SIAM J. APPL. MATH.
Vol. 32, No. 2, March 1977

SPECTRAL ANALYSIS OF NETWORKS WITH RANDOM TOPOLOGIES*

ULF GRENANDER AND JACK W. SILVERSTEIN†

Abstract. A class of neural models is introduced in which the topology of the neural network has been generated by a controlled probability model. It is shown that the resulting linear operator has a spectral measure that converges in probability to a universal one when the size of the net tends to infinity: a law of large numbers for the spectra of such operators. The analytical treatment is accompanied by computational experiments.

1. The network. It is known that neural networks exhibit a great deal of regularity. The topology of the network, which describes how its cells are connected to each other, seems to be genetically determined at least on a global level, so that the connections are certainly not completely random. On the other hand, the details of the topology may vary from individual to individual within the same species for higher animals. A model of such networks will therefore involve controlled probabilities for connection, where probabilities are not all the same but are controlled by distance and possibly other characteristics.

We shall investigate the spectral properties of networks based on such a model and viewed as linear operators. When a signal is applied to the network the response can be expressed in terms of the spectral properties of the operator, and we will show that the spectral measure converges for increasing size to a universal limit under weak conditions that will be made precise below.

The practical implication of this is that a *law of large numbers for spectra* exists for such networks, so that when the size becomes large the influence of the randomness in the topology will tend to zero. Before we begin the proof of this result (the theorem in § 2) we shall make some preliminary observations. Our initial guess, based on these observations, was that the limit was a one-point measure, a single spike. Computational experiments indicated that this was not the case, however, and we are led to a quite different conjecture formulated in the theorem.

Let us consider a network consisting of two sets of nodes, the inputs enumerated by $i = 1, 2, \cdots, n$, and the outputs enumerated by $j = 1, 2, \cdots, m$. These nodes are the cells, the neurons, of the network. Synaptic connections are established between some of the i-nodes and some of the j-nodes. There are no connections between i-nodes and none between j-nodes.

Obviously real neural nets are quite heterogeneous and have a characteristic three-dimensional architecture. To simplify the analysis, which will not be easy anyway, we neglect heterogeneity and the three-dimensional aspect in this paper, and hope to return to the more realistic case in a later publication.

The values of n and m are very large and $m = dn$, where the *divergence coefficient* d is assumed to be an integer. We are particularly interested in diverging nets, $d > 1$.

* Received by the editors February 25, 1976.

† Division of Applied Mathematics and Center for Neural Studies, Brown University, Providence, Rhode Island 02912. This work was supported by the National Science Foundation under Grant MPS75-15153.

A set of probabilities $\{p_h^n;\ h=-s,-s+1,\cdots,s\}$ controls the establishment of connections. For a given value of i there will be a connection $i\to j=(i+h)\times d-e$ $(e=0,1,2,\cdots,d-1)$ with probability

$$P_{ij}^n=\begin{cases} p_{h \bmod n}^n & \text{if} \quad h \bmod n \leqq s,\\ p_{(h \bmod n)-n} & \text{if} \quad (h \bmod n)-n \geqq -s,\\ 0 & \text{otherwise.}\end{cases}$$

The number $s<n/2$ is called the *spread* of the network. Figure 1.1 gives an example of connections with $d=3$. Figure 1.2 gives the matrix of probabilities of connections for the case when $n=5$, $d=2$, $s=1$. Notice the cyclic arrangement of the probabilities. It is imposed only for mathematical convenience so that unnecessary complications can be avoided.

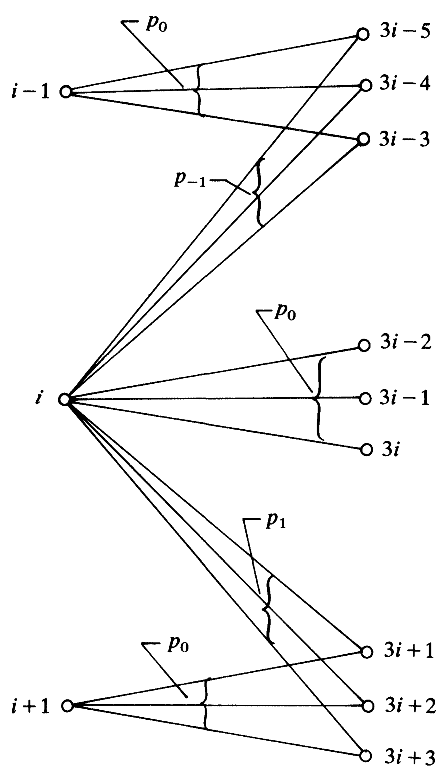

FIG. 1.1

$$\begin{array}{cc} & j \\ i & \begin{bmatrix} p_0 & p_0 & p_1 & p_1 & 0 & 0 & 0 & 0 & p_{-1} & p_{-1}\\ p_{-1} & p_{-1} & p_0 & p_0 & p_1 & p_1 & 0 & 0 & 0 & 0\\ 0 & 0 & p_{-1} & p_{-1} & p_0 & p_0 & p_1 & p_1 & 0 & 0\\ 0 & 0 & 0 & 0 & p_{-1} & p_{-1} & p_0 & p_0 & p_1 & p_1\\ p_1 & p_1 & 0 & 0 & 0 & 0 & p_{-1} & p_{-1} & p_0 & p_0 \end{bmatrix}\end{array}$$

FIG. 1.2

The *average connectivity* C_n is then

$$C_n = d \sum_{h=-s}^{s} p_h^n \tag{1.1}$$

and we shall assume that $C_n \to \infty$ with n. If i is connected to j we shall let the *strength* v_{ij} of this connection be $+1$ or -1 with probability $\frac{1}{2}$ each. All the choices will be carried out independently of each other.

Once this has been done we have a rectangular matrix denoted by $V = (v_{ik})$ which can be viewed as a linear operator carrying the input vector $x = \{x_i\}$ to the processors. If $d = 1$, then V is a square matrix. The jth processor takes the input from the ith node, $\pm x_i$ or zero, and amplifies it by a factor a_i, which is some positive number not necessarily greater than one, and then transmits it back along the same wire and with the same change of sign if any.

This implies that the returned vector can be written as $y = VAV^T x$ where A is diag $[a_1, a_2, \cdots, a_m]$. The a_j's represent long term memory and they will slowly change due to the varying input patterns. In this paper we shall not study the development of long term memory and we shall put $A = I$, the identity operator. This is the *tabula rasa* hypothesis: the young network does not carry out any nontrivial information processing. The development of the network in time, in particular learning and adaptation, will be discussed in a future paper.

A more realistic model is one in which the backward transmission property is not assumed, that is, when $y = AV^T x$. However, we return to the original network if we consider another aspect of the *tabula rasa* hypothesis: the young network is not biased toward any input. It will be biased if $\|V^T x\| = x^T VV^T x$ is much larger for some vector x than other vectors having the same length as x. In either case we are led to the investigation of the spectral measure of the operator VV^T which is the main goal of this paper.

The controlled randomness of the network is in accordance with current biological theories. It is believed that all neural connections in the brain cannot be genetically encoded, and that some must be left to chance. But, if this is so, how can one account for the similar capabilities among animals in a species? We attempt to reconcile this question by investigating common properties of realizations of our model. By making use of the high dimensionality of the network, we try to formulate and prove limit theorems that can be applied to it and that imply that within the same species individuals have (asymptotically) the same spectral properties for their operator.

At first we were concerned only with nondiverging connections, i.e., $d = 1$. If the *tabula rasa* hypothesis is fully satisfied one would have $VV^T \cong I$. This was, in fact, our first conjecture, that for n large enough, VV^T with high probability was close to I. Since $E((1/n) \times \text{sum of all eigenvalues of } VV^T) = E((1/n) \operatorname{tr} VV^T) = C_n$, we were led to investigating the normalized operator $W = (1/C_n) VV^T$. We did this in a series of mathematical experiments in which we generated several V's and for each one plotted the eigenvalue distribution of W, hoping the eigenvalues would cluster around 1. They clearly did not, as can be seen from the graphs in Fig. 1.3, so that we had to discard the conjecture. But we noticed the consistency between the distributions, which suggested to us that perhaps some other type of limit law was governing the eigenvalues.

It was then hypothesized that perhaps we could get W close to I if the connections diverged, V now being a $n \times dn$ matrix. Subsequent simulations seems to support this (Fig. 1.4). Again it appeared that a limit law was involved. So we finally conjectured that we could with high probability get W as close to I as we please by making first n and then d large enough.

Let us begin our study with some preliminary observations. We could attempt to measure how close $(1/\sqrt{C_n})V^T$ is to being an isometric operator, since an operator $A: R^n \to R^{dn}$ is isometric iff $A^TA = I$. For each $x \in R^n$ we calculate

$$\operatorname{var}\left(\frac{1}{C_n}\|V^Tx\|^2\right) = E\left(\frac{1}{C_n^2}\|V^Tx\|^4\right) - \left[E\left(\frac{1}{C_n}\|V^Tx\|^2\right)\right]^2.$$

We get

$$\begin{aligned} E(\|V^Tx\|^2) &= E\Big(\sum_i (V^Tx)_i^2\Big) = E\Big(\sum_i\Big(\sum_k v_{ki}x_k\Big)^2\Big) \\ &= E\Big(\sum_i \sum_{kk'} v_{ki}v_{k'i}x_kx_{k'}\Big) \qquad (1.2) \\ &= \sum_i \sum_{kk'} E(v_{ki}v_{k'i})x_kx_{k'}. \end{aligned}$$

Since $E(v_{ki}v_{k'i}) = 0$ if $k \neq k'$ and $= P_{ki}$ if $k = k'$ we have

$$E\Big(\frac{1}{C_n}\|V^Tx\|^2\Big) = \frac{1}{C_n}\sum_k x_k^2 \sum_i P_{ki} = \|x\|^2 \qquad (1.3)$$

since for each k, $\sum_i P_{ki} = C_n$ (see Fig. 1.2). We also have

$$\begin{aligned} E(\|V^Tx\|^4) &= E\Big(\Big(\sum_i\Big(\sum_k v_{ki}x_k\Big)^2\Big)^2\Big) \\ &= E\Big(\sum_{\substack{i_1i_2\\k_1k_2k_3k_4}} v_{k_1i_1}v_{k_2i_1}x_{k_1}x_{k_2}v_{k_3i_2}v_{k_4i_2}x_{k_3}x_{k_4}\Big) \qquad (1.4) \\ &= \sum_{\substack{i_1\neq i_2\\k_1k_2}} P_{k_1i_1}P_{k_2i_2}x_{k_1}^2x_{k_2}^2 + 3\sum_{\substack{i\\k_1\neq k_2}} P_{k_1i}P_{k_2i}x_{k_1}^2x_{k_2}^2 + \sum_{ik} P_{ki}x_k^4 \\ &= \sum_{\substack{i_1\\k_1k_2}} P_{k_1i_1}x_{k_1}^2x_{k_2}^2(C_n - P_{k_2i_1}) + 3\sum_{k_1\neq k_2} a_{k_1k_2}x_{k_1}^2x_{k_2}^2 + C_n\sum_k x_k^4 \end{aligned}$$

(where $a_{k_1k_2} \equiv \sum_i P_{k_1i}P_{k_2i}$)

$$\begin{aligned} &= C_n^2 \sum_{k_1k_2} x_{k_1}^2x_{k_2}^2 - \sum_{k_1k_2} a_{k_1k_2}x_{k_1}^2x_{k_2}^2 \\ &\quad + 3\sum_{k_1} x_{k_1}^2\Big(\sum_{k_2} a_{k_1k_2}x_{k_2}^2 - ax_{k_1}^2\Big) + C_n\sum_k x_k^4 \end{aligned}$$

(where $a \equiv a_{00} = d \sum (p_h^n)^2$)

$$= C_n^2 \sum_{k_1k_2} x_{k_1}^2 x_{k_2}^2 + 2 \sum_{k_1k_2} a_{k_1k_2} x_{k_1}^2 x_{k_2}^2 + (C_n - 3a) \sum_k x_k^4.$$

Therefore

$$\text{var}\left(\frac{1}{C_n}\|V^T x\|^2\right) = \frac{1}{C_n^2}\left[2 \sum_{k_1k_2} a_{k_1k_2} x_{k_1}^2 x_{k_2}^2 + (C_n - 3a) \sum_k x_k^4\right]. \tag{1.5}$$

Assume $C_n - 3a \geqq 0$ for n sufficiently large. For these n the maximum value of (1.5) over unit vectors is attained for $x = (1, 0, \cdots, 0)$.

Therefore

$$\begin{aligned} \max_{\|x\|=1} \text{var}\left(\frac{1}{C_n}\|V^T x\|^2\right) &= \frac{1}{C_n^2}[2a + C_n - 3a] = \frac{C_n - a}{C_n^2} \\ &= \frac{1}{d}\left(\frac{\sum p_h^n - \sum (p_h^n)^2}{(\sum p_h^n)^2}\right). \end{aligned} \tag{1.6}$$

Since (1.6) $\to 0$ as $n \to \infty$ we can conclude that for n sufficiently large, and for $x \in R^n$ we can with high probability get $\|(1/\sqrt{C_n})\, V^T x\|^2$ as close as we want to $\|x\|^2$. But this does not imply $(1/\sqrt{C_n})\, V^T$ can be made nearly isometric. We should

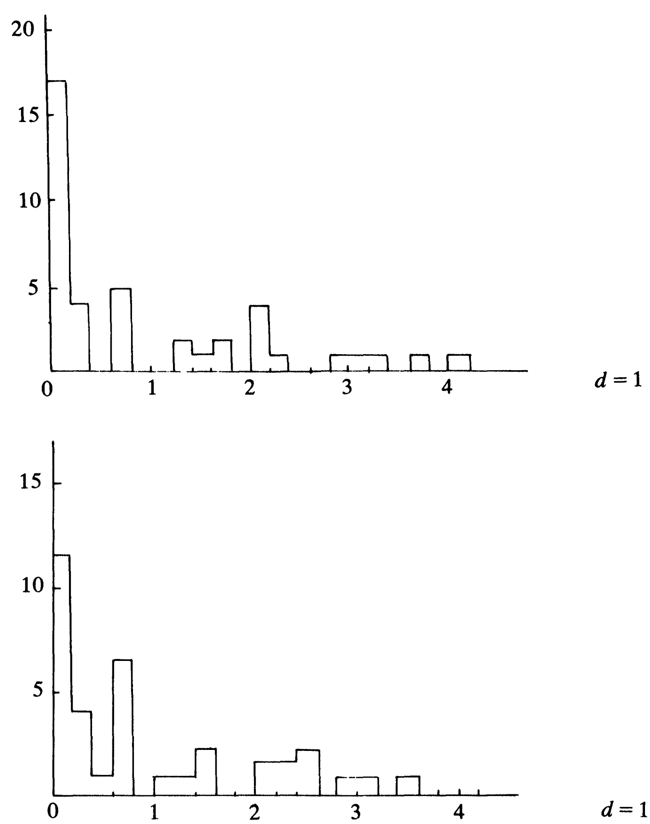

FIG. 1.3

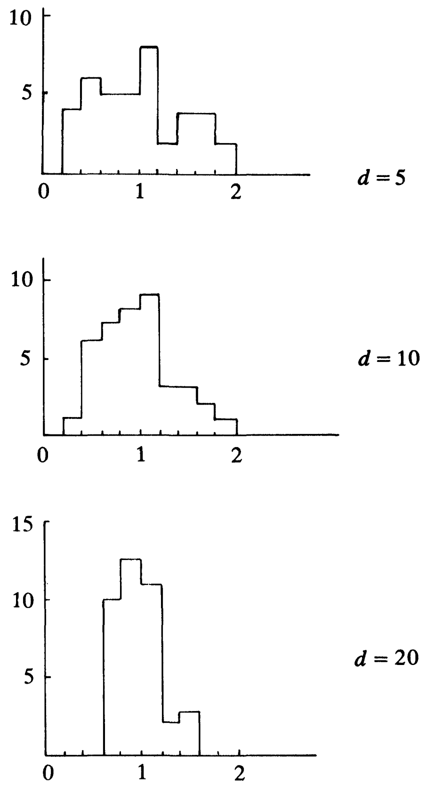

FIG. 1.4. *Note: for all graphs,* $n=40, s=2, p_h^n=\frac{1}{2}-\frac{1}{5}\times|h|, -2\leqq h\leqq 2$

consider n orthonormal vectors, say, $x_1, \cdots, x_n$, and see whether $(1/C_n)\|V^T x_i\|^2 \cong 1$ for each of them. For any $\varepsilon > 0$ we get

$$\begin{aligned}
&\operatorname{prob}\left(\left|\frac{1}{C_n}\left\|V^T x_i\right\|^2-1<\varepsilon \ \forall_i\right)\right. \\
(1.7)\qquad &\geqq 1-\sum_i \operatorname{prob}\left(\left|\frac{1}{C_n}\left\|V^T x_i\right\|^2-1\right| \geqq \varepsilon\right) \\
&\geqq 1-\frac{n}{\varepsilon^2 d}\left(\frac{\sum p_h^n-\sum (p_h^n)^2}{(\sum p_h^n)^2}\right),
\end{aligned}$$

using Chebyshev's inequality. Since $\sum p_h < bn$ where $b > 0$ we have for n sufficiently large (assuming again $C_n - 3a \geqq 0$)

$$(1.8)\qquad n\left(\frac{\sum p_h^n-\sum (p_h^n)^2}{(\sum p_h^n)^2}\right) \geqq n\left(\frac{\sum p_h^n-\frac{1}{3}\sum p_h^n}{(\sum p_h^n)^2}\right)=\frac{2n}{3\sum p_h^n}>\frac{2}{3b}>0.$$

Under the above assumption we cannot get (1.7) as close to one as desired unless d is large enough too.

Clearly this argument is not sufficient and we need a deeper analysis of the problem. To start this, let us note that our present object of study is the spectral properties of the W operator for large values of n. It is a symmetric and nonnegative definite random operator. Denote its eigenvalues by $\lambda_i^{(n)}, i = 1, 2, \cdots, n$ and the spectral measure characterized in terms of the spectral d.f.

$$F_n(\lambda) = \frac{1}{n} \#(\lambda_i^{(n)} \leqq \lambda) \tag{1.9}$$

on the nonnegative real line. Of course $F_n(\lambda)$ is a stochastic process. The main problem here is whether F_n converges and whether this limit is universal, i.e., does not depend upon the controlling probabilities.

Before proceeding to the analysis let us mention that in our work on long term memory we are mainly interested in many-layered models, where there are many sets of processors connected in a cascade fashion. The present one-layer model will however bring out the relevant mathematical facts clearly enough.

2. Convergence of spectrum. With $C = C_n$, $p_h = p_h^n$, $P_{ij} = P_{ij}^n$ we prove the following

THEOREM. *Assume $C \to \infty$ as $n \to \infty$. Given d, let*

$$b_1 = \frac{d+1-2\sqrt{d}}{d}, \quad b_2 = \frac{d+1+2\sqrt{d}}{d} \quad \textit{and} \quad g(u) = \frac{d\sqrt{(u-b_1)(b_2-u)}}{2\pi u},$$

$$b_1 \leqq u \leqq b_2.$$

Then, for each λ, $\{F_n(\lambda)\}$ converges in probability to:

$$F(\lambda) = \begin{cases} 0 & \textit{for} \quad u \leqq b_1, \\ \displaystyle\int_{b_1}^{\lambda} g(u)\, du & \textit{for} \quad b_1 \leqq \lambda \leqq b_2, \\ 1 & \textit{for} \quad u \geqq b_2. \end{cases} \tag{2.1}$$

The proof is not easy but we see no hope of simplifying it at present. To do this some radically different approach would be needed.

Our strategy of proof is based on 8 lemmas. We first derive the limiting values of the expected moments of the eigenvalues, i.e.

$$\frac{1}{n} \sum_{i=1}^{n} \lambda_i^r = \frac{1}{n} (\operatorname{tr} W^r).$$

Using a combinatorial argument we then show that these values satisfy a difference equation. Next we prove the convergence of the moments in probability to these values. We then solve the difference equation with the aid of the generating function which we invert in the complex domain, arriving at $F(\lambda)$. We finish by showing the convergence in probability of $F_n(\lambda)$ to $F(\lambda)$ using a standard argument from measure theory.

Proof of theorem: We will begin with the case of no divergence, i.e., $d = 1$. Our first task is to find, for each integer $r \geqq 1$, the limiting value of:

$$
\begin{aligned}
E\left(\frac{1}{n}\operatorname{tr} W^r\right) &= \frac{1}{nC^r} E\left(\sum_{\substack{i_1 \cdots i_r \\ k_1 \cdots k_r}} v_{i_1k_1} v_{i_2k_1} v_{i_2k_2} v_{i_3k_2} \cdots v_{i_rk_r} v_{i_1k_r} \right) \\
&= \frac{1}{nC^r} \sum_{\substack{i_1 \cdots i_r \\ k_1 \cdots k_r}} E(v_{i_1k_1} v_{i_2k_1} \cdots v_{i_rk_r} v_{i_1k_r}).
\end{aligned} \tag{2.2}
$$

Many of these terms are zero. As in (1.3) the only nonzero terms are those in which each v_{ik} appearing in the term is repeated an even number of times. Consider one way of grouping the v_{ik}'s in this manner. For example, one way is to take all terms $(i_1k_1) = (i_2k_1)$, $(i_2, k_2) = (i_3, k_2) = \cdots = (i_1, k_r)$ and $(i_1, k_1) \neq (i_2, k_2)$. In this example there are two groups. Taking all terms where the v_{ik}'s are equal is another way of grouping. We can perform the sum in (2.2) by summing on each type of grouping. Given one of these we have

$$
\sum_{\substack{a_1 \cdots a_{r'} \\ b_1 \cdots b_{r'}}} P_{a_1b_1} P_{a_2b_2} \cdots P_{a_{r'}b_{r'}} \tag{2.3}
$$

where $r' \leqq r$ and constraints on $a_1 \cdots a_{r'}$, $b_1 \cdots b_{r'}$ due to the original restrictions. Notice that we do not sum on all values of $a_1 \cdots a_{r'}$, $b_1 \cdots b_{r'}$ since, for example, (a_1, b_1) cannot equal (a_2, b_2). We illustrate for the case when $r = 3$. Then (2.2) becomes

$$
E\left(\frac{1}{n} W^3\right) = \frac{1}{nC^3} \sum_{\substack{i_1 \cdots i_r \\ k_1 \cdots k_r}} E(v_{i_1k_1} v_{i_2k_1} v_{i_2k_2} v_{i_3k_3} v_{i_3k_3} v_{i_1k_3}). \tag{2.4}
$$

Notice that some groupings cannot be done. For example, it is not possible to group $v_{i_1k_1}$ with $v_{i_2k_2}$ and have $v_{i_2k_1}$ in another group since (i_1, k_1) will equal (i_2, k_2) which will make $v_{i_2k_1} = v_{i_1k_1} = v_{i_2k_2}$. One legitimate way of grouping is to take $(i_1, k_1) = (i_2, k_1) = (i_2, k_2) = (i_3, k_2)$, $(i_3, k_3) = (i_1, k_3)$ *and* $(i_3, k_3) \neq (i_1, k_1)$. Letting (a_1, b_1) denote indices for the first group and (a_2, b_2) for the second group, we find that (2.3) becomes

$$
\sum_{\substack{a_1a_2 \\ b_1b_2}} P_{a_1b_1} P_{a_1b_2}. \tag{2.5}
$$

Here, $r' = 2$ and we must constrain a_1 with a_2. It is clear we do not take terms where $(a_1b_1) = (a_2, b_2)$. Notice that b_1 and b_2 are free in the sense that they are not constrained to each other.

Given r, let S_r be the set of all groupings, where each $s \in S_r$ will contribute a nonnegligible amount to $E((1/n) \operatorname{tr} W^r)$ in the limit.

LEMMA 1. *For each $s \in S_r$, $r' = r$, i.e., $s \in S_r$ pairs up the v_{ik}'s exactly.*

Proof. We can bound (2.3) by first removing all pairs a_jb_j where either a_j or b_j is not constrained (note that for each P_{ab} either a or b must be repeated in another factor). Each time we remove any single ones we can bound (2.3) by removing the $P_{a_jb_j}$, along with the indices a_jk_j, and multiplying the resulting sum by C. After this is done, choose any b_j and sum on it. If there are t factors $P_{a_{j'}b_{j'}}$ such that b_j is constrained with $b_{j'}$, we can bound the sum by removing these t factors and their indices, and multiplying the result by $\sum_h p_h^t$ (since $\sum e_{i_1}\cdots e_{i_t} \leqq \prod_{j=1}^t (\sum_{i_j} e_{i_j}^t)^{1/t}$). Note that for each $P_{a_jb_j}$ that is removed there is still a $P_{a_{j'}b_{j'}}$ where a_j and $a_{j'}$ are constrained to each other, that is not removed, since all free indices are eliminated before. We then remove all single $P_{a_jn_j}$'s, then factors, and so on. In the end we will have

$$(2.6)\qquad (2.3) \leqq (nC^{r''})\left(\prod_{j=1}^{q}\left(\sum_h p_h^{t_j}\right)\right) \leqq nC^{r''}\left(\sum p_h^2\right)^q$$

where $\sum t_j + r'' = r'$, $t_j \geqq 2$. If only single $P_{a_jb_j}$'s are removed at each step, then we obviously will not have the second factor. In this case define $q = 0$. Since q is obviously $< \sum t_j$ we see that $r' < r$ implies $(1/(nC^r)) \times (2.3) \to 0$ as $n \to \infty$.

Note in the proof at any time we clumped we could have done so on an a_j instead of a b_j. It is also evident that for $s \in S_r$, $q = 0$. So we conclude that S_r is the set of all constraints that exactly pair up the v_{ik}'s and at each step (as done in Lemma 1) a single $P_{a_jb_j}$ can be removed.

LEMMA 2. *For each* $s \in S_r$ (2.3) $\sim nC^r$ *so that* $E((1/n)\operatorname{tr} W^r) \to f(r) \equiv$ *number of elements in* S_r, *as* $n \to \infty$.

Proof. Given $s \in S_r$, either some a_j or some b_j is free. Suppose b_j is free (and, of course, a_j is not). Then

$$(2.7)\qquad (2.3) = \sum_{\substack{a_1\cdots a_{j-1},a_{j+1}\cdots a_r\\ b_1\cdots b_{j-1},b_{j+1}\cdots b_r}} P_{a_1b_1}\cdots P_{a_{j-1}b_{j-1}}P_{a_{j+1}b_{j+1}}\cdots P_{a_rb_r}\left(C - \sum_{j'} P_{a_{j'}b_{j'}}\right)$$

where for each term j' ranges on indices where $(a_j, b_j) = (a_{j'}, b_{j'})$. But since from (2.6)

$$\sum_{\substack{a_1\cdots a_{j-1}a_{j+1}\cdots a_r\\ b_1\cdots b_{j-1}b_{j+1}\cdots b_r}} P_{a_1b_1}\cdots P_{a_{j-1}b_{j-1}}P_{a_{j+1}b_{j+1}}\cdots P_{a_rb_r} = O(nC^{r-1})$$

and since $\sum_j P_{a_{j'}b_{j'}} \leqq r$ we have

$$(2.8)\qquad \frac{1}{nC^r}\times(2.3) \sim \frac{1}{nC^r}\sum_{\substack{a_1\cdots a_{j-1}a_{j+1}\cdots a_r\\ b_1\cdots b_{j-1}b_{j+1}\cdots b_r}} P_{a_1b_1}\cdots P_{a_{j-1}b_{j-1}}P_{a_{j+1}b_{j+1}}\cdots P_{a_rb_r}C.$$

It is clear we can do this at every step. Therefore (2.3) $\sim nC^r$.

We introduce an essential tool that will enable us to get a recursion relation for $\{f(r)\}_{r=1}^{\infty}$. We call it clock notation. The basic notation is in Fig. 2.1. Each number corresponds to a v_{ik} and the indices show the original constraints. We can characterize an $s \in S_r$ completely using clock notation. For each pair of v_{ik}'s draw a line between the corresponding pair of numbers. Each number will then have one and only one line from it. Each line will correspond to a $P_{a_jb_j}$ and vice versa. It is easy to see that the only way a single $P_{a_jb_j}$ can be removed is when the line representing it connects two adjacent numbers on the clock. Whenever a pair is removed the two places on the clock nearest the pair can be considered adjacent and we then have a clock corresponding to an $s \in S_{r-1}$. For example, if 1 and 2 are paired, then they can be removed and $2r$ and 3 are then

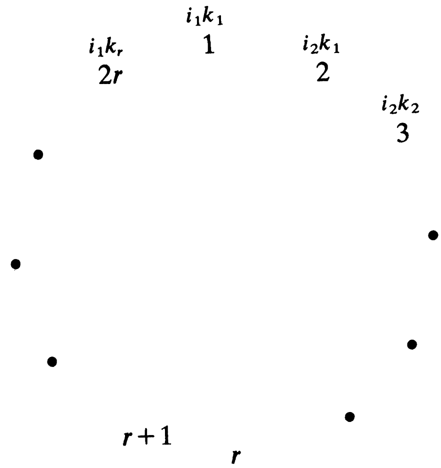

FIG. 2.1

adjacent, since i_1 and i_2 must be constrained. We can continue in this way until all pairs are removed.

However, an $s \in S_r$ cannot have any lines (in its clock representation) crossing, since a pair can only be removed if all numbers between them (on either side of the clock) are removed. This is impossible if two lines cross. We have therefore

LEMMA 3. *$S_r \subseteq B_r$, where B_r is the set of all constraints (given r), with clock representations having exact pairings and no crossing.*

It is easy to see that for any $b \in B_r$ such that a $P_{a_jb_j}$ can be removed the resulting constraint will be in B_{r-1}.

LEMMA 4. *$S_r = B_r$.*

Proof. It remains to show that for any $r' \leqq r$, $b \in B_{r'} \Rightarrow$ some $P_{a_jb_j}$ can be removed. Then $b \in B_r$ must be in S_r.

CLAIM. *For $r \geqq 2$, $b \in B_r$ has at least* two *$P_{a_jb_j}$'s that can be removed.*

We prove this by using induction on r and clock notation. $r = 2$ is obvious. Assume its true for all $r' \leqq r$. Given $b \in B_{r+1}$ draw it in clock notation. Suppose 1 is paired with m.

Case 1. The line will split the clock up into 2 parts, one representing a $b \in B_{r_1}$, and the other representing one in B_{r_2}, $r_1, r_2 \leqq r$. If both r_1 and r_2 are $\geqq 2$, then the inductive hypothesis will imply that each part will have at least two adjacent pairs, at least one pair on each side must be on the original clock. If either r_1 or r_2 or both are 1, then it is clear we will have two pairs on the original clock.

Case 2. $m = r$ or 2. Then $1 - m$ is one pair and the remaining clock which represents a constraint in B_r has two pairs, at least one being on the original clock.

LEMMA 5. *With $f(0) \equiv 1$ we have*

$$f(r) = \sum_{j=1}^{r} f(r-j)f(j-1), \qquad r \geqq 1. \tag{2.9}$$

Proof. We just count how many clock configurations there are. 1 can be connected to 2, 4, 6, $\cdots$, $2r$. When 1–2 we get $f(0)f(r-1)$. When 1–4 we get $f(1)f(r-2)$, etc. Therefore $f(r)=\sum_{j=1}^r f(r-j)f(j-1)$.

We next look at

$$E\left(\left(\frac{1}{n}\operatorname{tr} W^r\right)^2\right)$$

$$(2.10)\qquad =\frac{1}{n^2C^2}\sum_{\substack{i_1\cdots i_k,k_1\cdots k_r\\ i_1'\cdots i_r',k_1'\cdots k_r'}} E(v_{i_1k_1}v_{i_2k_1}\cdots v_{i_rk_r}v_{i_1k_r}v_{i_1'k_1'}v_{i_2'k_1'}\cdots v_{i_1'k_r'}).$$

Again, we see that the only nonvanishing terms are those whose v_{ik}'s pair up. Given one type of pairing we can write

$$(2.11)\qquad \sum_{\substack{a_1\cdots a_{r'}\\ k_1\cdots k_{r'}}} P_{a_1b_1}\cdots P_{a_{r'}b_{r'}}$$

with constraints on $a_1\cdots a_{r'}$, $b_1\cdots b_{r'}$ due to the original restrictions. As in Lemma 1 it is not difficult to see that this pairing can contribute a nonnegligible amount to

$$E\left(\left(\frac{1}{n}\operatorname{tr} W^r\right)^2\right)$$

in the limit only if $r'=2r$. Also at any step we cannot remove any $P_{a_jb_j}$ that is a pairing of a v_{ik} and a $v_{i'k'}$ since neither a_j nor b_j will be free. We can only remove a $P_{a_jb_j}$ that pairs adjacent v_{ik}'s or adjacent $v_{i'k'}$'s. By induction we conclude that

$$(2.12)\qquad \begin{aligned} E\left(\left(\frac{1}{n}\operatorname{tr} W^r\right)^2\right) &\sim \frac{1}{nC^{2r}}\sum_{\substack{i_1\cdots i_r,k_1\cdots k_r\\ i_1'\cdots i_r',k_1'\cdots k_r'}} E(v_{i_1k_1}\cdots v_{i_rk_r}v_{i_1k_r})\\ &\qquad\qquad E(v_{i_1'k_1'}\cdots v_{i_r'k_r'}v_{i_1'k_r'})\\ &=\left[E\left(\frac{1}{n}\operatorname{tr} W^r\right)\right]^2. \end{aligned}$$

Therefore var $((1/n)\operatorname{tr} W^r)\to 0$.

LEMMA 6. *$(1/n)\operatorname{tr} W^r$ converges in probability to $f(r)$.*

Proof. Given $\varepsilon>0$ we have

$$(2.13)\qquad \begin{aligned} P\left(\left|\frac{1}{n}\operatorname{tr} W^r-f(r)\right|>\varepsilon\right) &\leqq P\left(\left|\frac{1}{n}\operatorname{tr} W^r-E\left(\frac{1}{n}\operatorname{tr} W^r\right)\right|>\varepsilon\right)\\ &\quad+P\left(\left|E\left(\frac{1}{n}\operatorname{tr} W^r\right)-f(r)\right|>\varepsilon\right). \end{aligned}$$

Using Chebyshev's inequality we get

$$P\left(\left|\frac{1}{n}\operatorname{tr} W^r - f(r)\right| < \varepsilon\right) \leqq \frac{\operatorname{var}((1/n)\operatorname{tr} W^r)}{\varepsilon^2} + P\left(\left|E\left(\frac{1}{n}\operatorname{tr} W^r\right) - f(r)\right| > \varepsilon\right). \tag{2.14}$$

Since both terms go to 0 as $n \to \infty$, we are done.

We generalize now to the case where the connections diverge by a factor of $d \geqq 1$. The matrix $V = (v_{ij})$ will be $n \times dn$.

As before we connect in a circular fashion. Recall that C is now $d \sum_h p_h$. Given a row, when the probabilities are summed across the columns we get C. Given a column, the sum of the probabilities across the rows is C/d (see Fig. 1.2). We follow along the same way as in the case $d = 1$. The relevant constraints on the indices when evaluating $E((1/n)\operatorname{tr} W^r)$ are the same, except the value on each constraint will not be 1. Whenever a $P_{a_j b_j}$ is removed we get C if the removal is done by summing on b_j, and C/d if it is done on a_j, except when we are down to the end, since $\sum_{ab} P_{ab} = nC$. Let $f(r) = \lim_{n\to\infty} E((1/n)\operatorname{tr} W^r)$.

We will use a slightly different notation we call zig-zag notation. The natural constraints on the i's and k's can clearly be seen in Fig. 2.2.

In this example $r = 5$. $v_{i_1k_1}$ is represented by the pair of circles labeled i_1 and k_1, $v_{i_2k_1}$ by i_2 and k_1, etc. Any constraint can be given by filling in the circles with symbols where a symbol appearing twice means a constraint on the two indices. For example the constraint given in Fig. 2.3 can be represented by:

$$\begin{array}{ccccccccccc} a & & b & & a & & c & & d & & a. \\ & x & & x & & y & & y & & y & \end{array}$$

Removing a $P_{a_j b_j}$ and reducing to an $s \in S_{r-1}$ corresponds to removing a symbol on one level whose two symbols next to it on the other level are the same, and then closing up the gap. For example, after removing 2–3 in Fig. 2.3, we get

$$\begin{array}{ccccccccc} a & & a & & c & & d & & a. \\ & x & & y & & y & & y & \end{array}$$

Note that $\sum_{ab} P_{ab}$ will correspond to either

$$\begin{array}{ccc} a & & a \\ & x & \end{array} \quad \text{or} \quad \begin{array}{ccc} & a & \\ x & & x. \end{array}$$

So, for $s \in S_r$ we find out how many $1/d$'s we have by counting how many times we have reduced the zig-zag figure by removing a symbol on the top and closing up

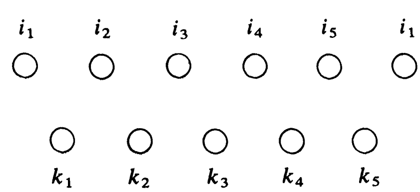

FIG. 2.2

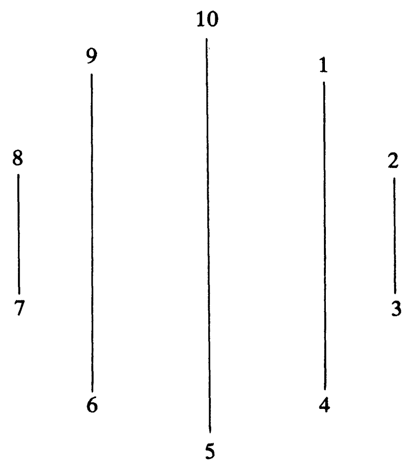

FIG. 2.3

the gap on the bottom, until we reach

$$a\,{}_{x}\,a \quad \text{or} \quad {}_{x}\,a\,{}_{x}.$$

If $N(s)$ is the number of times this is done for s, then the value we get for the sum is $(1/d)^{N(s)}$, and $f(r)=\sum_{s\in S_r}(1/d)^{N(s)}$.

Let $h(r)$ be the value we get if we get a $1/d$ every time we sum on a b_j instead of an a_j. We could have just as well put the k's on the top and the i's on the bottom in zig-zag notation. For this reason $f(r)=h(r)$. Note that $f(1)=1$. With $f(0)\equiv 1$ we have

LEMMA 7. $f(r)=((d-1)/d)f(r-1)+(1/d)\sum_{j=1}^{r} f(j-1)f(r-j)$, $d=1,2,3,\cdots, r\geqq 1$.

Note. For both $d=1$ and $d>1$ we have a quadratic difference equation. The crucial fact is that it is of convolution type so that we can use a generating function technique to solve it.

Proof. It is easy to see that the formula holds for $r=1$. As in Lemma 4 we consider what values we get when, in clock notation, 1 is connected to $2,4,6,\cdots,2r$. When 1–2, in zig-zag notation we will get as shown in Fig. 2.4. For each $s\in S_{r-1}$ we will get $N(s)$ reductions from the top until we are at

$$a\,{}_{x}\,a\,{}_{y}\,a.$$

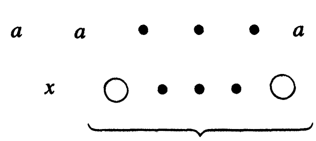

FIG. 2.4

Since there are no more reductions from the top we sum up and get $f(r-1)$. When $1\text{–}2j$, $r \geqq j \geqq 2$, we get Fig. 2.5 and so we must take into account the combining of every pair s_1 and s_2 where $s_1 \in S_{j-1}$ and $s_2 \in S_{r-j}$ ($s \in S_0 \Rightarrow N(s) = 1$). If $M(s)$ is defined in the same way as $N(s)$ except considering k's instead of i's, we will get $M(s_1) \times N(s_2)$ removals before the figure is reduced to

$$a\;{}_x\;x\;{}_x\;a\;{}_y\;a,$$

and we will get one more reduction from the top. Summing, we get $(1/d)h(j-1)f(r-j) = (1/d)f(j-1)f(r-j)$. Therefore,

$$f(r) = f(r-1) + \frac{1}{d}\sum_{j=2}^{r} f(j-1)f(r-j) = \frac{d-1}{d} f(r-1) + \frac{1}{d}\sum_{j=1}^{r} f(j-1)f(r-j).$$

LEMMA 8. *The generating function* $G: \mathbb{C} \to \mathbb{C}$ *defined by* $G(z) = \sum_{j=0}^{\infty} f(r)z^r$ *is well defined and analytic in a neighborhood of* 0.

Proof. The auxiliary function

$$b(r) \equiv \left[\frac{d+1}{d}\frac{r^2}{(r-1)^2} + \frac{2}{d}\frac{r^2}{(r-1)^2} + \frac{r^2}{d(r-1)^3}\log r(r-2)\right]$$

is bounded above for all $r \geqq 3$ so we can find $a \geqq 1$ such that $f(2) \leqq a^2/4$ and $b(r) \leqq a$. We show by induction that $f(r) \leqq a^r/r^2$ for $r \geqq 1$. Since $f(1) = 1$ the cases $r = 1, 2$ are already satisfied. Assume the statement is true for all $r' < r$, $r \geqq 3$. Then,

$$\begin{aligned}
f(r) &= \frac{d-1}{d} f(r-1) + \frac{1}{d}\sum_{j=1}^{r} f(j-1)f(r-j) \\
&= \frac{d+1}{d} f(r-1) + \frac{2}{d} f(r-2)f(1) + \frac{1}{d}\sum_{j=3}^{r-2} f(j-1)(r-j) \\
&\leqq \frac{d+1}{d}\frac{a^{r-1}}{(r-1)^2} + \frac{2}{d}\frac{a^{r-1}}{(r-2)^2} + \frac{a^{r-1}}{d}\left(\sum_{j=3}^{r-2}\frac{1}{(j-1)^2(r-j)^2}\right) \\
&\leqq \frac{a^{r-1}}{r^2}\left(\frac{d+1r^2}{d(r-1)^2} + \frac{2r^2}{d(r-2)^2} + \frac{r^2}{d}\int_2^{r-1}\frac{dx}{(x-1)^2(r-x)^2}\right) \\
&= \frac{a^{r-1}}{r^2}\left(\frac{d+1}{d}\frac{r^2}{(r-1)^2} + \frac{2}{d}\frac{r^2}{d(r-2)^2} + \frac{r^2}{d(r-1)^3}\log r(r-2)\right) \\
&\leqq \frac{a^r}{r^2}.
\end{aligned} \tag{2.15}$$

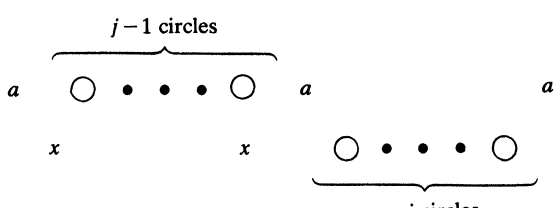

FIG. 2.5

We can easily find $G(x)$, $x \in R$ in a neighborhood of the origin. Let $\mathcal{F}$ be the convolution function of $\{f(r)\}_{r=0}^{\infty}$ with itself, and $\mathcal{G}$ the generating function of $\{\mathcal{F}(r)\}_{r=0}^{\infty}$. Then,

$$f(r) = \frac{d-1}{d} f(r-1) + \frac{1}{d} \mathcal{F}(r-1) \tag{2.16}$$

and

$$G(x) - 1 = \frac{x(d-1)}{d} G(x) + \frac{x}{d} \mathcal{G}(x) = \frac{x(d-1)}{d} G(x) + \frac{x}{d} G^2(x), \tag{2.17}$$

$$xG^2(x) + (x(d-1)-d)G(x) + d = 0 \tag{2.18}$$

and

$$G(x) = \frac{d - x(d-1) \pm \sqrt{(x(d-1)-d)^2 - 4\,dx}}{2x}. \tag{2.19}$$

Since $G(0) = 1$ we conclude that for $z \in \mathbb{C}$ in a neighborhood of the origin we have

$$G(z) = \frac{d - z(d-1) - \sqrt{(z(d-1)-d)^2 - 4\,dz}}{2z}. \tag{2.20}$$

Note that because of the square root function in (2.19) we must be careful how we extend G analytically into $\mathbb{C}$, that is, we must continue G on the right branch of the Riemann surface.

We determine the function

$$\phi(t) = \sum_{r=0}^{\infty} \frac{f(r)(it)^r}{r!} \tag{2.21}$$

by invoking the theorem (Titchmarsh [1]) that says

$$a(z) = \sum_{r=0}^{\infty} a_n z^n, \qquad b(z) = \sum_{r=0}^{\infty} b_n z^n,$$

both analytic in a neighborhood of 0 imply $c(y) = \sum_{r=0}^{\infty} a_n b_n y^n$ is analytic in a neighborhood of 0 and can be given by

$$\frac{1}{2\pi i} \oint a(z) b\left(\frac{y}{z}\right) \frac{dz}{z} \tag{2.22}$$

where the contour is around the origin where both $a(z)$ and $b(y/z)$ are analytic. With $a(z) = G(z)$, $b(z) = e^z$ we have

$$\phi(t) = \frac{1}{2\pi i} \oint G(z) e^{it/z} \frac{dz}{z}. \tag{2.23}$$

For $d \geqq 2$,

$$
\begin{aligned}
(z(d-1)-d)^2-4\,dz &= (d-1)^2\left[z^2-\frac{2dz}{(d-1)}+\frac{d^2}{(d-1)^2}-\frac{4dz}{(d-1)^2}\right] \\
&= (d-1)^2\left[z^2-\frac{2d(d+1)z}{(d-1)^2}+\frac{d^2}{(d-1)^2}\right] \\
&= (d-1)^2[(z-a_1)(z-a_2)]
\end{aligned}
\tag{2.24}
$$

where a_1 and a_2 (with $a_1 < a_2$) are given by

$$
\begin{aligned}
&\left(\frac{2d(d+1)}{(d-1)^2}\pm\sqrt{\frac{4d^2(d+1)^2}{(d-1)^4}-\frac{4d^2}{(d-1)^2}}\right)\div 2 \\
&\quad =\frac{d(d+1)}{(d-1)^2}\pm\frac{d}{(d-1)}\sqrt{\frac{(d+1)^2}{(d-1)^2}-1}=\frac{d}{(d-1)^2}[(d+1)\pm 2\sqrt{d}].
\end{aligned}
\tag{2.25}
$$

It is not difficult to see that $0 < a_1 < d/(d-1) < a_2$. Therefore, for $d \geqq 2$

$$
G(z)=\frac{(d-1)\left(-\left(z-\dfrac{d}{d-1}\right)-\sqrt{(z-a_1)(z-a_2)}\right)}{2z}.
\tag{2.26}
$$

For $d=1$ we get

$$
G(z)=\frac{1-\sqrt{1-4z}}{2z}.
\tag{2.27}
$$

We will work first with $d \geqq 2$. $G(z)$ can be rewritten as

$$
\begin{aligned}
&\frac{(d-1)(-r_3e^{i\theta_3}-(\sqrt{r_1r_2}\,e^{[i(\theta_1+\theta_2)]/2+i\pi}))}{2R\,e^{i\theta}} \\
&\quad =\frac{(d-1)}{2}\left(\frac{-r_3}{R}e^{i(\theta_3-\theta)}+\frac{\sqrt{r_1r_2}}{R}e^{[i((\theta_1-\theta)+(\theta_2-\theta))]/2}\right)
\end{aligned}
\tag{2.28}
$$

where $R, r_1, r_2, r_3, \theta, \theta_1, \theta_2, \theta_3$ can be seen in Fig. 2.6. The reason we have π in the exponential is because the square root function must be positive on the real line to the left of a_1, i.e., when $\theta=\theta_1=\theta_2=\theta_3=\pi$. Elementary trigonometry tells us that as $R\to\infty$, $r_1/R\to 1$, and $\theta_i\to\theta$, $i=1,2,3$. From (2.28) we see, as $R\to\infty$, $G(z)\to 0$ for each θ.

We integrate along the path in Fig. 2.7.

Noting the discontinuity of the square root function across the line between a_1 and a_2 we get in the limit

$$
\begin{aligned}
&\phi(t)-\frac{1}{2\pi i}\int_0^{2\pi} G(R\,e^{i\theta})\,e^{(it/R)e^{-i\theta}}\,d\theta \\
&\quad -\frac{1}{2\pi}\int_{a_1}^{a_2}(d-1)\frac{\sqrt{-(x-a_1)(x-a_2)}\,e^{it/x}}{x^2}\,dx=0.
\end{aligned}
\tag{2.29}
$$

Since $G(R\,e^{i})\,e^{(it/R)e^{-i\theta}}$ is uniformly bounded for all R we use the bounded convergence theorem to conclude that the second term in (2.29) converges to zero.

Therefore

$$\phi(t)=\frac{(d-1)}{2\pi}\int_{a_1}^{a_2}\frac{\sqrt{-(x-a_1)(x-a_2)}\,e^{it/x}}{x^2}\,dx. \tag{2.30}$$

Letting $u=1/x$ we get

$$\begin{aligned}\phi(t)&=\frac{(d-1)}{2\pi}\int_{1/a_2}^{1/a_1}\sqrt{-(1/u-a_1)(1/u-a_2)}\,e^{itu}\,du\\ &=\frac{(d-1)}{2\pi}\int_{1/a_2}^{1/a_1}\frac{\sqrt{-a_1a_2(u-1/a_1)(u-1/a_2)}}{u}\,e^{itu}\,du\\ &=\frac{d}{2\pi}\int_{b_1}^{b_2}\frac{\sqrt{(u-b_1)(b_2-u)}}{u}\,e^{itu}\,du\end{aligned} \tag{2.31}$$

where $b_1=(d+1-2\sqrt{d})/d$, $b_2=(d+1+2\sqrt{d})/d$.

For $d=1$ we proceed similarly, noting that the square root function must be $-2i\sqrt{r_1}\,e^{i\theta_1/2}$ (see Fig. 2.8)

Since $G(R\,e^{i\theta})\to 0$ uniformly in θ as $R\to\infty$ we get

$$\begin{aligned}\phi(t)&=\frac{1}{\pi}\int_{1/4}^{\infty}\frac{\sqrt{x-1/4}}{x^2}\,e^{it/x}\,dx=\frac{1}{\pi}\int_0^4\sqrt{1/u-1/4}\,e^{itu}\,du\\ &=\frac{1}{2\pi}\int_0^4\frac{\sqrt{4-u}}{\sqrt{u}}\,e^{itu}\,du.\end{aligned} \tag{2.32}$$

Combining the two results we get

$$\begin{gathered}\phi(t)=\frac{d}{2\pi}\int_{b_1}^{b_2}\frac{\sqrt{(u-b_1)(b_2-u)}}{u}\,e^{itu}\,du\\ b_1=\frac{d+1-2\sqrt{d}}{d},\qquad b_2=\frac{d+1+2\sqrt{d}}{d},\qquad d=1,2,3,\cdots.\end{gathered} \tag{2.33}$$

It is not difficult to prove that

$$\frac{d}{2\pi}\int_{b_1}^{b_2}\frac{\sqrt{(u-b_1)(b_2-u)}}{u}\,du=1,\qquad d=1,2,\cdots. \tag{2.34}$$

Since it is evident that $\phi(t)$ can be analytically extended onto all of R, $\phi(t)$ is the characteristic function of a probability distribution G with density

$$g(u)=\begin{cases}\dfrac{d}{2\pi}\,\dfrac{\sqrt{(u-b_1)(b_2-u)}}{u}, & b_1\leqq u\leqq b_2,\\ 0 & \text{otherwise.}\end{cases} \tag{2.35}$$

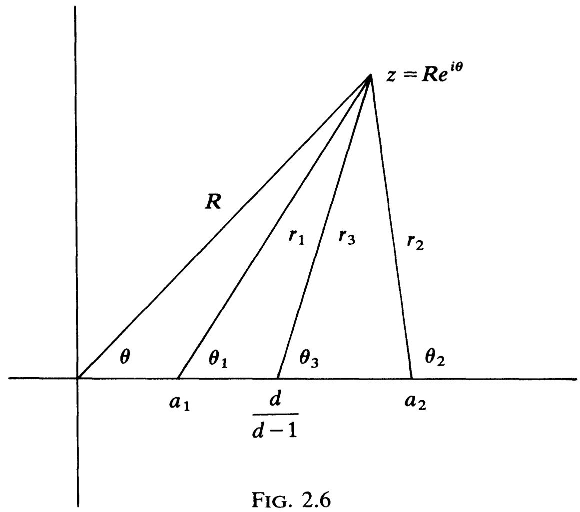

FIG. 2.6

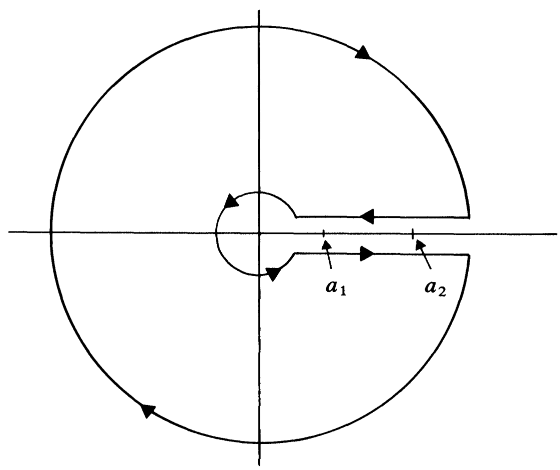

FIG. 2.7

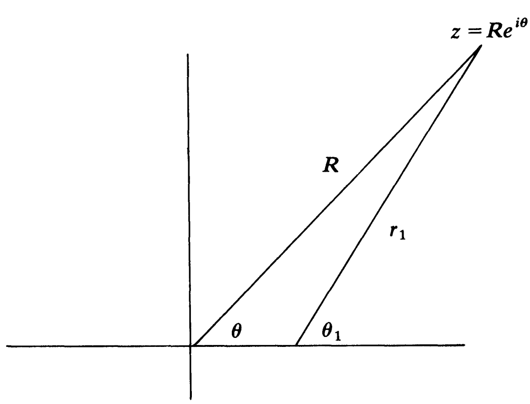

FIG. 2.8

Since $\phi(t)$ is analytic in a neighborhood of 0, G is the *only* distribution with moments

$$f(r)=\int_{b_1}^{b_2} u^r g(u)\,du, \qquad r=0,1,2,\cdots, \tag{2.36}$$

so that F defined in (2.1) is uniquely determined by its moments.

We can now complete the theorem.

Let $\mu_r^{(n)}=(1/n)\operatorname{tr} W^r=(1/n)\sum_{i=1}^{n}\lambda_i^r$. We know that

$$\mu_r^{(n)}\to\int_{b_1}^{b_2} u^r g(u)\,du \tag{2.37}$$

in probability as $n\to\infty$. Given any subsequence $\{n_j'\}_{j=1}^{\infty}$ we can find, using a diagonal argument, a subsequence $\{n_j'\}\subseteq\{n_j\}_{j=1}^{\infty}$ such that

$$\mu_r^{(n_j)}\overset{\text{a.s.}}{\to}\int_{b_1}^{b_2} u^r g(u)\,du \tag{2.38}$$

for all r. Since F is continuous we have for each λ

$$F_{n_j'}(\lambda)\overset{\text{a.s.}}{\to}F(\lambda). \tag{2.39}$$

Since this could be done for any subsequence of the natural numbers we conclude that

$$F_n(\lambda)\to F(\lambda) \tag{2.40}$$

in probability.

This concludes the proof of the theorem.

Figure 2.9 contains plots of $g(u)$ for $d=1,5,10,20$. They should be compared with the graphs in Figs. 1.3 and 1.4.

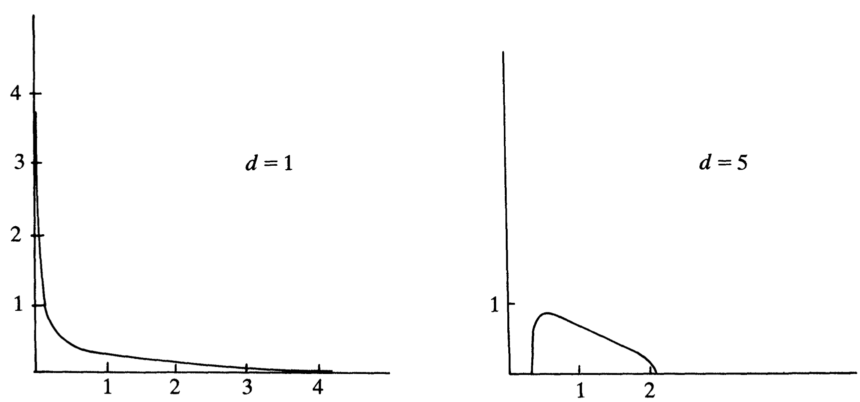

FIG. 2.9

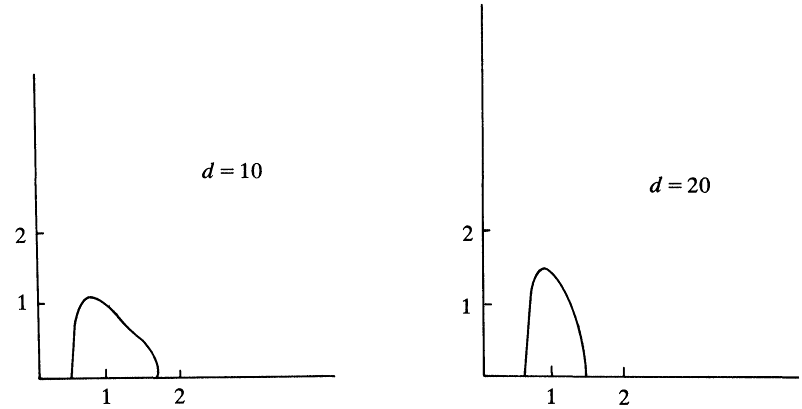

FIG. 2.9. Continued

3. Remarks and conclusions. Symmetric random matrices have been dealt with in the literature in two contexts. One is multivariate analysis. The other is physics, where the pioneering work of Wigner [2], [3] led to the results which have some resemblance to our theorem. He investigated the limiting spectra of symmetric matrices having independent and symmetrically distributed (symmetric about 0) elements, all sharing a common second moment, and proved the famous semi-circle law. Wigner's work has been extended by Grenander [5] and Arnold [6]; see also Dyson [4].

In our situation we have dependent elements in the matrix that complicates the analysis. A partial result in this case can be found in Arharov [7].

The simplest types of Toeplitz forms are those whose entries a_{ij} depend only on $|i-j|$: a stationarity condition. Such matrices have been studied in depth and the limiting spectral measures are known (Grenander and Szego [8]). In our case the matrix of probabilities generating the random matrix satisfies a stationarity condition and this might lead one to expect a relationship between the present analysis and Toeplitz theory. However, there seems to be no connection in substance.

As far as we know this is the first attempt to analyze randomly connected networks in terms of the spectrum they induce. We consider an ensemble (population) of organisms whose network topologies may differ drastically on the microscopic level. Our main result shows that in spite of this the global, spectral properties are essentially the same: for large networks the local randomness does not influence the spectral measure so that a law of large numbers has been established. This could be compared with the classical way of deriving thermodynamic laws from statistical principles applied to the underlying mechanical systems.

To continue this metaphor: it could be said that we have investigated only the situation of (statistical) equilibrium. Evolutionary situations on the other hand correspond, in our context, to learning and adaptation. We are currently investigating how the spectrum will develop in time in such cases.

We do not claim any biological evidence to support the information processing in the forward-backward manner implied by the W operator. Instead our study should be viewed as a "Gedanken-experiment," where a postulated controlled randomness generates the physiology of connections. A mathematical technique is then developed for the spectral analysis of the network. It is remarkable that although (under our hypothesis) two members of the species have little in common on the microscopic level, they tend to have the same spectrum. We believe that the technique used will be of value for more realistic models.

REFERENCES

[1] E. C. TITCHMARSH, *The Theory of Functions*, 2nd ed., Oxford University Press, London, 1939.

[2] E. P. WIGNER, *Characteristic vectors of bordered matrices with infinite dimensions*, Ann. Math., 62 (1955), pp. 548–564.

[3] ———, *On the distribution of the roots of certain symmetric matrices*, Ibid., 67 (1958), pp. 325–327.

[4] F. J. DYSON, *A Brownian-motion model for the eigenvalues of a random matrix*, J. Math. Phys., 3 (1962), pp. 1191–1198.

[5] U. GRENANDER, *Probabilities on Algebraic Structures*, Almqvist and Wiksell, Stockholm, 1963.

[6] L. ARNOLD, *On the asymptotic distribution of the eigenvalues of random matrices*, J. Math. Anal., 20 (1967), pp. 262–268.

[7] L. V. ARHAROV, *Limit theorems for the characteristic roots of a sample covariance matrix*, Soviet Math. Dokl., 12 (1971), pp. 1206–1209.

[8] U. GRENANDER AND G. SZEGÖ, *Toeplitz forms and their applications*, University of California Press, Berkeley and Los Angeles, 1958.