SIAM J. MATRIX ANAL. APPL.
Vol. 9, No. 4, October 1988

008

EIGENVALUES AND CONDITION NUMBERS OF RANDOM MATRICES*

ALAN EDELMAN†

Abstract. Given a random matrix, what condition number should be expected? This paper presents a proof that for real or complex $n \times n$ matrices with elements from a standard normal distribution, the expected value of the log of the 2-norm condition number is asymptotic to $\log n$ as $n \to \infty$. In fact, it is roughly $\log n + 1.537$ for real matrices and $\log n + 0.982$ for complex matrices as $n \to \infty$. The paper discusses how the distributions of the condition numbers behave for large n for real or complex and square or rectangular matrices. The exact distributions of the condition numbers of $2 \times n$ matrices are also given.

Intimately related to this problem is the distribution of the eigenvalues of Wishart matrices. This paper studies in depth the largest and smallest eigenvalues, giving exact distributions in some cases. It also describes the behavior of all the eigenvalues, giving an exact formula for the expected characteristic polynomial.

Key words. characteristic polynomial, condition number, eigenvalues, random matrices, singular values, Wishart distribution

AMS(MOS) subject classification. 15A52

1. Introduction. What is the condition number of a random matrix? Though we were originally motivated by this question, the problem quickly becomes one of studying the eigenvalues of a related random matrix.

This application of random eigenvalues originally appeared in a classic paper by von Neumann and Goldstine [22]. Further applications can be found in statistics and physics (see, e.g., [7], [25]). Statisticians use random eigenvalues in principal component analysis, multiple discriminant analysis, and canonical correlation analysis. Physicists model nuclear levels with eigenvalues.

When speaking of a random matrix, we will focus on the Gaussian and Wishart distributions. We say that a matrix X has the Gaussian distribution if each element of the matrix comes from an independent standard normal distribution. We obtain Wishart matrices from Gaussian matrices by forming XX^T. Wishart matrices are of intrinsic interest because they are essentially the sample covariance matrices for multivariate Gaussian distributions, as discussed in books on multivariate statistics such as [25].

Various researchers have investigated the eigenvalues of a Wishart matrix from a number of points of view. If we take a large matrix from a Wishart distribution, we may sort and plot the eigenvalues against their position index. A theory of what the picture should be is developed in [13], [16], [21], and [23]. Estimates of the largest and smallest eigenvalues are given in [9] and [17]. A complicated expression for the distribution of the largest eigenvalue is given in [19] and for the smallest eigenvalue in [15].

Our question about condition numbers was introduced in a precise format in [18]. In effect, Smale asks for the expected geometric mean of the condition number of a Gaussian matrix. Precisely, let X_n be an $n \times n$ matrix whose elements are independent standard normal random variables. Let $\kappa_{X_n} = \|X_n\| \|X_n^{-1}\|$ be its condition number in the 2-norm. What is the expected value of $\log \kappa_{X_n}$? The reason we use $\log \kappa_{X_n}$ is that this quantity is the measure of the loss of numerical precision (see [6]). The result of directly averaging the condition number, on the other hand, is known to be infinite. Kostlan and

* Received by the editors November 25, 1987; accepted for publication (in revised form) April 4, 1988. This research was supported by a Hertz Foundation Fellowship and by an IBM Faculty Development Award to Lloyd N. Trefethen.

† Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139.

Ocneanu (see [18]) obtained some estimates showing that for all $\varepsilon > 0$, when n is sufficiently large,

$$\frac{2}{3} - \varepsilon \leqq \frac{E(\log \kappa_{X_n})}{\log n} \leqq \frac{5}{2} + \varepsilon.$$

Kostlan has communicated to me a new result that raises the lower bound to 1 [14]. In the present paper, we show that this new result is sharp: $E(\log \kappa_{X_n}) \sim \log n$ as $n \to \infty$. The same leading behavior holds for complex matrices, but we have more precise estimates. We also explore asymptotic results for rectangular matrices.

A natural first step in conducting this investigation was to run some numerical experiments. In Table 1.1, we list the result of averaging log condition numbers from random samples of 1000 square matrices of dimension equal to various powers of 2. Also listed are the results for 1000 matrices of dimensions 100×200. The data for square matrices clearly suggest $E(\log \kappa_{X_n}) \sim \log n$ for both the real and complex cases, and we might perhaps predict that $E(\log \kappa_{X_n}) = \log n + c + o(1)$ for some constant c. In § 6, we derive the constant c (≈ 1.537 for real matrices and ≈ 0.982 for complex matrices). We also show that for large (real or complex) matrices the condition number depends on the ratio of rows to columns m/n. For example, matrices with twice as many columns as rows have an expected log condition number asymptotic to 1.76. It is of interest that this value is finite. In the table we see that the asymptotic result gives a usable approximation for the finite case.

In Table 1.2, we summarize our results about condition numbers in the limit $n \to \infty$. (Please consult the text for details not explained here.) The values listed are the exponentials of the expected logarithms of three random variables: the condition number of the Gaussian matrix and the largest and smallest eigenvalues of the related Wishart matrix. Note that this first quantity is the ratio of the square root of the other two quantities. As a kind of table of contents, the table lists where these results are stated explicitly in the text. K_2 is in fact $2e^{\gamma/2}$, where γ is Euler's constant, ≈ 0.5772. K_1 is a little more complicated. It is given by γ and a readily evaluated definite integral. For the rectangular matrices, the variable y denotes the ratio m/n, where $0 < y < 1$.

For the special case of real and complex $2 \times n$ matrices we can specify exactly the distributions of condition numbers and eigenvalues; these results are reported in § 7. We comment about the tail of the condition number distribution in § 8. We look at the complete spectrum of a Wishart matrix in § 9 and derive further exact distributions in § 10.

TABLE 1.1
Average log condition numbers.

	Real		Complex	
n	Avg.	Avg. $-\log n$	Avg.	Avg. $-\log n$
2	1.53	0.84	1.19	0.49
4	2.63	1.25	2.09	0.70
8	3.46	1.38	2.91	0.83
16	4.24	1.47	3.65	0.88
32	4.93	1.47	4.35	0.88
64	5.64	1.48	5.06	0.90
128	6.44	1.59	5.78	0.93
256	7.04	1.49	6.50	0.96
100×200	1.72		1.67	

TABLE 1.2
Exponentials of expected logs ($K_1 \approx 4.65$, $K_2 \approx 2.67$).

		Real	Complex
Square	κ	$K_1 n$ Thm. 6.1	$K_2 n$ Thm. 6.2
	$\lambda_{\max}$	$4n$ Prop. 4.1	$8n$ Prop. 4.2
	$\lambda_{\min}$	$\frac{4}{K_1^2 n}$ Cor. 3.2	$\frac{8}{K_2^2 n}$ Cor. 3.4
Rectangular	κ	$\frac{1+\sqrt{y}}{1-\sqrt{y}}$ Thm. 6.3	$\frac{1+\sqrt{y}}{1-\sqrt{y}}$ Thm. 6.3
	$\lambda_{\max}$	$n(1+\sqrt{y})^2$ Prop. 4.1	$2n(1+\sqrt{y})^2$ Prop. 4.2
	$\lambda_{\min}$	$n(1-\sqrt{y})^2$ Prop. 5.1	$2n(1-\sqrt{y})^2$ Prop. 5.2

2. Gaussian and Wishart matrices. We are interested in rectangular Gaussian matrices, that is, $m \times n$ matrices all of whose components are independent standard normal variables. We denote such a random matrix (or its distribution) by $G(m, n)$. $G(m, n)$ has the symmetry property that it is invariant under orthogonal transformations (i.e., isotropic).

A derived random matrix is the $m \times m$ Wishart matrix $W(m, n)$ defined by $M = XX^T$, where X has the distribution $G(m, n)$. We will focus on the eigenvalues of M, $\lambda_{\max} = \lambda_1 \geqq \cdots \geqq \lambda_m = \lambda_{\min} \geqq 0$, since they are the squares of the singular values of X, and the 2-norm condition number of X is $\sqrt{\lambda_{\max}/\lambda_{\min}}$.

Remarkably enough, the exact joint density function for the m eigenvalues of M can be written as

$$K_{n,m} \exp\left(-\frac{1}{2}\sum_{i=1}^{m}\lambda_i\right)\prod_{i=1}^{m}\lambda_i^{(n-m-1)/2}\prod_{i<j}(\lambda_i-\lambda_j)\,d\lambda_1\cdots d\lambda_m, \tag{1}$$

where

$$K_{n,m}^{-1} = \left(\frac{2^n}{\pi}\right)^{m/2}\prod_{i=1}^{m}\Gamma\left(\frac{n-i+1}{2}\right)\Gamma\left(\frac{m-i+1}{2}\right) \tag{2}$$

(see [12] or [25]).

We may further define complex Wishart matrices $\tilde{M} = \tilde{X}\tilde{X}^T$, where $\tilde{X}$ is of the form $X_1 + iX_2$, with X_1, X_2 each independent and with distribution $G(m, n)$. Let $\tilde{G}(m, n)$ and $\tilde{W}(m, n)$ denote the distributions of $\tilde{X}$ and $\tilde{M}$, respectively. In this case also, the exact joint density function for the m eigenvalues is known [12]:

$$\tilde{K}_{n,m} \exp\left(-\frac{1}{2}\sum_{i=1}^{m}\lambda_i\right)\prod_{i=1}^{m}\lambda_i^{n-m}\prod_{i<j}(\lambda_i-\lambda_j)^2\,d\lambda_1\cdots d\lambda_m, \tag{3}$$

where

$$\tilde{K}_{n,m}^{-1} = 2^{mn}\prod_{i=1}^{m}\Gamma(n-i+1)\Gamma(m-i+1). \tag{4}$$

3. The smallest eigenvalue of $W(n, n)$ and $\tilde{W}(n, n)$. In Theorem 3.1, we show that the probability density function (pdf) for the smallest eigenvalue, $\lambda_{\min}$ of a matrix from $W(n, n)$ is given exactly by a confluent hypergeometric function of a single variable. The exact distributions of the largest and smallest eigenvalues of Wishart matrices are known in certain cases (see [15] and [19]), but these distributions are given as zonal polynomials or hypergeometric functions of matrix arguments that are computationally unwieldy. In contrast, the function described below in Theorem 3.1 is readily calculated numerically by equations 13.1.2 and 13.1.3 in [2].

In Corollary 3.1, we will observe that $n\lambda_{\min}$ converges in distribution to a random variable whose distribution has a simple form. From the limiting distribution, we will analyze the asymptotic behavior of $\log \lambda_{\min}$, which is the key factor in analyzing $E(\log \kappa)$, the expected log condition number.

THEOREM 3.1. *If M_n has the distribution $W(n, n)$, $n \geqq 1$, then the* pdf *of $\lambda_{\min}$ is given by*

$$f_{\lambda_{\min}}(\lambda) = \frac{n}{2^{n-1/2}} \frac{\Gamma(n)}{\Gamma(n/2)} \lambda^{-1/2} e^{-\lambda n/2} U\left(\frac{n-1}{2}, -\frac{1}{2}, \frac{\lambda}{2}\right).$$

When $a > 0$ and $b < 1$, the Tricomi function, $U(a, b, z)$, is the unique solution to Kummer's equation

$$z\frac{d^2w}{dz^2} + (b-z)\frac{dw}{dz} - aw = 0 \tag{5}$$

satisfying $U(a, b, 0) = \Gamma(1-b)/\Gamma(1+a-b)$ and $U(a, b, \infty) = 0$.

Proof. Integrating (1), we obtain

$$f_{\lambda_{\min}}(\lambda) = K_n \lambda^{-1/2} e^{-\lambda/2} \int_{R_\lambda} \exp\left(-\sum_{i=1}^{n-1} \frac{\lambda_i}{2}\right) \prod_{1 \leq i < j \leq n-1} (\lambda_i - \lambda_j) \prod_{i=1}^{n-1} (\lambda_i - \lambda) \lambda_i^{-1/2} \, d\lambda_i,$$

where $R_\lambda = \{\lambda_1 \geqq \cdots \geqq \lambda_{n-1} \geqq \lambda\} \subseteq R^{n-1}$ and $K_n^{-1} = \pi^{-n/2} 2^{n^2/2} \prod_{i=1}^{n} \Gamma(i/2)^2$.

The first trick is the transformation $x_i = \lambda_i - \lambda$,

$$f_{\lambda_{\min}}(\lambda) = \frac{K_n}{(n-1)!} \lambda^{-1/2} e^{-\lambda n/2} \int_{R_+^{n-1}} \prod_{i=1}^{n-1} (x_i + \lambda)^{-1/2} \Delta \, d\mu(x_1) \cdots d\mu(x_{n-1}),$$

where $\Delta = \prod_{1 \leqq i < j \leqq n-1} |x_i - x_j|$, $d\mu(x) = xe^{-x/2}$, and the integration takes place over $R_+^{n-1} = \{(x_1, \cdots, x_{n-1}) : x_i \geqq 0\}$. Let $w(\lambda)$ denote the integral above. Our goal is to show that w satisfies (5).

Let $\Delta = \delta\Delta_2$, where $\delta = \prod_{i=2}^{n-1} |x_1 - x_i|$ and $\Delta_2 = \prod_{2 \leqq i < j \leqq n-1} |x_i - x_j|$. Further, let $f_j^{a,b} = x_j^a (x_j + \lambda)^b$ and $g_j = \prod_{i=j}^{n-1} (x_i + \lambda)^{-1/2}$. Last, let $d\Omega = d\mu(x_1) \cdots d\mu(x_{n-1})$ and $d\Omega_2 = d\mu(x_2) \cdots d\mu(x_{n-1})$. Below we express w, w', and w'' using this notation. All the integrations are over R_+^{n-1}, and symmetry is used when possible.

$$w = \int g_1 \Delta \, d\Omega,$$

$$w' = -\frac{n-1}{2} \int f_1^{0,-3/2} g_2 \Delta \, d\Omega,$$

$$w'' = \frac{(n-1)(n-2)}{4} \int f_1^{0,-3/2} f_2^{0,-3/2} g_3 \Delta \, d\Omega + \frac{3}{4}(n-1) \int f_1^{0,-5/2} g_2 \Delta \, d\Omega.$$

Since $g_1 = (\lambda + x_1) f_1^{0,-3/2} g_2$, we have

$$
\begin{aligned}
w &= \int x_1 f_1^{0,-3/2} g_2 \Delta \, d\Omega + \lambda \int f_1^{0,-3/2} g_2 \Delta \, d\Omega \\
&= -\frac{2\lambda}{n-1} w' + \int f_1^{1,-3/2} g_2 \Delta \, d\Omega \\
&= -\frac{2\lambda}{n-1} w' + \int f_1^{2,-3/2} g_2 e^{-x_1/2} \Delta \, dx_1 \, d\Omega_2 \\
&= -\frac{2\lambda}{n-1} \omega' - 2 \int f_1^{2,-3/2} g_2 \frac{d}{dx_1} \{e^{-x_1/2}\} \Delta \, dx_1 \, d\Omega_2 \\
&= -\frac{2\lambda}{n-1} w' + 2 \int \frac{d}{dx_1} \{ f_1^{2,-3/2} \delta \} e^{-x_1/2} g_2 \Delta_2 \, dx_1 \, d\Omega_2 .
\end{aligned}
$$

The last line is the result of integration by parts. The differentiation gives three terms, so that

$$
\begin{aligned}
w &= -\frac{2\lambda}{n-1} w' + 4 \int f_1^{0,-3/2} g_2 \Delta \, d\Omega - 3 \int f_1^{1,-5/2} g_2 \Delta \, d\Omega \\
&\quad + 2(n-2) \int \frac{x_1}{x_1 - x_2} f_1^{0,-3/2} g_2 \Delta \, d\Omega \qquad (6) \\
&= -\frac{(2\lambda + 8) w'}{n-1} - 3 \int f_1^{1,-5/2} g_2 \Delta \, d\Omega + 2(n-2) \int \frac{x_1}{x_1 - x_2} f_1^{0,-3/2} g_2 \Delta \, d\Omega .
\end{aligned}
$$

Investigating each of the above two integrals, we find

$$
\int f_1^{1,-5/2} g_2 \Delta \, d\Omega = \int f_1^{0,-3/2} g_2 \Delta \, d\Omega - \lambda \int f_1^{0,-5/2} g_2 \Delta \, d\Omega , \qquad (7)
$$

and

$$
\begin{aligned}
\int \frac{x_1}{x_1 - x_2} f_1^{0,-3/2} g_2 \Delta \, d\Omega &= \int \frac{x_1 (x_2 + \lambda)}{x_1 - x_2} f_1^{0,-3/2} f_2^{0,-3/2} g_3 \Delta \, d\Omega \\
&= \lambda \int \frac{x_1}{x_1 - x_2} f_1^{0,-3/2} f_2^{0,-3/2} g_3 \Delta \, d\Omega ,
\end{aligned}
$$

because $x_1 x_2/(x_1 - x_2)$ is antisymmetric. We can use the identity $x_1/(x_1 - x_2) + x_2/(x_2 - x_1) = 1$ and symmetry to integrate this last expression. We obtain

$$
\begin{aligned}
\int \frac{x_1}{x_1 - x_2} f_1^{0,-3/2} g_2 \Delta \Omega &= \lambda \int \frac{x_1}{x_1 - x_2} f_1^{0,-3/2} f_2^{0,-3/2} g_3 \Delta \Omega \qquad (8) \\
&= \frac{\lambda}{2} \int f_1^{0,-3/2} f_2^{0,-3/2} g_3 \Delta \Omega .
\end{aligned}
$$

We substitute (7) and (8) into (6), replacing the integrals with the expressions for w' and w'', and finally rescale $z = \lambda/2$ to obtain (5). Equation (1) gives $w(0) = K_{n+1,n-1}^{-1}(n-1)!$ and clearly $w(\infty) = 0$. The constant term in the pdf is then

$$
\frac{K_n}{K_{n+1,n-1}} \frac{\Gamma(n/2+1)}{\Gamma(3/2)} = \frac{n}{2^{n-1/2}} \frac{\Gamma(n)}{\Gamma(n/2)} ,
$$

and the theorem is proved. This proof was inspired by [3]. See §§ 9 and 10 for further applications of the techniques used.

Though the pdf given in Theorem 3.1 is readily computed, the distribution of $n\lambda_{\min}$ is far simpler as $n \to \infty$.

COROLLARY 3.1. *If M_n has the distribution $W(n, n)$, then as $n \to \infty$, $n\lambda_{\min}$ converges in distribution to a random variable whose* pdf *is given by*

$$f(x) = \frac{1+\sqrt{x}}{2\sqrt{x}} e^{-(x/2+\sqrt{x})}.$$

Proof. From Theorem 3.1, the pdf of $n\lambda_{\min}$ is

$$f_{n\lambda_{\min}}(x) = \frac{n^{1/2}}{2^{n-1/2}} \frac{\Gamma(n)}{\Gamma(n/2)} x^{-1/2} e^{-x/2} U\left(\frac{n-1}{2}, -\frac{1}{2}, \frac{x}{2n}\right).$$

We recall that x_n converges to x in distribution if, for all α, $\lim_{n\to\infty} P(x_n < \alpha) = P(x < \alpha)$. We obtain pointwise convergence of the pdfs on $(0, \infty)$ with the aid of Stirling's formula and the following limiting expression:

$$\lim_{n\to\infty} 2\pi^{-1/2}\Gamma\left(\frac{n+2}{2}\right) U\left(\frac{n-1}{2}, -\frac{1}{2}, \frac{x}{2n}\right) = (1+\sqrt{x})e^{-\sqrt{x}},$$

which is a valid variation of equation 13.3.3 in [2].

In Fig. 3.1, we illustrate the speed of this convergence. We plot the ratio of the pdf of $n\lambda_{\min}$ for $n = 10$ against the function given by Corollary 3.1. We do the same for $n = 50$. Note for $n = 50$ the ratio is nearly 1 throughout the whole interval shown.

COROLLARY 3.2. *If M_n has the distribution $W(n, n)$, then as $n \to \infty$,*

$$E(\log(n\lambda_{\min})) \to -1.68788\cdots.$$

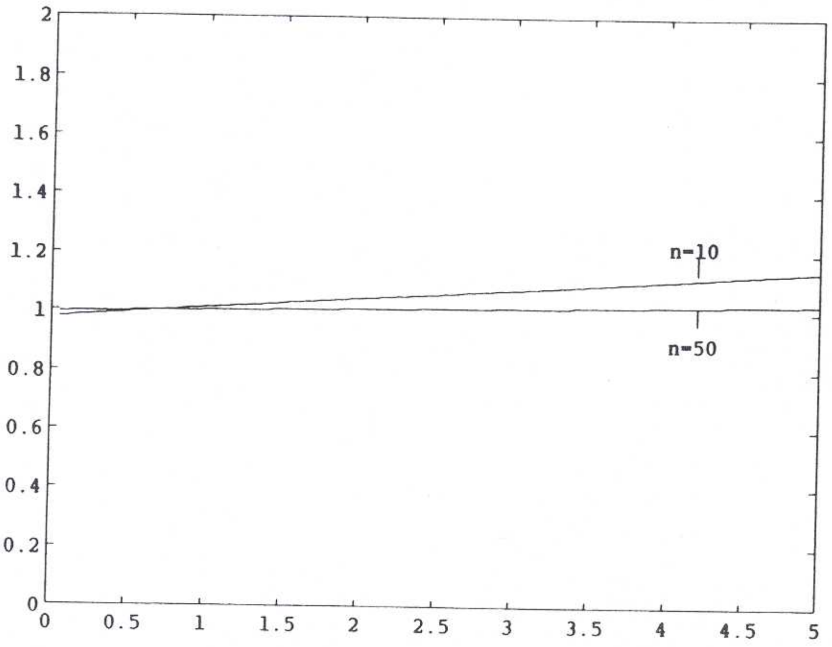

FIG. 3.1. *Speed of convergence of the* pdf *of* $n\lambda_{\min}$.

Proof. In light of the previous corollary and proper convergence of the integrals, the number we seek is

$$\int_0^\infty \log x \frac{1+\sqrt{x}}{2\sqrt{x}} e^{-(x/2+\sqrt{x})}\, dx.$$

This integral can be manipulated into

$$-2\gamma - 2e^{1/2}\int_1^\infty \frac{e^{-1/2y^2}}{y+1}\, dy$$

via a change of variables and equation 4.331.1 in [10], but we know of no simpler form. In this form, however, numerical integration is trivial. $\gamma \approx 0.5772$ in Euler's constant.

We now give the analogous results for complex matrices. The complex case turns out to be simpler.

THEOREM 3.2. *If M_n has the distribution $\tilde{W}(n, n)$, then the* pdf *of $\lambda_{\min}$ is given by*

$$f_{\lambda_{\min}}(\lambda) = \frac{n}{2} e^{-\lambda n/2}.$$

Proof. Let $f_{\lambda_{\min}}(\lambda)$ be the pdf of $\lambda_{\min}$. From (1) we have

$$f_{\lambda_{\min}}(\lambda) = \tilde{K}_{n,n} e^{-\lambda/2} \int_{R_\lambda} \exp\left(-\frac{1}{2}\sum_{i=1}^{n-1} \lambda_i\right) \prod_{i<j} (\lambda_i - \lambda_j)\, d\lambda_1 \cdots d\lambda_{n-1}.$$

By making the transformation $x_i = \lambda_i - \lambda$, we may conclude that $f_{\lambda_{\min}}(\lambda) = ce^{-n\lambda/2}$ for some constant c.

COROLLARY 3.3. *If M_n has the distribution $\tilde{W}(n, n)$, then for all n, $n\lambda_{\min}$ has the distribution χ_2^2.*

Although this corollary immediately follows from the theorem, we might only have guessed it immediately for $n = 1$. This result may be observed experimentally in Fig. 3.2, where we have computed $n\lambda_{\min}$ for 1000 matrices, each 100×100. After sorting these 1000 numbers, let η_i denote the ith value obtained. In Fig. 3.2, we plot η_i versus i/n. This gives the empirical fraction that is less than or equal to η_i. Note that this empirical cumulative density function (cdf) (also known as the empirical distribution function) wiggles around the theoretical cdf plotted as a solid line.

COROLLARY 3.4. *If M_n has the distribution $\tilde{W}(n, n)$, then for all n,*

$$E(\log n\lambda_{\min}) = \log 2 - \gamma = 0.11593\cdots.$$

Proof. We can use equation 4.352 in [10] to compute the appropriate integral.

4. The largest eigenvalue of $W(m, n)$ and $\tilde{W}(m, n)$. In this section we discuss the largest eigenvalue, $\lambda_{\max}$, of $W(n, n)$ and $\tilde{W}(n, n)$, but it requires little extra effort to consider a more general case. Specifically, consider a sequence of Wishart matrices $W(m_n, n)$ or $\tilde{W}(m_n, n)$ such that $m_n/n \to y$ as $n \to \infty$. Loosely speaking, we are looking at large matrices XX^T, where the ratio of number of rows to columns in X is roughly y. Clearly, $y = 1$ covers the cases of $W(n, n)$ and $\tilde{W}(n, n)$.

We start with a known result concerning the convergence in probability of the largest eigenvalues. As a reminder, to say $x_n \xrightarrow{p} x$ means for all $\epsilon > 0$,

$$\lim_{n\to\infty} \Pr(|x - x_n| > \epsilon) = 0.$$

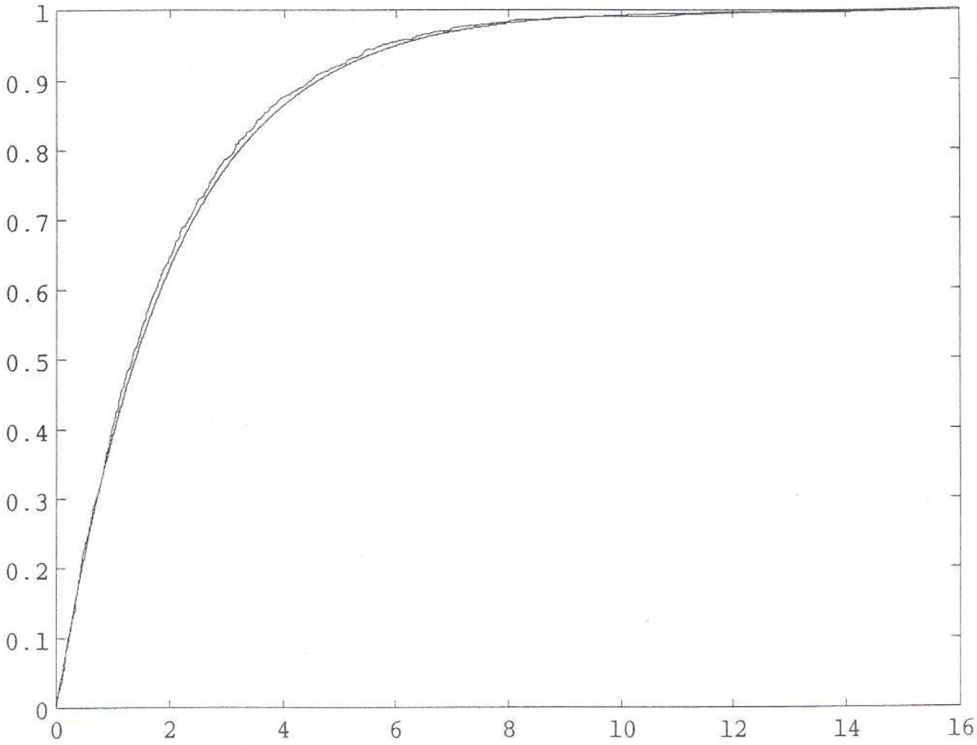

FIG. 3.2. *Theoretical and empirical* cdf *of* $n\lambda_{\min}$ *for* $W(n, n)$.

LEMMA 4.1. *If* M_n *has the distribution* $W(m_n, n)$, *where* $\lim_{n\to\infty} m_n/n = y$, $0 \leqq y < \infty$, *then*

$$(9) \qquad (1/n)\lambda_{\max} \xrightarrow{p} (1+\sqrt{y})^2 \quad \textit{and for } 0 \leqq y \leqq 1, \quad (1/n)\lambda_{\min} \xrightarrow{p} (1-\sqrt{y})^2.$$

Proof. A stronger result (almost sure convergence) can be found in [17].

It is interesting to check Lemma 4.1 experimentally. When we take $y = 1$, the lemma states that, $(1/n)\lambda_{\max}$ converges in probability to 4. With $n = 100$, we computed $\lambda_{\max}/n$ for 1000 matrices. In Fig. 4.1, we plot the empirical cumulative density function (cdf), which is quite close to a step function with step at 4.

We would like to be able to readily conclude from Lemma 4.1 that

$$E(\log \lambda_{\max}/n) \to \log(1+\sqrt{y})^2.$$

It would be that simple if the logarithm were a bounded function; however, since $\log x$ has singularities at zero and infinity, we must carefully investigate the convergence at the singularities. To be precise, we must show that the sequence of random variables $\log \lambda_{\max}/n$ is uniformly integrable [5]. In the following lemma we estimate the pdf.

LEMMA 4.2. *If* M *has the distribution* $W(m, n)$, *then the* pdf $f_{\lambda_{\max}}(x)$ *satisfies*

$$(10) \quad f_{\lambda_{\max}}(x) \leqq \frac{K_{n,m}}{K_{n-1,m-1}} x^{(n+m-3)/2} e^{-x/2} = \frac{\pi^{1/2} 2^{(1-n-m)/2}}{\Gamma(n/2)\Gamma(m/2)} x^{(n+m-3)/2} e^{-x/2}.$$

Proof. This was shown for $m = n$ in [22] by manipulating the expression (1). The same techniques work in the general case.

We can now prove the result that we expect.

PROPOSITION 4.1. *If* M_n *satisfies the hypotheses of Lemma* 4.1 *then* $E(\log \lambda_{\max}) = \log n + \log(1+\sqrt{y})^2 + o(1)$ *as* $n \to \infty$.

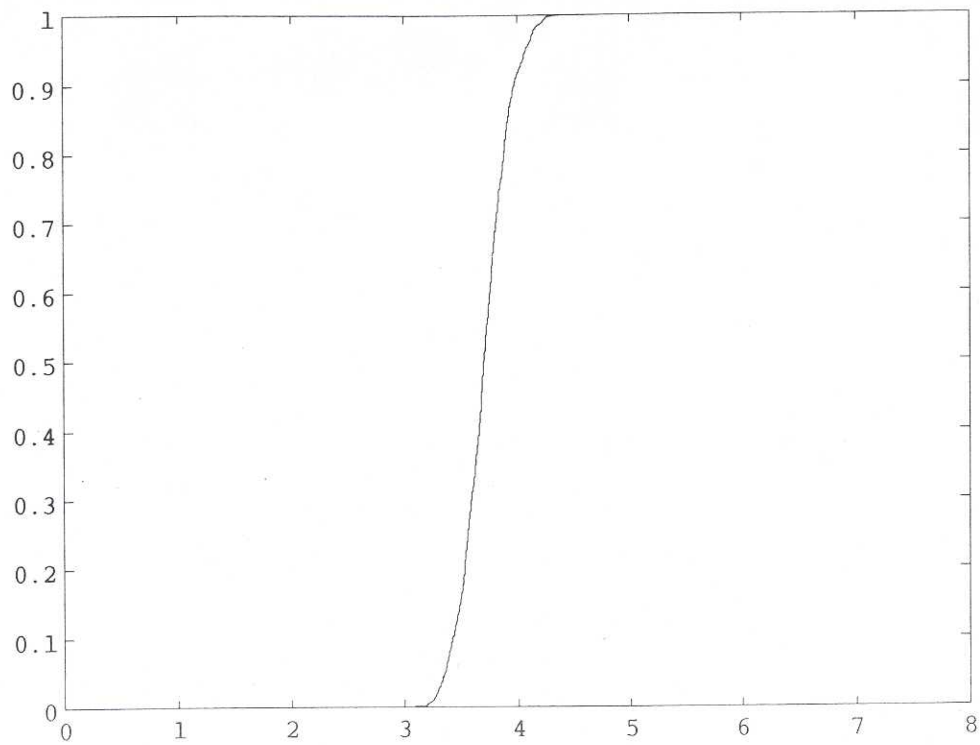

FIG. 4.1. *Empirical* cdf *of* $(1/n\lambda_{\max})$ *for* $W(n, n)$ $(n = 100)$.

Proof. Let σ denote $\lambda_{\max}/n$, and let $f_\sigma(x)$, $F_\sigma(x)$ be the corresponding probability density function and cumulative density function. We break up

$$E(\log \sigma) = \int_0^\infty \log x f_\sigma(x)\, dx$$

into three integrals:

$$\int_0^\varepsilon + \int_\varepsilon^r + \int_r^\infty$$

for values of ε and r depending on y, but not n. By Lemma 4.1, the middle integral approaches $\log\,(1 + \sqrt{y})^2$, and we proceed to show that the other integrals vanish in the limit.

Step 1. $\int_0^\varepsilon$.

We will need a fact that is also of independent interest. We have available another distribution of random matrices whose singular values are distributed exactly as that of $G(m, n)$. We perform a series of Householder transformations to obtain this distribution. (See [17] or [21] for details.) The conclusion is that if X has the distribution $G(m, n)$, then X is orthogonally similar to an $m \times n$ matrix

$$\begin{pmatrix} x_n & & & & 0 \cdots 0 \\ y_{m-1} & x_{n-1} & & & \vdots \quad \vdots \\ & \ddots & \ddots & & \vdots \quad \vdots \\ & & y_1 & x_{n-(m-1)} & 0 \cdots 0 \end{pmatrix}, \tag{11}$$

where x_i^2 and y_i^2 are distributed as χ^2 variables with i degrees of freedom (i.e., χ_i^2). The elements here are all nonnegative and independent.

Let τ be the random variable defined by $(1/n)(x_n^2 + y_{m-1}^2)$. Considering the first column of (11), we have $\|M_n\| = \|X\|^2 = \lambda_{\max} \geqq x_n^2 + y_{m-1}^2$, i.e., $\sigma \geqq \tau$. It follows that $F_\sigma(x) \leqq F_\tau(x)$. Integrating by parts, we obtain

$$0 \geqq \int_0^1 \log x f_\sigma(x)\,dx = -\int_0^1 \frac{F_\sigma(x)}{x}\,dx \geqq -\int_0^1 \frac{F_\tau(x)}{x}\,dx = \int_0^1 \log x f_\tau(x)\,dx.$$

The terms $\log x\, F_\tau(x)$ and $\log x\, F_\sigma(x)$ produced by the integration by parts vanish as $x \to 0$. The former can be verified by using the fact that τ has the distribution $n^{-1}\chi^2_{n+m-1}$, and the latter follows from the former.

To complete the argument we take $m = m_n$, and let $k = n + m_n - 1$, so that τ has the distribution χ_k^2/n, and $f_\tau(x) = ((n/2)^{k/2}/\Gamma(k/2))x^{k/2-1}e^{-nx/2}$. Then,

$$0 \geqq \int_0^\varepsilon \log x f_\tau(x)\,dx \geqq \frac{(n/2)^{k/2}}{\Gamma(k/2)}\int_0^\varepsilon (\log x)x^{k/2-1} \approx \left(\frac{e\varepsilon}{1+y}\right)^{k/2}.$$

Here the $\approx$ indicates that only the exponential behavior is kept as $n \to \infty$. (Computing the asymptotics of this integral is routine but not obvious. A good reference is [4, Chap. 6].) By choosing any $\varepsilon < (1+y)/e$, we have the desired result.

Step 2. $\int_r^\infty$.

For the singularity of the logarithm at ∞ we use Lemma 4.1, the fact that $f_\sigma(x) = n f_{\lambda_{\max}}(nx)$, and a standard asymptotic analysis.

For $r > 1 + y$,

$$\begin{aligned}\int_r^\infty f_\sigma(x)\log x\,dx &\leqq \int_r^\infty x f_\sigma(x)\,dx = \int_{rn/2}^\infty \left(\frac{4x}{n}\right) f_{\lambda_{\max}}(2x)\,dx \\ &\leqq \frac{(2/n)\pi^{1/2}}{\Gamma(n/2)\Gamma(m_n/2)}\int_{rn/2}^\infty x^{(n+m_n-1)/2}e^{-x}\,dx \\ &\approx (e^{-r}(er)^{1+y}y^{-y})^{n/2}.\end{aligned}$$

Here again, $\approx$ indicates that only the exponential behavior is kept as $n \to \infty$. By taking r (depending on y) sufficiently large, we conclude Step 2.

All of these results have analogues for the complex case.

LEMMA 4.3. *If M_n has the distribution $\tilde{W}(m_n, n)$, where $\lim_{n\to\infty} m_n/n = y$, $0 \leqq y < \infty$, then*

$$(12)\qquad (1/n)\lambda_{\max} \xrightarrow{p} 2(1+\sqrt{y})^2 \text{ and for } 0 \leqq y \leqq 1, (1/n)\lambda_{\min} \xrightarrow{p} 2(1-\sqrt{y})^2.$$

PROPOSITION 4.2. *If M_n satisfies the hypotheses of Lemma 4.3, then $E(\log \lambda_{\max}) = \log n + \log 2(1+\sqrt{y})^2 + o(1)$ as $n \to \infty$.*

The proofs are similar and are omitted, but we think it is of interest to mention the analogue of formula (11). If $\tilde{X}$ has the distribution $\tilde{G}(m, n)$, then $\tilde{X}$ is orthogonally similar to an $m \times n$ matrix

$$(13)\qquad \begin{pmatrix} x_{2n} & & & & 0\cdots 0 \\ y_{2(m-1)} & x_{2(n-1)} & & & \vdots \quad \vdots \\ & \ddots & \ddots & & \vdots \quad \vdots \\ & & y_2 & x_{2(n-(m-1))} & 0\cdots 0 \end{pmatrix},$$

where the notation is as in (11). From this we can immediately read that in the square complex case $\det \tilde{M}$ has the distribution $\chi_{2n}^2\chi_{2(n-1)}^2\cdots\chi_2^2$, while in the square real case it is well known (and can be seen from (11)) that $\det M$ has the distribution $\chi_n^2\chi_{n-1}^2\cdots\chi_1^2$.

5. The smallest eigenvalue of $W(m, n)$ and $\tilde{W}(m, n)$.

PROPOSITION 5.1. *If M_n satisfies the hypotheses of Lemma 4.1 and $0 < y < 1$, then $E(\log \lambda_{\min}) = \log n + \log(1 - \sqrt{y})^2 + o(1)$.*

Proof. As in the proof of Proposition 4.1, we must check that $\int_0^\varepsilon \log \lambda f_{\lambda_{\min}}(\lambda)\, d\lambda$ and $\int_r^\infty \log \lambda f_{\lambda_{\min}}(\lambda)\, d\lambda$ vanish as $n \to \infty$. We use the same notation as in the proof of Proposition 4.1 and abbreviate m_n as m:

$$
\begin{aligned}
f_{\lambda_{\min}}(\lambda) & \\
&= K_{n,m}\lambda^{(n-m-1)/2}e^{-\lambda/2}\int_{R_\lambda} \exp\left(-\sum_{i=1}^{m-1}\frac{\lambda_i}{2}\right)\prod_{i<j}(\lambda_i-\lambda_j)\prod_{i=1}^{m-1}(\lambda_i-\lambda)\lambda_i^{(n-m-1)/2}\,d\lambda_i \\
&\leqq K_{n,m}\lambda^{(n-m-1)/2}e^{-\lambda/2}\int_{R_0} \exp\left(-\sum_{i=1}^{m-1}\frac{\lambda_i}{2}\right)\prod_{i<j}(\lambda_i-\lambda_j)\prod_{i=1}^{m-1}\lambda_i^{(n-m+1)/2}\,d\lambda_i \\
&= \frac{K_{n,m}}{K_{n+1,m-1}}\lambda^{(n-m-1)/2}e^{-\lambda/2},
\end{aligned}
$$

and from (2),

$$
\frac{K_{n,m}}{K_{n+1,m-1}} = \pi^{1/2}2^{-(n-m+1)/2}\Gamma\left(\frac{n+1}{2}\right)\Big/\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n-m+1}{2}\right)\Gamma\left(\frac{n-m+2}{2}\right).
$$

Let $\sigma = \lambda_{\min}/n$, so that $f_\sigma(x) = n f_{\lambda_{\min}}(nx)$. For $\varepsilon < 1 - y$,

$$
\begin{aligned}
0 \geqq \int_0^\varepsilon \log x f_\sigma(x)\,dx &\geqq \frac{K_{n,m}}{K_{n+1,m-1}} n^{(n-m+1)/2}\int_0^\varepsilon (\log x)x^{(n-m-1)/2}e^{-nx/2}\,dx \\
&\approx \left(\left(\frac{e\varepsilon}{n(1-y)^2}\right)^{1-y} e^{-\varepsilon}\right)^{n/2}.
\end{aligned}
$$

On the other hand, as in the proof of Proposition 4.1, $\sigma \leqq \tau$, which has the distribution χ^2_{n+m-1}. It then follows that $F_\sigma(x) \geqq F_\tau(x)$. For $r > 1$,

$$
\begin{aligned}
\int_r^\infty \log x f_\sigma(x)\,dx &= \log x(F_\sigma(x) - 1)\big|_r^\infty + \int_r^\infty \frac{1 - F_\sigma(x)}{x} \\
&\leqq \log x(F_\sigma(x) - 1)\big|_r^\infty - \log x(F_\tau(x) - 1)\big|_r^\infty + \int_r^\infty \log x f_\tau(x)\,dx.
\end{aligned}
$$

The same kind of asymptotic analysis as above shows that as $n \to \infty$, each of the terms vanishes.

Of course, we have the complex result as well.

PROPOSITION 5.2. *If M_n satisfies the hypotheses of Lemma 4.3 and $0 < y < 1$, then $E(\log \lambda_{\min}) = \log n + \log 2(1 - \sqrt{y})^2 + o(1)$.*

6. Limiting condition number distributions and expected logarithms. We can now combine all the results of the previous section to describe the condition number distributions and the expected logarithms.

THEOREM 6.1. *If κ_n is the condition number of a matrix from the distribution $G(n, n)$, then κ_n/n converges in distribution to a random variable whose* pdf *is given by*

$$
f(x) = \frac{2x+4}{x^3}e^{-2/x-2/x^2}.
$$

Moreover,

$$E(\log \kappa_n) = \log n + c + o(1) \approx \log n + 1.537$$

as $n \to \infty$.

Proof. From Lemma 4.1, we know $(1/n)\lambda_{\max} \xrightarrow{p} 4$ and Corollary 3.1 gives the limiting distribution for $n\lambda_{\min}$. The ratio of these quantities, κ_n^2/n^2, converges in distribution by a standard probability argument. The appropriate change of variables gives the limiting pdf of κ_n/n. The expected logarithm follows from Corollary 3.2 and Proposition 4.1.

THEOREM 6.2. *If* κ_n *is the condition number of a matrix from the distribution* $\tilde{G}(n, n)$, *then* κ_n/n *converges in distribution to a random variable whose* pdf *is given by*

$$f(x) = \frac{8}{x^3} e^{-4/x^2}.$$

Moreover,

$$E(\log \kappa_n) = \log n + \tfrac{1}{2}\gamma + \log 2 + o(1) \approx \log n + 0.982$$

as $n \to \infty$.

Proof. As in the proof of Theorem 6.1, the pdf follows from Lemma 4.3 and Corollary 3.3, and the expected logarithm follows from Corollary 3.4 and Proposition 4.2.

THEOREM 6.3. *If* κ_n *is the condition number of a matrix from the distribution* $G(m_n, n)$ *or* $\tilde{G}(m_n, n)$, *where* $\lim_{n\to\infty} m_n/n = y$ *and* $0 < y < 1$, *then* κ_n *converges in probability to* $(1 + \sqrt{y})/(1 - \sqrt{y})$. *Moreover,*

$$E(\log \kappa_n) = \log \frac{1+\sqrt{y}}{1-\sqrt{y}} + o(1)$$

as $n \to \infty$.

The convergence follows trivially from Lemma 4.1 and Lemma 4.2 and, of course, the statement could be strengthened to almost sure convergence. The expected logarithm follows from Propositions 4.1, 4.2, 5.1, and 5.2.

7. Exact expressions for $m = 2$. It is possible to integrate expressions (1) and (3) against the condition number to get the exact distributions of the condition numbers of real and complex $2 \times n$ matrices. We spare the reader the details and give only the results.

The pdf of the condition number of matrices that have the distribution $G(2, n)$ is given by

$$f_\kappa(x) = (n-1)2^{n-1} \frac{x^2 - 1}{(x^2+1)^n} x^{n-2}. \tag{14}$$

Similarly, when the matrices have the distribution $\tilde{G}(2, n)$, we have

$$f_{\tilde{\kappa}}(x) = 2 \frac{\Gamma(2n)}{\Gamma(n)\Gamma(n-1)} \frac{x^{2n-3}(x^2-1)^2}{(x^2+1)^{2n}}. \tag{15}$$

We can use (14) and (15) to evaluate the integrals giving the expected condition numbers, and the result is the following theorem.

THEOREM 7.1. *If* X_n *has the distribution* $G(2, n)$, *then*

$$E(\log \kappa_{X_n}) = \frac{1}{2}\sqrt{\pi}\,\Gamma\left(\frac{n-1}{2}\right) \Big/ \Gamma\left(\frac{n}{2}\right).$$

If $\tilde{X}_n$ has the distribution $\tilde{G}(2, n)$, then

$$E(\log \kappa_{\tilde{X}_n}) = \log 2 + \frac{1}{2} - \sum_{k=2}^{n-1} \frac{1}{4^k}\binom{2k}{k}\frac{1}{k-1}.$$

We can also obtain the exact distribution for the smaller and the larger eigenvalues:

THEOREM 7.2. *If M_n has the distribution $W(2, n)$ and β denotes $(n-1)/2$, then*

$$f_{\lambda_{\min}}(\lambda) = K_{n,2}e^{-\lambda}(2\lambda^\beta e^{-\lambda/2} + 2^\beta(2\beta - \lambda)\Gamma(\beta, \lambda/2))$$

and

$$f_{\lambda_{\max}}(\lambda) = K_{n,2}e^{-\lambda/2}\lambda^{\beta-1}(2\lambda^\beta e^{-\lambda/2} - 2^\beta(2\beta - \lambda)\gamma(\beta, \lambda/2)).$$

A similar result for $\tilde{W}(2, n)$ could be calculated.

8. The tails of the condition number distributions. In the previous sections, we described the behavior of the condition numbers but said nothing about the probability that a matrix with a large condition number may appear. Here we will approximate the condition numbers for square matrices in order to get a sense of the tails of the distributions.

There are four condition numbers that we find interesting. Let κ and $\tilde{\kappa}$ denote the random variables, which are the 2-norm condition number of a matrix having the distribution $G(n, n)$ and $\tilde{G}(n, n)$, respectively. Since we are only considering $n \times n$ matrices, we omit the dependence on n in the notation. The other two condition numbers were introduced by Demmel [8]. Let $\|X\|_F$ denote the Frobenius norm of X, defined as $\sqrt{\sum_{i,j} X_{ij}^2} = \sqrt{\text{trace}\,(XX^T)}$. Demmel's condition number is defined by $\|X\|_F\|X^{-1}\|_2$. Let κ_D and $\tilde{\kappa}_D$ denote the random variables that are the Demmel condition number in the real and complex cases as above. We chart the condition numbers and relate them to the eigenvalues of the corresponding Wishart matrix in the table below.

$\kappa = \sqrt{\lambda_{\max}/\lambda_{\min}}$	$\tilde{\kappa} = \sqrt{\lambda_{\max}/\lambda_{\min}}$
$\kappa_D = \sqrt{\sum \lambda_i/\lambda_{\min}}$	$\tilde{\kappa}_D = \sqrt{\sum \lambda_i/\lambda_{\min}}$

In the tables that follow, we consistently use the above ordering: real versus complex in the columns, and 2-norm versus Demmel's norm in the rows.

The numbers in the table below are the values that the indicated expressions converge to in probability as $n \to \infty$.

	$W(n,n)$	$\tilde{W}(n,n)$
$\frac{1}{n}\lambda_{\max}$	4	8
$\frac{1}{n^2}\sum_{i=1}^n \lambda_i$	1	2

The first row is Lemmas 4.1 and 4.3. The second row is derived from the law of large numbers and the observation that the trace of a Wishart matrix has the $\chi^2_{n^2}$ distribution in the real case and the $\chi^2_{2n^2}$ in the complex case. Replacing these convergence results with equality, we define four approximate condition numbers:

$\kappa' = \sqrt{4n/\lambda_{\min}}$	$\tilde{\kappa}' = \sqrt{8n/\lambda_{\min}}$
$\kappa'_D = \sqrt{n^2/\lambda_{\min}}$	$\tilde{\kappa}'_D = \sqrt{2n^2/\lambda_{\min}}$

Directly from the definition of these condition numbers we have the following justification of our approximation.

LEMMA 8.1. *As* $n \to \infty$, κ/κ', κ_D/κ'_D, $\tilde{\kappa}/\tilde{\kappa}'$, *and* $\tilde{\kappa}_D/\tilde{\kappa}'_D$ *all converge in probability to* 1.

The approximate condition numbers only depend on $\lambda_{\min}$. Thus it becomes necessary to investigate the probability that $\lambda_{\min}$ is small.

LEMMA 8.2. *As* $\lambda \to 0$, $P(\lambda_{\min} < \lambda) \sim \sqrt{\lambda n}$ *if M has the distribution* $W(n, n)$ *and* $P(\lambda_{\min} < \lambda) \sim \lambda n/2$ *if M has the distribution* $\tilde{W}(n, n)$.

Proof. The real result comes from analyzing the formula given in Theorem 3.1. The complex result is trivial since $n\lambda_{\min}$ has the distribution χ_2^2 according to Corollary 3.3.

THEOREM 8.1. *As* $x \to \infty$,

$P(\kappa' > x) \sim 2n/x$	$P(\tilde{\kappa}' > x) \sim 4n^2/x^2$
$P(\kappa'_D > x) \sim n^{3/2}/x$	$P(\tilde{\kappa}'_D > x) \sim n^3/x^2$

Proof. Combine the small λ behavior described in Lemma 8.2 with the definitions of our condition numbers. The results follow from the obvious change of variables.

In one case we can compare our results with those known for the exact condition number. Demmel showed that for all n, $P(\tilde{\kappa}_D > x) \sim (n^3 - n)/x^2$ as $x \to \infty$, while we have $P(\tilde{\kappa}'_D > x) \sim n^3/x^2$ as $x \to \infty$. The difference is negligible for all but very small n.

9. All the eigenvalues of a Wishart matrix. We would like to describe the complete spectrum of a Wishart matrix. The m eigenvalues of a matrix from $W(m, n)$ and $\tilde{W}(m, n)$ are, of course, random, but what can we say about them? We have already mentioned their joint density function in (1) and (3), but this does not give much insight into the total picture. Here, we contrast three descriptions of the complete set of eigenvalues. The first two are well known and the third is, we believe, new.

(1) *Mode*. The m-tuple $(\lambda_1, \cdots, \lambda_m)$ that maximizes (1) or (3) (when there is a maximum) consists of the roots of the Laguerre polynomial

$$L_m^{(2(\alpha/\beta)-1)}(x/\beta),$$

where $\alpha = \frac{1}{2}(n - m - 1)$ and $\beta = 1$ in the real case, while $\alpha = n - m$ and $\beta = 2$ in the complex case.

(2) *Empirical distribution function*. Take a large Wishart matrix and plot the $(\lambda_i, i/n)$. The picture will be a curve the limiting form of which is well known and listed for reference in Propositions 9.1 and 9.2.

(3) *Expected characteristic polynomial*. The expected characteristic polynomials of Wishart matrices can be computed precisely. They are

$$(-\beta)^m m! L_m^{(n-m)}(t/\beta);$$

$\beta = 1$ in the real case and 2 in the complex case.

We now discuss these ideas in detail.

9.1. Mode. The mode is related to an electrostatic interpretation of the zeros of the classical polynomials given in [20]. Note that there is an infinite density in the real case when $m = n$ and $\lambda_m = 0$, so the formula does not apply.

9.2. Empirical distribution function. The empirical distribution function $W_M(x)$ of a matrix M is the fraction of eigenvalues of M that are less than or equal to x. One

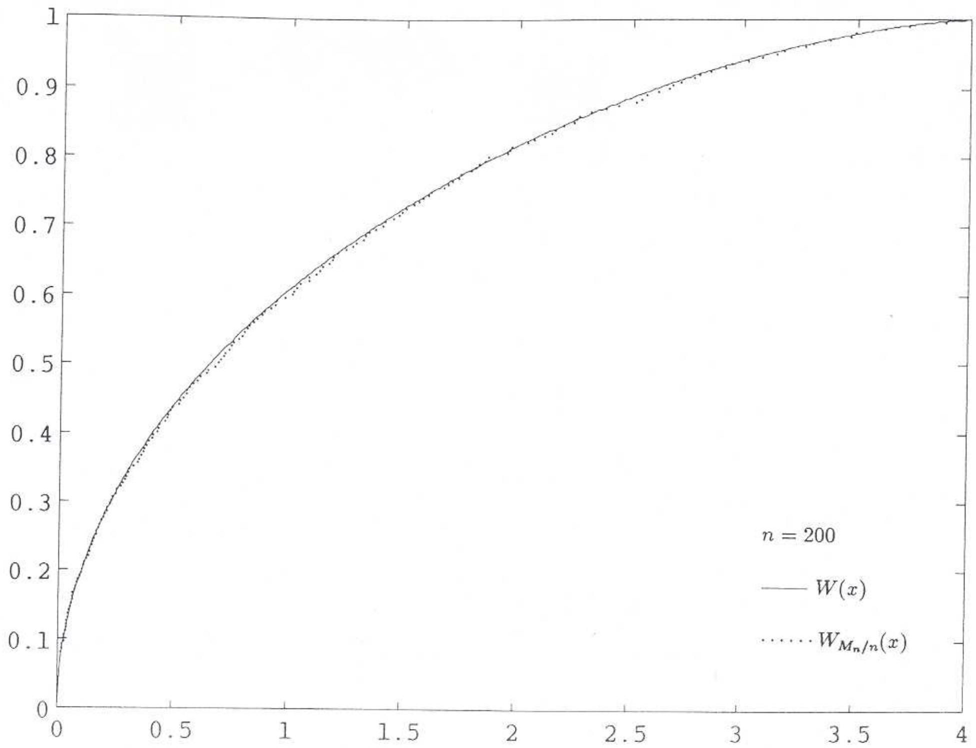

FIG. 9.1. *Empirical* cdf *of the eigenvalues of* $W(n, n)$.

way to view this is that if the eigenvalues are thought of as being chosen from a random sample, $W_M(x)$ is its empirical cdf. Computationally, we simply sort the eigenvalues and plot λ_i against i/n. We do this for a matrix M_n/n, where M_n was generated from the distribution $W(200, 200)$ and plot $W_{M_n}(x)$ in Figure 9.1 as a dotted line. It is well known that $W_{M_n}(x)$ converges almost surely to a limiting function as $n \to \infty$. $W(x)$ is plotted in Fig. 9.1 as a solid line.

PROPOSITION 9.1. *If M_n satisfies the conditions of Lemma* 4.1, *then $W_{M_n/n}(x)$ converges almost surely to a fixed function $W(x)$ as $n \to \infty$. If $y = 1$, this function satisfies $W'(x) = (1/2\pi)((4 - x)/x)^{1/2}$ for $0 \leqq x \leqq 4$. More generally, for $0 < y \leqq 1$, we have almost sure convergence to a fixed function satisfying*

$$W'(x) = \frac{\sqrt{(x - a(y))(b(y) - x)}}{2\pi yx}$$

for $a(y) < x < b(y)$, where

$$a(y) = (\sqrt{y} - 1)^2 \quad \text{and} \quad b(y) = (\sqrt{y} + 1)^2.$$

For $y > 1$ the above result is modified by adding $(1 - 1/y)\delta(x)$ to $W'(x)$.

Proof. This proposition and the one to follow was proved in [23] in a very general setting. Convergence in probability was proved earlier in [16]. Other more recent proofs can be found in [13] and [21]. These last two proofs are not as general but are quite elegant.

PROPOSITION 9.2. *If $\tilde{M}_n$ has the distribution $\tilde{W}(m_n, n)$, where $\lim_{n \to \infty} m_n/n = y$ and $0 \leqq y < \infty$, then $W_{\tilde{M}/n}(x)$ converges almost surely to a fixed function $\tilde{W}(x)$ as $n \to \infty$. If $y = 1$, this function satisfies $\tilde{W}'(x) = (1/4\pi)((8 - x)/x)^{1/2}$ for $0 \leqq x \leqq 8$. More generally, $\tilde{W}(2x) = W(x)$, as defined in Proposition* 9.1.

The source of the extra factor of 2 is simple. It is merely the variance of the elements of the matrices that are 1 in the real case but 2 in the complex case.

9.3. Characteristic polynomial. We can derive exactly the expected characteristic polynomial of a Wishart matrix. This could be thought of as the average of all the coefficients (which are of course symmetric functions of the eigenvalues) or as the average value of the characteristic polynomial at a given point. This is of interest here because the roots of the average polynomial deserve to be thought of as "typical" values for the eigenvalue.

Computing the expected characteristic polynomial is a special case of a multivariate integration of the form

$$\int_S f(\lambda_1, \cdots, \lambda_m)\Delta^k \, d\mu_1 \cdots d\mu_m,$$

where $d\mu_i = e^{-(1/2)\lambda_i}\lambda_i^\alpha d\lambda_i$, $\Delta = \prod_{i<j}(\lambda_i - \lambda_j)$, and the region of integration S is defined by $\lambda_1 \geqq \cdots \geqq \lambda_m \geqq 0$. Any expected value calculations involving the eigenvalues of Wishart matrices has exactly this form with $k = 1$ in the real case and $k = 2$ in the complex case. (See § 2.)

To compute the expected characteristic polynomial, take $f(\lambda_1, \cdots, \lambda_m) = \prod_{i=1}^p (t - \lambda_j)$, where t may be thought of as a variable. We make use of a recent result due to Aomoto [3].

LEMMA 9.1. *Let*

$$I_f = \int_{S_1} \prod_{i=1}^m (t - \lambda_i)\Delta^k \, d\nu_1 \cdots d\nu_m, \tag{16}$$

where $d\nu_i = \lambda_i^\alpha (1-\lambda_i)^\beta d\lambda_i$, *and the region of integration*, S_1 *is defined by* $1 \geqq \lambda_1 \geqq \cdots \geqq \lambda_m \geqq 0$. *Then*

$$\frac{I_f}{I_1} = \binom{\alpha' + \beta' + 2n}{n}^{-1} P_m^{(\alpha',\beta')}(1-2t), \tag{17}$$

where $P_m^{(\alpha',\beta')}$ *denotes a Jacobi polynomial*, $\alpha' = -1 + 2(\alpha+1)k$, $\beta' = -1 + 2(\beta+1)/k$ *and* $I_1 = \int_{S_1} \Delta^k \, d\nu_1 \cdots d\nu_m$.

This lemma is proved in [3]. The value of I_1 was first computed by Selberg in 1944, but his original paper is unavailable in many libraries. His results and argument, however, can be found in § 5.4 of [1]. We have derived an alternative proof to this lemma and to Lemma 9.2 by proving that the integrals satisfy the correct second-order differential equation for the Jacobi and Laguerre polynomials. This proof closely resembles the proof of Theorem 3.1.

LEMMA 9.2.

$$\int_S \prod_{i=1}^m (t - \lambda_i)\Delta^k \, d\mu_1 \cdots d\mu_m = c_{\alpha,m}^{(k)} L_m^{(\alpha')}\left(\frac{t}{k}\right), \tag{18}$$

where $L_m^{(\alpha')}$ *denotes a Laguerre polynomial*, $(c_{\alpha,m}^{(1)})^{-1} = (-1)^m \binom{m+\alpha}{m} K_{2\alpha+m+3,m}$ *and* $(c_{\alpha,m}^{(2)})^{-1} = (-1)^m \binom{m+\alpha}{m} \tilde{K}_{\alpha+m+1,m}$.

Proof. In (16) make the substitutions $\lambda_i \to \lambda_i/2\beta$ and $t \to t/2\beta$. The value of (17) becomes a multiple of

$$P_m^{(\alpha',\beta')}\left(1 - \frac{2t}{2\beta}\right) = P_m^{(\alpha',\beta')}\left(1 - \frac{2t}{k(\beta'+1)-2}\right).$$

To compute (18), let $\beta' \to \infty$. Using standard formulas about orthogonal polynomials, we can verify

$$\lim_{\beta' \to \infty} P_m^{(\alpha', \beta')}\left(1 - \frac{2t}{k(\beta'+1)-2}\right) = L_m^{(\alpha')}\left(\frac{t}{k}\right).$$

We get the constants $c_{\alpha,m}^{(k)}$ by setting $t = 0$ in (18). The right-hand side is $c_{\alpha,m}^{(k)}\binom{m+\alpha}{m}$. The left-hand integral is an integral of the expressions (1) and (3) up to a constant. Since (1) and (3) are joint density functions they integrate to 1. For a suitable choice of n we get the values (2) and (4). (Note that we computed the constant for $k = 1$ or 2 since we had (2) and (4) handy. We could have obtained $c_{\alpha,m}^{(k)}$ from scratch for all k, by evaluating the integral (18) when $t = 0$ by using a limiting process on the value of Selberg's integral.)

THEOREM 9.1. *Let* $P_M(t) = \det(tI - M)$ *be the characteristic polynomial of* M. *Then* $E(P_M(t)) = (-1)^m m! L_m^{(n-m)}(t)$ *if* M *has the distribution* $W(m, n)$ *and* $E(P_M(t)) = (-2)^m m! L_m^{(n-m)}(t/2)$ *if* M *has the distribution* $\tilde{W}(m, n)$.

Proof. Each of the expected values we are computing here has the form (18). In the real case $k = 1$ and $\alpha = (n - m - 1)/2$, so $\alpha' = n - m$. In the complex case, $k = 2$ and $\alpha = n - m$, so again $\alpha' = n - m$. The easy way to check that the constant is correct is to compare the highest coefficient of t, which is unity on both sides.

10. The probability density function of $\lambda_{\min}$ for $W(m, m+3)$. The smallest eigenvalue of a matrix from $W(m, m+1)$ behaves exactly like the one in $\tilde{W}(m, m)$, that is, $m\lambda_{\min}$ has the χ_2^2 distribution. The proof is similar to that of Theorem 3.2.

In fact, the pdf of $\lambda_{\min}$ for any matrix from $W(m, n)$ for $n - m$ odd or any matrix from $\tilde{W}(m, n)$ is given by

$$e^{-\lambda m/2} P(\lambda),$$

where P is a polynomial. This was pointed out in the real case in [15] and in fact can be seen directly from the integral.

To illustrate another application of Lemma 9.2, we derive the polynomial for the special case of $W(m, m+3)$. A similar result is given in [15], where the distribution is expressed as a hypergeometric function of a matrix argument. The two results are in fact equivalent, but we give a more explicit expression.

THEOREM 10.1. *If* M *has the distribution* $W(m, m+3)$, *then*

$$f_{\lambda_{\min}}(\lambda) = \frac{1}{2(m+1)} e^{-\lambda m/2} \lambda L_{m-1}^{(3)}(-\lambda).$$

Proof. From (1) we know that

$$f_{\lambda_{\min}}(\lambda) = K_{n,m} e^{-\lambda/2} \lambda \int_{S'} \prod_{i=1}^{m-1} (\lambda_i - \lambda) \Delta \prod_{i=1}^{m-1} \lambda_i \, d\lambda_1 \cdots d\lambda_{m-1},$$

where S' is defined by $\lambda_1 \geqq \lambda_2 \geqq \cdots \geqq \lambda_{m-1} \geqq 0$. Letting $\lambda_i \to \lambda_i - \lambda$, we obtain

$$f_{\lambda_{\min}}(\lambda) = K_{n,m} e^{-\lambda m/2} \lambda \int_{S} \prod_{i=1}^{m-1} (\lambda_i + \lambda) \Delta \, d\mu_1 \cdots d\mu_{m-1}.$$

Here the notation is as in the previous section and $\alpha = 1$, so that $\alpha' = 3$. The conclusion follows from Lemma 9.2.

Acknowledgments. My sincere thanks go to Nick Trefethen, whose door was always open when I needed to describe an idea. His insightful comments have meant a great deal. I also wish to acknowledge the creators of an excellent software package, MATLAB, which made running experiments as easy as thinking of them.

REFERENCES

[1] G. E. ANDREWS, *q-Series: Their Development and Application in Analysis, Number Theory, Combinatorics, Physics, and Computer Algebra*, Regional Conference Series in Mathematics 66, American Mathematical Society, Providence, RI, 1986.

[2] M. ABRAMOWITZ AND I. A. STEGUN, EDS., *Handbook of Mathematical Functions*, Dover, New York, 1970.

[3] K. AOMOTO, *Jacobi polynomials associated with Selberg integrals*, SIAM J. Math. Anal., 18 (1987), 545–549.

[4] C. M. BENDER AND S. A. ORSZAG, *Advanced Mathematical Methods for Scientists and Engineers*, McGraw-Hill, New York, 1978.

[5] P. BILLINGSLEY, *Probability and Measure*, John Wiley, New York, 1979.

[6] L. BLUM AND M. SHUB, *Evaluating rational functions: infinite precision is finite cost and tractable on average*, SIAM J. Comput., 15 (1986), pp. 384–398.

[7] J. E. COHEN, H. KESTEN, AND C. M. NEWMAN, EDS., *Random Matrices and Their Applications*, Contemporary Mathematics, Vol. 50, American Mathematical Society, Providence, RI, 1986.

[8] J. DEMMEL, *The probability that a numerical analysis problem is difficult*, Math. Comp., 50 (1988), pp. 449–480.

[9] S. GEMAN, *A limit theorem for the norm of random matrices*, Ann. Probab., 8 (1980), pp. 252–261.

[10] I. S. GRADSHTEYN AND I. W. RYZHIK, *Table of Integrals, Series, and Products*, Fourth edition, Academic Press, New York, 1965.

[11] R. D. GUPTA AND D. S. P. RICHARDS, *Hypergeometric functions of scalar matrix argument are expressible in terms of classical hypergeometric functions*, SIAM J. Math. Anal., 16 (1985), pp. 852–858.

[12] A. T. JAMES, *Distributions of matrix variates and latent roots derived from normal samples*, Ann. Math. Statist., 35 (1964), pp. 475–501.

[13] D. JONSSON, *Some limit theorems for the eigenvalues of a sample covariance matrix*, J. Multivariate Anal., 12 (1982), pp. 1–38.

[14] E. KOSTLAN, *Numerical linear algebra and multivariant analysis*, in preparation.

[15] P. R. KRISHNAIAH AND T. C. CHENG, *On the exact distribution of the smallest roots of the Wishart Matrix using zonal polynomials*, Ann. Inst. Statist. Math., 23 (1971), pp. 293–295.

[16] V. A. MARCENKO AND L. A. PASTUR, *Distributions of eigenvalues for some sets of random matrices*, Math. USSR-Sb., 1 (1967), pp. 457–483.

[17] J. W. SILVERSTEIN, *The smallest eigenvalue of a large-dimensional Wishart matrix*, Ann. Probab., 13 (1985), pp. 1364–1368.

[18] S. SMALE, *On the efficiency of algorithms of analysis*, Bull. Amer. Math Soc., 13 (1985), pp. 87–121.

[19] T. SUGIYAMA, *On the distribution of the largest latent root of the covariance matrix*, Ann. Math. Statist., 38 (1967), pp. 1148–1151.

[20] G. SZEGÖ, *Orthogonal Polynomials*, American Mathematical Society, Providence, RI, 1939.

[21] H. F. TROTTER, *Eigenvalue distributions of large hermitian matrices; Wigner's semi-circle law and a theorem of Kac, Murdock, and Szegö*, Adv. in Math., 54 (1984), pp. 67–82.

[22] J. VON NEUMANN AND H. H. GOLDSTINE, *Numerical inverting of matrices of high order*, in John von Neumann, Collected Works, Vol. 5: Design of Computers, Theory of Automata and Numerical Analysis, A. H. Taub, ed., Pergamon, New York, 1963.

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

[24] N. WEISS, G. W. WASILKOWSKI, H. WOŹNIAKOWSKI, AND M. SHUB, *Average condition number for solving linear equations*, Linear Algebra Appl., 83 (1986), pp. 79–102.

[25] S. WILKS, *Mathematical Statistics*, John Wiley, New York, 1967.