives the result.

THEOREM 3. *For each n let $x\in R^n$ be a fixed unit vector. Then for any $a\in[0,\infty)$*

$$x^T P_a^W x \xrightarrow{\text{i.p.}} F(a) \quad \text{as } n\to\infty. \tag{2.21}$$

Proof. Let g be the density of F defined on $[0,\infty)$ (in [1] g was defined only on the support of F'). As in [1] the rth moment of F is denoted by $f(r)$.

We have

$$E((W^r)_{ab})=\frac{1}{C^r}\sum_{\substack{i_2\cdots i_r\\ k_1\cdots k_r}} v_{ak_1}v_{i_2k_2}\cdots v_{i_rk_r}v_{bk_r} \tag{2.22}$$

when $a \neq b$ the number of v_{ik}'s in each term of (2.22) having $i = a$ must be odd. Therefore in this case each term in (2.22) is zero so that

$$E((W^r)_{ab}) = 0, \qquad a \neq b. \tag{2.23}$$

It is evident that the distribution of $(W^r)_{ii}$ is independent of i. Since $E((1/n) \operatorname{tr} W^r) \to f(r)$ as $n \to \infty$ we have $E((W^r)_{11}) \to f(r)$ also. Therefore

$$E(x^T W^r x) = E\Big(\sum_i x_i^2 (W^r)_{ii}\Big) = E((W^r)_{11}) \to f(r) \quad \text{as } n \to \infty. \tag{2.24}$$

We find that

$$E((x^T W^r x)^2) = \frac{1}{C^{2r}} \sum_{\substack{ab \\ a'b'}} x_a x_b x_{a'} x_{b'} \sum_{\substack{i_2 \cdots i_r \\ k_1 \cdots k_r \\ i_2' \cdots i_r' \\ k_1' \cdots k_r'}} E(v_{ak_1} v_{i_2 k_1} \cdots v_{bk_r} v_{a'k_1'} \cdots v_{b'k_r'}). \tag{2.25}$$

Again we see the only nonzero term occurs when a, b, a', b' pair up. Therefore

$$\begin{aligned} E((x^T W^r x)^2) = {} & \frac{1}{C^{2r}} \sum_{ab} x_a^2 x_b^2 \sum_{\substack{i_2 \cdots i_r \\ k_1 \cdots k_r \\ i_2' \cdots i_r' \\ k_1' \cdots k_r'}} E(v_{ak_1} v_{i_2 k_1} \cdots v_{ak_r} v_{bk_1'} \cdots v_{bk_r'}) \\ & + \frac{1}{C^{2r}} \sum_{a \neq b} x_a^2 x_b^2 \sum_{\substack{i_2 \cdots i_r \\ k_1 \cdots k_r \\ i_2' \cdots i_r' \\ k_1' \cdots k_r'}} E(v_{ak_1} \cdots v_{bk_r} v_{ak_1'} \cdots v_{bk_r'}) \\ & + \frac{1}{C^{2r}} \sum_{a \neq b} x_a^2 x_b^2 \sum_{\substack{i_2 \cdots i_r \\ k_1 \cdots k_r \\ i_2' \cdots i_r' \\ k_1' \cdots k_r'}} E(v_{ak_1} \cdots v_{bk_r} v_{bk_1} \cdots v_{ak_r}). \end{aligned} \tag{2.26}$$

We can use the same reasoning as in [1]. We consider the constraints on the indices that contribute a nonnegligible amount in the limit. From Lemma 1 of [1] we find that each of these constraints will not pair a v_{ik} with a $v_{i'k'}$, which implies that the last two terms in (2.26) do not contribute anything in the limit. Therefore

$$\begin{aligned} E((x^T W^r x)^2) & \sim \frac{1}{C^{2r}} \sum_{ab} x_a^2 x_b^2 \sum_{\substack{i_2 \cdots i_r \\ k_1 \cdots k_r \\ i_2' \cdots i_r' \\ k_1' \cdots k_r'}} E(v_{ak_1} \cdots v_{ak_r}) E(v_{ak_1'} \cdots v_{ak_r'}) \\ & = [E(x^T W^r x)]^2. \end{aligned} \tag{2.27}$$

We conclude that for $r = 1, 2, \cdots$

$$x^T W^r x \xrightarrow{\text{i.p.}} f(r) \quad \text{as } n \to \infty. \tag{2.28}$$

Since $f(r) = \int_0^\infty u^r g(u)\, du$ we have for any polynomial $P(x)$

$$x^T P(W) x \xrightarrow{\text{i.p.}} \int_0^\infty P(u) g(u)\, du. \tag{2.29}$$

If the eigenvalues of W were known to be uniformly bounded for all n almost surely, then we can simply approximate the indicator function on $[0, a]$ by polynomials and obtain the result. However, we are unable to show this. The following argument therefore seems necessary.

For $0 < a < b$ let $R_{a,b}(x)$ be the ramp function which is 1 for $0 \leqq x \leqq a$, zero for $x \geqq b$ and linearly decreasing from a to b. Let $\{U_m(x)\}$ be a sequence of nonnegative polynomials such that $U_m(x) \geqq R_{a,b}(x)$ for all $x \geqq 0$, and $U_m(x) \downarrow R_{a,b}(x)$ for $0 \leqq x \leqq b$. This sequence can be formed, for example, by first considering a sequence $\{U_m^1(x)\}$ of ramp functions, all upper translates of $R_{a,b}$, with $U_m^1 < U_{m-1}^1$ and $U_m^1 \downarrow R_{a,b}$ on $[0, b]$. We can then find polynomials $\{p_m\}$ where each p_m approximates $\sqrt{U_m^1}$ sufficiently closely so that we can let $U_m = p_m^2$.

It is also possible to construct a sequence of polynomials $\{L_m(x)\}$ such that $L_m(x) \leqq R_{a,b}(x)$ for $x \geqq 0$, and $L_m(x) \uparrow R_{a,b}(x)$ for $0 \leqq x \leqq b$. Start with a sequence $\{L_m^1\}$ of nonincreasing C^1 functions defined on $[0, b]$ such that $L_m^1 \uparrow R_{ab}$ and

$$\inf_{x \in [0,b]} |L_m^1(x) - L_{m-1}^1(x)| > 0 \quad \text{for } n > 1. \tag{2.30}$$

For fixed m let $\varepsilon > 0$ be the closest distance L_m^1 has with L_{m-1}^1 and L_{m+1}. Let p_m be a polynomial such that

$$|p_m(x) - \sqrt{-L_m^{1'}(x)}| < \min\left(1, \frac{\varepsilon}{2b(1 + 2\max_{x \in [0,b]} \sqrt{-L_m^{1'}(x)})}\right) \tag{2.31}$$

for all $x \in [0, b]$. Then $L_m(x) \equiv L_m^1(0) - \int_0^x p_m^2(y)\,dy$ is a nonincreasing polynomial with

$$\begin{aligned} |L_m(x) - L_m^1(x)| &= \left| -\int_0^x p_m^2(y)\,dy - \int_0^x L_m^{1'}(y)\,dy \right| \\ &= \left| \int_0^x p_m^2(y)\,dy - \int_0^x (\sqrt{-L_m^{1'}(y)})^2\,dy \right| \\ &\leqq \int_0^x |p_m(y) - \sqrt{-L_m^{1'}(y)}|\,|p_m(y) + \sqrt{-L_m^{1'}(y)}|\,dy. \end{aligned} \tag{2.32}$$

Since $|p_m - \sqrt{-L_m^{1'}}| < 1$, we have

$$\begin{aligned} |p_m(y) + \sqrt{-L_m^{1'}(y)}| &\leqq |p_m(y)| + |\sqrt{-L_m^{1'}(y)}| \\ &\leqq 2\sqrt{-L_m^{1'}(y)} + 1 \end{aligned} \tag{2.33}$$

for $y \in [0, b]$. Therefore

$$|L_m(x) - L_m^1(x)| \leqq b\frac{\varepsilon}{2b} = \frac{\varepsilon}{2} \tag{2.34}$$

so that $L_m(x) < L_{m+1}(x) < R_{a,b}(x)$. With $\{L_m\}$ defined in this way we have $L_m(x) \uparrow R_{a,b}(x)$ for all $x \in [0, b]$.

We have for all m, n

$$x^T L_m(W) x \leqq x^T R_{a,b}(W) x \leqq x^T U_m(W) x. \tag{2.35}$$

Given any subsequence $\{n_j\}_{j=1}^{\infty}$ we can find, using a diagonal selection argument, a subsequence $\{n_j'\} \subseteq \{n_j\}$ such that for each m

$$x^T L_m(W)x \xrightarrow{\text{a.s.}} \int_0^\infty L_m(u)g(u)\,du \tag{2.36}$$

and

$$x^T U_m(W)x \xrightarrow{\text{a.s.}} \int_0^\infty U_m(u)g(u)\,du \tag{2.37}$$

along $\{n_j'\}$. We have, on this subsequence outside of a set of measure zero

$$\int_0^\infty L_m(u)g(u)\,du \leqq \underline{\lim}\, x^T R_{a,b}(W)x \leqq \overline{\lim}\, x^T R_{a,b}x \leqq \int_0^\infty U_m(u)g(u)\,du. \tag{2.38}$$

Using the monotone convergence theorem the two extremes approach the same value, so that

$$x^T R_{a,b}(W)x \xrightarrow{\text{a.s.}} \int_0^\infty R_{a,b}(u)g(u)\,du \quad \text{on } \{n_j'\}. \tag{2.39}$$

Since the original subsequence was arbitrary we have

$$x^T R_{a,b}(W)x \xrightarrow{\text{i.p.}} \int_0^\infty R_{a,b}(u)g(u)\,du. \tag{2.40}$$

For a_1 such that $0 \leqq a_1 < a$ we have

$$x^T R_{a_1,a}(W)x \leqq x^T P_a^W x \leqq x^T R_{a,b}(W)x. \tag{2.41}$$

Using a similar argument as above we can therefore conclude that

$$x^T P_a^W x \xrightarrow{\text{i.p.}} \int_0^a g(u)\,du. \tag{2.42}$$

THEOREM 4. *Let W_1 and W_2 be two independent generations of W. Then, for every pair of nonnegative a_1, a_2*

$$\frac{1}{n}\operatorname{tr}(P_{a_1}^{W_1} - P_{a_2}^{W_2})^2 \xrightarrow{\text{i.p.}} F(a_1) + F(a_2) - 2F(a_1)F(a_2) \quad \text{as } n \to \infty. \tag{2.43}$$

Proof. We have

$$\begin{aligned} \frac{1}{n}\operatorname{tr}(P_{a_1}^{W_1} - P_{a_2}^{W_2})^2 &= \frac{1}{n}(\text{no. of eigenvalues of } W_1 \leqq a_1) \\ &\quad + \frac{1}{n}(\text{no. of eigenvalues of } W_2 \leqq a_2) \\ &\quad - \frac{1}{n}\operatorname{tr} P_{a_1}^{W_1} P_{a_2}^{W_2} - \frac{1}{n}\operatorname{tr} P_{a_2}^{W_2} P_{a_1}^{W_1}. \end{aligned} \tag{2.44}$$

From [1] we know that the first two terms on the right hand side of (2.44) converge in probability to $F(a_1)$ and $F(a_2)$ respectively. The remaining two terms can be handled with the aid of methods similar to those used in Theorem 3.

For any pair of positive integers r_1 and r_2 we have

$$E\left(\frac{1}{n}\operatorname{tr} W_1^{r_1}W_2^{r_2}\right)=\frac{1}{n}\operatorname{tr} E(W^{r_1})E(W^{r_2}) \tag{2.45}$$
$$=\frac{1}{n}\sum_{i=1}^{n} E(W_{ii}^{r_1})E(W_{ii}^{r_2})$$
$$=E(W_{11}^{r_1})E(W_{11}^{r_2})\to f(r_1)f(r_2) \quad \text{as } n\to\infty,$$

the second equality holding because of (2.23). We also have

$$E\left(\left(\frac{1}{n}\operatorname{tr} W_1^{r_1}W_2^{r_2}\right)^2\right) \tag{2.46}$$
$$=\frac{1}{n^2}\sum_{\substack{i_1k_1\\ i_2k_2}} E((W_1^{r_1})_{i_1k_1}(W_1^{r_1})_{i_2k_2}(W_2^{r_2})_{i_1k_1}(W_2^{r_2})_{i_2k_2}$$
$$=\frac{1}{n^2}\sum_{\substack{i_1k_1\\ i_2k_2}} E(W_{i_1k_1}^{r_1}W_{i_2k_2}^{r_1})E(W_{i_1k_1}^{r_2}W_{i_2k_2}^{r_2}).$$

As in [1] and in Theorem 3 we can break down the last expression in (2.46) in terms of the v_{ik}'s and consider constraints on the indices contributing a nonnegligible amount in the limit. Lemma 1 of [1] will enable us to conclude that

$$E\left(\left(\frac{1}{n}\operatorname{tr} W_1^{r_1}W_2^{r_2}\right)^2\right)\sim\frac{1}{n^2}\sum_{ab} E(W_{aa}^{r_1})E(W_{bb}^{r_1})E(W_{aa}^{r_2})E(W_{bb}^{r_2}) \tag{2.47}$$
$$=\left(\frac{1}{n}\sum_{a} E(W_{aa}^{r_1})E(W_{aa}^{r_2})\right)^2\to(f(r_1)f(r_2))^2.$$

Therefore

$$\frac{1}{n}\operatorname{tr} W_1^{r_1}W_2^{r_2}\xrightarrow{\text{i.p.}} f(r_1)f(r_2) \quad \text{as } n\to\infty \tag{2.48}$$

and for arbitrary polynomials P_1, P_2

$$\frac{1}{n}\operatorname{tr} P_1(W_1)P_2(W_2)\xrightarrow{\text{i.p.}}\left(\int_0^\infty P_1(u)g(u)\,du\right)\left(\int_0^\infty P_2(u)g(u)\,du\right). \tag{2.49}$$

The following statement can easily be proven: given A, B, A', B' all $n\times n$ symmetric, positive semi-definite matrices with the eigenvalues of A and B being less than or equal to the eigenvalues of A' and B' respectively. Then

$$\operatorname{tr} AB\leqq\operatorname{tr} A'B'. \tag{2.50}$$

Therefore we can use the same polynomials approximating ramp functions in Theorem 3 and use (2.50) to assert inequalities analogous to (2.35) and (2.41) and use the same argument to conclude that

$$\frac{1}{n}\operatorname{tr} P_{a_1}^{W_1}P_{a_2}^{W_2}\xrightarrow{\text{i.p.}} F(a_1)F(a_2) \quad \text{as } n\to\infty. \tag{2.51}$$

This together with (2.44) yields (2.43).

REFERENCES

[1] U. GRENANDER AND J. W. SILVERSTEIN, *Spectral analysis of networks with random topologies*, this Journal, 32 (1977), pp. 499–519.

[2] J. W. SILVERSTEIN, *Asymptotics applied to a neural network*, Biol. Cybernetics, 22 (1976), pp. 73–84.

[3] U. GRENANDER, *Pattern Analysis: Lectures in Pattern Theory*, vol. 2, Springer-Verlag, Berlin, 1978.

[4] C. MÜLLER, *Spherical Harmonics*, Lecture Notes in Mathematics 17, Springer-Verlag, Berlin, 1966.