The Annals of Probability
1980, Vol. 8, No. 2, 252–261

A LIMIT THEOREM FOR THE NORM OF RANDOM MATRICES

BY STUART GEMAN

Brown University

This paper establishes an almost sure limit for the operator norm of rectangular random matrices: Suppose $\{v_{ij}\} i = 1, 2, \cdots, j = 1, 2, \cdots$ are zero mean i.i.d. random variables satisfying the moment condition $E|v_{11}|^n \leq n^{\alpha n}$ for all $n \geq 2$, and some α. Let $\sigma^2 = Ev_{11}^2$ and let V_{pn} be the $p \times n$ matrix $\{v_{ij}\}_{1 \leq i \leq p;\, 1 \leq j \leq n}$. If p_n is a sequence of integers such that $p_n/n \to y$ as $n \to \infty$, for some $0 < y < \infty$, then $1/n|V_{p_n n} V_{p_n n}^T| \to (1 + y^{\frac{1}{2}})^2\sigma^2$ almost surely, where $|A|$ denotes the operator ("induced") norm of A. Since $1/n|V_{p_n n} V_{p_n n}^T|$ is the maximum eigenvalue of $1/nV_{p_n n} V_{p_n n}^T$, the result relates to studies on the spectrum of symmetric random matrices.

Generate an increasing sequence of matrices by choosing new elements i.i.d. We will discuss the existence of a limit for the induced norm of such matrices. The motivation is twofold: (1) Several authors (cf. [1]–[11]) have identified limiting spectral distributions for sequences of symmetric real-valued random matrices. (See discussion following the theorem.) In connection with these results, it is natural to ask for the behavior of the largest eigenvalue, but this can not be inferred from the limiting spectral distribution. In many cases this behavior is identified by the theorem below, since, for symmetric matrices, the induced norm coincides with the largest eigenvalue. (2) We have been studying large and randomly connected systems in the hope of finding limit laws (CLT's and LLN's) for the behavior of the system as its size grows to ∞. The "connectivity" of these systems is described by a random matrix of the type studied here, and in several examples the desired limit can be established based on the theorem below.

To illustrate the theorem, let $\{v_{ij}\} i = 1, 2, \cdots, j = 1, 2 \cdots$ be i.i.d. $N(0, 1)$. Define V_{nn} to be the $n \times n$ matrix

$$\{v_{ij}\}_{1 \leq i \leq n;\, 1 \leq j \leq n},$$

and, for any $n \times n$ matrix M, define

$$|M| = \sup_{x \in R^n;\, |x| = 1} |Mx|$$

(using Euclidean norm in the supremum). How does $1/n|V_{nn} V_{nn}^T|$ (i.e., $|1/n^{\frac{1}{2}} V_{nn}|^2$) behave for large n?

An upper bound is

$$\Sigma_{ij}\left(\frac{v_{ij}}{n^{\frac{1}{2}}}\right)^2,$$

Received July 18, 1978.

AMS 1970 *subject classifications*. Primary 60F15; secondary 60C05.

Key words and phrases. Random matrices, maximum eigenvalue of symmetric random matrices, norm of random matrices.

but this goes to ∞ almost surely. On the other hand, if $u_i(n) = 1/n^{\frac{1}{2}}(v_{i1}, \cdots, v_{in})$, then $u_i(n) \cdot u_j(n) \to \delta_{ij}$ a.s., so $1/n^{\frac{1}{2}}V_{nn}$ "looks" orthogonal when n is large. This suggests $|1/n^{\frac{1}{2}}V_{nn}| \to 1$. In fact, $|1/n^{\frac{1}{2}}V_{nn}| \to 2$ a.s., which is a special case of:

THEOREM. *Suppose* $\{v_{ij}\} i = 1, 2, \cdots, j = 1, 2, \cdots$ *are* i.i.d. *random variables with*

(a) $Ev_{11} = 0$, *and*

(b) $E|v_{11}|^n \leqslant n^{\alpha n}$ *for all* $n \geqslant 2$, *some* α.

Let $\sigma^2 = Ev_{11}^2$ *and*

$$V_{pn} = \{v_{ij}\}_{1 \leqslant i \leqslant p;\ 1 \leqslant j \leqslant n}.$$

If p_n *is a sequence of integers such that*

$$\lim_{n\to\infty} \frac{p_n}{n} \to y \qquad \text{for some} \qquad 0 < y < \infty,$$

then

$$\frac{1}{n}|V_{p_n n} V_{p_n n}^T| \to \left(1 + y^{\frac{1}{2}}\right)^2 \sigma^2 \qquad \text{a.s.}$$

Of course $1/n|V_{p_n n} V_{p_n n}^T| = \lambda_{\max}(n)$ (the maximum eigenvalue of $1/nV_{p_n n}V_{p_n n}^T$), so the theorem is relevant to studies of the spectrum of symmetric random matrices. If $M_p, p = 1, 2, \cdots$ is a sequence of such matrices, with eigenvalues $\lambda_i(p), i = 1, 2, \cdots p$, then we can define a sequence of random distribution functions by

$$(1) \qquad F_p(x) = \frac{1}{p}\{\text{number of } \lambda_i(p) \leqslant x\}.$$

Beginning with Wigner ([10] and [11]), and depending upon how M_p is constructed, the asymptotics of $F_p(x)$ have been the topic of numerous papers (cf. [1] through [9]). It can often be shown that $F_p(x)$ converges almost surely to a specified nonrandom distribution function. For the most up-to-date results in this direction, see the papers by Jonsson [4] and Wachter [8].

Let $\lambda_i(p_n) i = 1, 2, \cdots p_n$ be the eigenvalues of $1/nV_{p_n n}V_{p_n n}^T$, and define $F_{p_n}(x)$ by (1). Under conditions somewhat more relaxed than those of the theorem, a nonrandom distribution function $F_y(x)$ can be identified such that $F_{p_n}(x) \to F_y(x)$ uniformly in x with probability one (see Wachter [7], Theorem 7.7 or Jonsson [4], Theorem 3.2). Since the support of $F_y(x)$ is $[(1 - y^{\frac{1}{2}})^2\sigma^2, (1 + y^{\frac{1}{2}})^2\sigma^2]$, together with $\{0\}$ when $y > 1$, this result implies

$$\liminf \lambda_{\max}(n) \geqslant \left(1 + y^{\frac{1}{2}}\right)^2 \sigma^2 \qquad \text{a.s.}$$

But the other inequality,

$$\limsup \lambda_{\max}(n) \leqslant \left(1 + y^{\frac{1}{2}}\right)^2 \sigma^2 \qquad \text{a.s.},$$

requires special treatment.

PROOF OF THE THEOREM. Fix $z > (1 + y^{\frac{1}{2}})^2\sigma^2$. We will show

$$\limsup \lambda_{\max}(n) \leqslant z \qquad \text{a.s.}$$

by showing

(2) $$\Sigma_{n=1}^{\infty} E\left[\left(\frac{\lambda_{\max}(n)}{z}\right)^{k_n}\right] < \infty,$$

where

(3) $$k_n = [w \log n]$$

= greatest integer less than or equal to $w \log n$, and w is any constant satisfying

(4) $$w > \left[\frac{3}{\log\left(z/\left(1 + y^{\frac{1}{2}}\right)^2 \sigma^2\right)}\right]$$

and

(5) $$w > 5.$$

Usually, the subscript, n, in k_n and p_n will be dropped.

For a bound on $E\lambda_{\max}(n)^k$ we have

(6) $$E\lambda_{\max}(n)^k \leqslant E \operatorname{tr}\left(\frac{1}{n} V_{pn} V_{pn}^T\right)^k$$
$$= \left(\frac{1}{n}\right)^k \Sigma_{1 \leqslant i_1, i_2, \cdots i_k \leqslant p} \Sigma_{1 \leqslant j_1, j_2, \cdots j_k \leqslant n} E v_{i_1 j_1} v_{i_2 j_1} v_{i_2 j_2} \cdots v_{i_k j_k} v_{i_1 j_k}.$$

The discussion of (6) will make use of the following definitions:

1. A "V-path" is an ordered sequence of $2k$ elements of V_{pn} (with repetitions allowed) such that
   (a) the first element is arbitrary,
   (b) the second element is in the same column as the first element, the third element is in the same row as the second element, etc.,
   (c) the last element of the path is in the same row as the first element of the path,
   (d) every element appearing in the path appears more than once.

Notice that every nonzero contribution to (6) is from a V-path.

2. r (c) will denote the total number of rows (columns) entered by a given V-path.

Evidently, $r + c$ must satisfy $2 \leqslant r + c \leqslant k + 1$.

3. β_l is the number of V-paths such that $r + c = l$.
4. α_l will denote an upper bound on

$$E v_{i_1 j_1} v_{i_2 j_1} v_{i_2 j_2} \cdots v_{i_k j_k} v_{i_1 j_k},$$

given that the v_{ij}'s form a V-path with $r + c = l$.

With these definitions we can write

$$E\lambda_{\max}(n)^k \leqslant \left(\frac{1}{n}\right)^k \Sigma_{l=2}^{k+1} \alpha_l \beta_l.$$

Accept, for now, the following two lemmas:

LEMMA 1. *Two suitable choices for α_l are*
(a) $\alpha_l \leqslant (2k)^{2\alpha k}$, *and*
(b) $\alpha_l \leqslant (\sigma^2)^k (6k)^{6\alpha(k-l)+6\alpha}$.

LEMMA 2. *For all n sufficiently large:*
(a) *There exists a constant β such that*

$$\beta_l \leqslant \beta^l n^l l^{4k}$$

for all $2 \leqslant l \leqslant k+1$.
(b) *There exist constants ξ_1 and ξ_2 such that*

$$\beta_l \leqslant \left(1 + (p/n)^{\frac{1}{2}}\right)^{2k} n^l k^{\xi_1(k-l)+\xi_2}$$

whenever

$$[(w-1)\log n] \leqslant l \leqslant k+1.$$

Then

$$E\left(\frac{\lambda_{\max}(n)}{z}\right)^k \leqslant \left(\frac{1}{z}\right)^k \left(\frac{1}{n}\right)^k \left\{\sum_{l=2}^{[(w-1)\log n]} \alpha_l \beta_l + \sum_{l=[(w-1)\log n]}^{k+1} \alpha_l \beta_l\right\} \tag{7}$$

$$\leqslant \left(\frac{1}{z}\right)^k \left(\frac{1}{n}\right)^k \sum_{l=2}^{[(w-1)\log n]} (2k)^{2\alpha k} \beta^l n^l l^{4k}$$

$$+ \left(\frac{1}{z}\right)^k \left(\frac{1}{n}\right)^k \sum_{l=[(w-1)\log n]}^{k+1} (\sigma^2)^k (6k)^{6\alpha(k-l)+6\alpha} \left(1 + \left(\frac{p}{n}\right)^{\frac{1}{2}}\right)^{2k} n^l k^{\xi_1(k-l)+\xi_2} \tag{8}$$

for large n. Letting A_n denote the expression in (7), and B_n denote the expression in (8):

$$A_n \leqslant \left(\frac{1}{z}\right)^k \left(\frac{1}{n}\right)^k (2k)^{2\alpha k} \beta^k k^{4k} \sum_{l=2}^{[(w-1)\log n]} n^l$$

$$\leqslant \left(\frac{2^{2\alpha}\beta}{z}\right)^k k^{(2\alpha+4)k+1} n^{1-\log n}.$$

From (3),

$$\frac{\log A_n}{\log n} < -2 \qquad \text{when } n \text{ is large.}$$

So A_n is summable.

$$B_n = \left[\frac{\left(1 + \left(\frac{p}{n}\right)^{\frac{1}{2}}\right)^2 \sigma^2}{z}\right]^k 6^{6\alpha} k^{6\alpha+\xi_2} \sum_{l=[(w-1)\log n]}^{k+1} \left(\frac{6^{6\alpha} k^{6\alpha+\xi_1}}{n}\right)^{k-l}.$$

When n is large, replace the sum by k times its last term:

$$B_n \leqslant \left[\frac{\left(1 + \left(\frac{p}{n}\right)^{\frac{1}{2}}\right)^2 \sigma^2}{z}\right]^k 6^{6\alpha} k^{6\alpha+\xi_2+1} \frac{n}{6^{6\alpha} k^{6\alpha+\xi_1}}$$

$$= \left[\frac{\left(1 + \left(\frac{p}{n}\right)^{\frac{1}{2}}\right)^2 \sigma^2}{z}\right]^k k^{\xi_2-\xi_1+1} n.$$

Since $p/n \to y$, (3) and (4) imply

$$\frac{\log B_n}{\log n} < -2 \text{ for large } n.$$

Hence B_n is also summable.

It remains to prove the lemmas.

PROOF OF LEMMA 1.

(a) Suppose

$$v_{i_1, j_1} v_{i_2 j_1} v_{i_2 j_2} \cdots v_{i_k j_k} v_{i_1 j_k}$$

is a V-path with (say)

$$n_1 \; v_{k_1 l_1}\text{'s},$$

$$n_2 \; v_{k_2 l_2}\text{'s},$$

$$\vdots$$

and

$$n_f \; v_{k_f l_f}\text{'s}.$$

Then $n_1 + n_2 + \cdots + n_f = 2k$ and $n_i \geqslant 2 \qquad i = 1, 2, \cdots f$.
Hence

$$E v_{i_1 j_1} v_{i_2 j_1} v_{i_2 j_2} \cdots v_{i_k j_k} v_{i_1 j_k} \leqslant n_1^{\alpha n_1} \cdots n_f^{\alpha n_f}$$
$$\leqslant (2k)^{2\alpha k}.$$

(b) If a V-path has $r + c = l$, then it must contain at least $l - 1$ distinct v_{ij}'s. Hence, there must be at least

$$(l - 1) - \{2k - 2(l - 1)\} = 3l - 2k - 3$$

v_{ij}'s appearing exactly twice. It follows that

$$E v_{i_1 j_1} v_{i_2 j_1} v_{i_2 j_2} \cdots v_{i_k j_k} v_{i_1 j_k}$$
$$\leqslant (\sigma^2)^{\max(0,\, 3l-2k-3)} \{6(k - l) + 6\}^{6\alpha(k-l)+6\alpha}$$

(where we have used the reasoning of part (a) to bound the contribution from the

$6(k-l)+6v_{ij}$'s possibly not appearing in pairs)

$$\leqslant (\sigma^2)^k(6k)^{6\alpha(k-l)+6\alpha}.$$

PROOF OF LEMMA 2. Define a "canonical V-path" to be a V-path with the following properties:

1. the first element is v_{11},
2. each time a new row (column) is entered, it is the next available row (column) (i.e., that empty row (column) with the lowest index).

Let $m_{r,c}$ be the number of canonical V-paths having r rows and c columns.

If we relabel the rows and the columns of any V-path, in such a way that distinct rows and distinct columns remain distinct, then we have again a V-path. E.g., for $k=3$,

$$v_{57}v_{27}v_{21}v_{21}v_{27}v_{57}$$

is a V-path. Making the row associations

$$5\to 4 \qquad 2\to 3$$

and the column associations

$$7\to 2 \qquad 1\to 3$$

gives the new V-path

$$v_{42}v_{32}v_{33}v_{33}v_{32}v_{42}.$$

By labeling the first row (column) entered "1", and the next new row (column) entered "2", and etc., we can associate a unique canonical V-path with every V-path. In the above example, the V-path

$$v_{57}v_{27}v_{21}v_{21}v_{27}v_{57}$$

is associated with the canonical V-path

$$v_{11}v_{21}v_{22}v_{22}v_{21}v_{11}.$$

Assuming that p and n are greater than k, there are

$$\frac{p!}{(p-r)!}\,\frac{n!}{(n-c)!}$$

V-paths associated with every canonical V-path containing r rows and c columns. Hence

$$\beta_l = \Sigma_{r=1}^{l-1} m_{r,l-r}\frac{p!}{(p-r)!}\,\frac{n!}{(n-(l-r))!} \tag{9}$$
$$\leqslant \Sigma_{r=1}^{l-1} m_{r,l-r}p^r n^{l-r}$$
$$= n^l \Sigma_{r=1}^{l-1} m_{r,l-r}\left(\frac{p}{n}\right)^r.$$

So, for part (a), if we take $\beta \geqslant 1$ and

$$\beta \geqslant \sup_{n\geqslant 1}\frac{p_n}{n} \qquad \left(\text{recall that } \frac{p_n}{n}\to y\right),$$

then

$$\beta_l \leqslant \beta^l n^l \Sigma_{r=1}^{l-1} m_{r,\,l-r}.$$

Any canonical V-path with $r + c = l$ is contained in the upper left $l \times l$ submatrix of V_{pn}. Since each V-path contains $2k$ elements, there are certainly less than

$$(l^2)^{2k}$$

canonical V-paths with $r + c = l$:

$$\Sigma_{r=1}^{l-1} m_{r,\,l-r} \leqslant l^{4k},$$

which proves (a).

For (b), we will first develop a bound on $m_{r,\,l-r}$. Make the following definitions: An element of a canonical V-path is a "row innovation" ("column innovation") if it is the first entry into a row (column). The first element of a canonical V-path will be considered a column innovation. We distinguish 4 types of elements:

type 1: row innovations
type 2: column innovations
type 3: elements which are the first to repeat a row or column innovation
type 4: all other elements.

When $r + c = l$, there are

$r - 1$ type 1 elements,
$l - r$ type 2 elements,
$l - 1$ type 3 elements, and
$2(k - l) + 2$ type 4 elements.

The $r - 1$ type 1 elements must be distributed among the k elements which result from row moves (elements taken by choosing from the column of the previous element). The $l - r$ type 2 elements must be distributed among the k elements which result from column moves (to which category we assign the first element). And, finally, the $l - 1$ type 3 elements must be distributed among the remaining $2k - (r - 1) - (l - r) = 2k - l + 1$ elements. Hence, the number of ways in which the four types of elements can be distributed among the $2k$ elements forming a canonical V-path is bounded by

$$\binom{k}{r-1}\binom{k}{l-r}\binom{2k-l+1}{l-1}. \tag{10}$$

Next, we bound the number of canonical V-paths associated with each distribution of the four types of elements. Begin with the observation that a row (column) innovation is always unambiguous; it must be that element of the same column (row) as the previous element, but in the empty row (column) with lowest index.

Each of the type 4 elements can be chosen from at most k elements of V_{pn}, since the entire canonical V-path is contained in the $k \times k$ upper left submatrix of V, and any given entry in the path must be chosen from a particular row or column. Therefore, the type 4 elements introduce at most the factor

$$k^{2(k-l)+2}. \tag{11}$$

We can distinguish 3 types of type 3 elements:

(a) Those which follow an innovation element. These elements are unambiguous, as they must repeat the previous element (there being no other element in the relevant row or column).

(b) Those which follow a type 4 element. Repeat the argument used in the discussion of type 4 elements to show that these elements can introduce at most the factor

(12) $$k^{2(k-l)+2}.$$

(c) Those which follow type 3 elements.

Here, we reason as follows:

Fix some column "j". j is first entered by an innovation element. The next element must be taken from j, and if it too is an innovation, then j will have 2 unpaired innovation elements. Then, the following element either pairs the previous element or it is taken from a different column. In any case, at this point in the V-path there are at most two unpaired innovation elements in j. There can never be more than two unpaired innovation elements in j, unless j is first reentered from another column by a type 4 element. Then, once the V-path again leaves j, there will be at most one additional unpaired innovation element.

And, of course, the same discussion applies to rows.

Now consider a type 3 element of the c type. Suppose, for example, that it must be chosen from the same column as the previous element (which was also of type 3). If this choice is ambiguous, then before the preceding type 3 element there must have been 3 or more unpaired innovation elements in that column. Evidently, this ambiguity can arise at most as many times as there are type 4 elements. In other words, there are no more than $2(k - l) + 2$ type 3 elements of the c type for which there is an ambiguity.

For a given path, let i be the number of type 3 elements of the c type for which there is an ambiguous choice. These may be distributed among the $l - 1$ type 3 elements in no more than

$$\binom{l-1}{i}$$

ways. Each such element certainly represents less than k choices, from which we arrive at the factor

$$\binom{l-1}{i}k^i.$$

So, the factor introduced by type 3 elements of the c type is bounded by

(13) $$\sum_{i=0}^{2(k-l)+2}\binom{l-1}{i}k^i.$$

In light of (3) and (5), and the restriction $[(w - 1)\log n] \leqslant l \leqslant k + 1$, the largest term in (13), for large n, is the $i = 2(k - l) + 2$ term (since $2(k - l) + 2 < (l -$

1)/2 for large n, and $\binom{l-1}{i}$ is maximum near $(l-1)/2$). Hence (13) is bounded by

$$k\binom{l-1}{2(k-l)+2}k^{2(k-l)+2} \tag{14}$$

for large n.

Putting together (10), (11), (12) and (14):

$$m_{r,l-r} \leqslant \binom{k}{r-1}\binom{k}{l-r}\binom{2k-l+1}{l-1}(k^{2(k-l)+2})(k^{2(k-l)+2})$$

$$\times k\binom{l-1}{2(k-l)+2}k^{2(k-l)+2}$$

$$= \binom{k}{r-1}\binom{k}{l-r}\frac{(2k-l+1)!}{\{(2(k-l)+2)!\}^2(3l-2k-3)!}k^{6(k-l)+7}$$

$$\leqslant \binom{k}{r}^2 \frac{\{(k-r)!\}^2\{r!\}^2}{(k-r+1)!\,(r-1)!\,(k-l+r)!\,(l-r)!}\,\frac{(2k-l+1)!}{(3l-2k-3)!}k^{6(k-l)+7}$$

$$\leqslant \binom{k}{r}^2 \frac{(k-r)!\,r!\,r}{(l-r)!\,(k-l+r)!}\{(2k)^{4(k-l)+4}\}k^{6(k-l)+7}$$

$$\leqslant \binom{k}{r}^2 k^{12(k-l)+14}$$

for large n. Then, using (9),

$$\beta_l \leqslant n^l k^{12(k-l)+14}\Sigma_{r=1}^{l-1}\binom{k}{r}^2\left(\frac{p}{n}\right)^r$$

$$\leqslant n^l k^{12(k-l)+14}\left(\Sigma_{r=0}^{k}\binom{k}{r}\left(\frac{p}{n}\right)^{r/2}\right)^2$$

$$= n^l k^{12(k-l)+14}\left(1+\left(\frac{p}{n}\right)^{\frac{1}{2}}\right)^{2k},$$

which completes the proof.

Acknowledgment. The approach was suggested by Professor Grenander: try to show (2) by using the bound $\lambda_{\max}(n)^k \leqslant \mathrm{tr}(1/nV_{pn}V_{pn}^T)^k$. Moreover, his assurance that the conjecture was true made it much easier to prove.

REFERENCES

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

[2] ARNOLD, L. (1971). On Wigner's semicircle law for the eigenvalues of random matrices. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **19** 191–198.

[3] GRENANDER, U. and SILVERSTEIN, J. W. (1977). Spectral analysis of networks with random topologies. *SIAM J. Appl. Math.* **32** 499–519.

[4] JONSSON, D. (1976). Some limit theorems for the eigenvalues of a sample covariance matrix. Uppsala University Dept. of Maths. Rept. No. 1976:6, Uppsala

[5] MARCENKO, V. A. and PASTUR, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Math. USSR-Sbornik*, **1** 457–483.

[6] OLSON, W. H. and UPPULURI, V. R. R. (1972). Asymptotic distribution of eigenvalues of random matrices. *Sixth Berkeley Symposium on Math. Statist. Probability* **3** 615–644.

[7] WACHTER, K. W. (1974). Foundations of the asymptotic theory of random matrix spectra. Ph.D. Dissertation. Trinity College, Cambridge and St. Catherine's College, Oxford.

[8] WACHTER, K. W. (1978). The strong limits of random matrix spectra for sample matrices of independent elements. *Ann. Prob.* **6** 1–18.

[9] WEGMANN, R. (1976). The asymptotic eigenvalue-distribution for a certain class of random matrices. *J. Math. Anal. Appl.* **56** 113–132.

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

[11] WIGNER, E. P. (1958). On the distribution of the roots of certain symmetric matrices. *Annals Math.* **67** 325–327.

DIVISION OF APPLIED MATHEMATICS
BROWN UNIVERSITY
PROVIDENCE, RHODE ISLAND 02912