SIAM J. MATRIX ANAL. APPL.
Vol. 26, No. 2, pp. 426–440

UPPER AND LOWER BOUNDS FOR THE TAILS OF THE DISTRIBUTION OF THE CONDITION NUMBER OF A GAUSSIAN MATRIX*

JEAN-MARC AZAÏS† AND MARIO WSCHEBOR‡

Abstract. Let A be an $m\times m$ real random matrix with independently and identically distributed standard Gaussian entries. We prove that there exist universal positive constants c and C such that the tail of the probability distribution of the condition number $\kappa(A)$ satisfies the inequalities $\frac{c}{x} < \mathrm{P}\{\kappa(A) > mx\} < \frac{C}{x}$ for every $x > 1$. The proof requires a new estimation of the joint density of the largest and the smallest eigenvalues of A^TA which follows from a formula for the expectation of the number of zeros of a certain random field defined on a smooth manifold.

Key words. random matrices, condition number, eigenvalue distribution, Rice formulae

AMS subject classifications. 15A12, 60G15

DOI. 10.1137/S0895479803429764

1. Introduction and main result.

Let A be an $m \times m$ real matrix and denote by

$$\|A\| = \sup_{\|x\|=1} \|Ax\|$$

its Euclidean operator norm. $\|x\|$ denotes the Euclidean norm of x in $\mathbb{R}^m$. If A is nonsingular, its *condition number* $\kappa(A)$ is defined by

$$\kappa(A) = \|A\|\|A^{-1}\|$$

(von Neumann and Goldstine [18]; Turing [17]). The role of $\kappa(A)$ in a variety of numerical analysis problems is well established (see, for example, Wilkinson [20], Smale [14], Higham [9], and Demmel [6]). The purpose of the present paper is to prove the following.

THEOREM 1.1. *Assume that $A = ((a_{ij}))_{i,j=1,\dots,m}$, $m \geq 3$, and that the a_{ij}'s are independently and identically distributed (i.i.d.) Gaussian standard random variables.*

Then there exist universal positive constants c, C such that for $x > 1$,

$$\frac{c}{x} < \mathrm{P}\{\kappa(A) > mx\} < \frac{C}{x}. \tag{1.1}$$

Remarks. The following are remarks on the statement of Theorem 1.1:

1. It is well known that as m tends to infinity, the distribution of the random variable $\kappa(A)/m$ converges to a certain distribution (this follows easily, for example, from Edelman [7]). The interest of (1.1) lies in the fact that it holds true for all $m \geq 3$ and $x > 1$

*Received by the editors June 10, 2003; accepted for publication (in revised form) April 26, 2004; published electronically January 12, 2005. This work was supported by ECOS program U03E01.
http://www.siam.org/journals/simax/26-2/42976.html

†Laboratoire de Statistique et Probabilités, UMR-CNRS C55830, Université Paul Sabatier, 118, route de Narbonne, 31062 Toulouse Cedex 4, France (azais@cict.fr).

‡Centro de Matemática, Facultad de Ciencias, Universidad de la República, Calle Igua 4225, 11400 Montevideo, Uruguay (wschebor@cmat.edu.uy).

2. We will see below that $c = 0.13,\ C = 5.60$ satisfy (1.1) for every $m = 3, 4, \dots$ and $x > 1$. Using the same methods, one can obtain more precise upper and lower bounds for each m, but we will not detail these calculations here.
   The simulation study of section 3 suggests that $\mathrm{P}\{\kappa(A) > mx\}$ is increasing with m, so that c should be the value corresponding to $m = 3$, i.e., $c \approx 0.88$, and C the one derived from the mentioned asymptotic result in Edelman [7], i.e., $C = 2$.
3. In Sankar, Spielman, and Teng [13] it was conjectured that

$$\mathrm{P}\{\kappa(A) > x\} = O\left(\frac{m}{\sigma x}\right)$$

   when the a_{ij}'s are independent Gaussian random variables having a common variance $\sigma^2 \le 1$ and $\sup_{i,j} |\mathrm{E}\,(a_{ij})| \le 1$.
   The upper bound part of (1.1) implies that this conjecture holds true in the centered case. The lower bound shows that, up to a constant factor, this is the exact order of the behavior of the tail of the probability distribution of $\kappa(A)$. See Wschebor [22] for the noncentered case.
4. This theorem, and related ones, can be considered as results on the Wishart matrix $A^T A$ (A^T denotes the transpose of A). Introducing some minor changes, it is possible to use the same methods to study the condition number of $A^T A$ for rectangular $n \times m$ matrices A having i.i.d. Gaussian standard entries, $n > m$. This will be considered elsewhere.

Some examples of related results on the random variable $\kappa(A)$ are the following.

THEOREM 1.2 (see [7]). *Under the same hypothesis as that of Theorem* 1.1, *one has*

$$E\,(\log \kappa(A)) = \log m + C_1 + \varepsilon_m,$$

where C_1 is a known constant ($C_1 \approx 1.537$) and $\varepsilon_m \to 0$ as $m \to +\infty$.

THEOREM 1.3 (see [4]). *Let $A = ((a_{ij}))_{i,j=1,\dots,m}$ and assume that the a_{ij}'s are independent Gaussian random variables with a common variance σ^2 and $m_{ij} = \mathrm{E}(a_{ij})$. Denote by $M = ((m_{ij}))_{i,j=1,\dots,m}$ the nonrandom matrix of expectations. Then*

$$\mathrm{E}\,(\log \kappa(A)) \le \log m + \log\left[\frac{\|M\|}{\sigma\sqrt{m}} + 4\right] + C_1',$$

where C_1' is a known constant.

Next, we introduce some notation. Given A, an $m \times m$ real matrix, we denote by $\lambda_1, \dots, \lambda_m$, $0 \le \lambda_1 \le \dots \le \lambda_m$, the eigenvalues of $A^T A$. If $X : S^{m-1} \to \mathbb{R}$ is the quadratic polynomial $X(x) = x^T A^T A x$, then

- $\lambda_m = \|A\|^2 = \max_{x \in S^{m-1}} X(x)$,
- in case $\lambda_1 > 0$, $\lambda_1 = \frac{1}{\|A^{-1}\|^2} = \min_{x \in S^{m-1}} X(x)$.

It follows that

$$\kappa(A) = \left(\frac{\lambda_m}{\lambda_1}\right)^{\frac{1}{2}}$$

when $\lambda_1 > 0$. We put $\kappa(A) = +\infty$ if $\lambda_1 = 0$. Note also that $\kappa(A) \ge 1$ and $\kappa(rA) = \kappa(A)$ for any real r, $r \neq 0$.

There is an important difference between the proof of Theorem 1.1 and those of the two other theorems mentioned above. In the latter cases, one puts

$$\log \kappa(A) = \frac{1}{2}\log \lambda_m - \frac{1}{2}\log \lambda_1,$$

and if one takes expectations, the joint distribution of the random variables λ_m, λ_1 does not play any role; the proof uses only the individual distributions of λ_m and λ_1. On the contrary, the proof below of Theorem 1.1 depends essentially on the joint distribution of the pair (λ_m, λ_1). A general formula for the joint density of $\lambda_1, \dots, \lambda_m$ has been well known for a long time (see, for example, Wilks [21], Wigner [19], Krishnaiah and Chang [11], Kendall, Stuart, and Ord [10], and the references therein), but it seems to be difficult to adapt this to our present requirements. In fact, we will use a different approach, based on the expected value of the number of zeros of a random field parameterized on a smooth manifold.

We have also applied this technique to give a new proof of the known result of Lemma 2.2, a lower bound for $\mathrm{P}\{\lambda_1 < a\}$.

One can ask if Theorem 1.1 follows from the well-known exponential bounds for the concentration of the distribution of λ_m together with known bounds for the distribution of λ_1 (see, for example, Szarek [15], Davidson and Szarek [5], and Ledoux [12] for these types of inequalities).

More precisely, consider the upper bound in Theorem 1.1. For $\varepsilon > 0$ one has

$$\begin{aligned}
\mathrm{P}\big\{\kappa(A) > mx\big\} &= \mathrm{P}\left\{\frac{\lambda_m}{\lambda_1} > m^2x^2\right\} \\
&\le \mathrm{P}\big\{\lambda_m > (4+\varepsilon)m\big\} + \mathrm{P}\left\{\lambda_m \le (4+\varepsilon)m, \frac{\lambda_m}{\lambda_1} > m^2x^2\right\} \\
&\le \mathrm{P}\big\{\lambda_m > (4+\varepsilon)m\big\} + \mathrm{P}\left\{\lambda_1 < \frac{(4+\varepsilon)}{mx^2}\right\} \\
&\le C_1 \exp\left[-C_2 m\varepsilon^2\right] + C_3\frac{\sqrt{4+\varepsilon}}{x},
\end{aligned} \tag{1.2}$$

where C_1, C_2, C_3 are positive constants. From (1.2), making an adequate choice of ε one can get an upper bound for $\mathrm{P}\{\kappa(A) > mx\}$ of the form $(\mathrm{const})\frac{1}{x}\big(\frac{\log x}{m}\big)^\alpha$ for some $\alpha > 0$ and x large enough. However, this kind of argument does not lead to the precise order given by our Theorem 1.1.

On the other hand, using known results for the distribution of other functions of the spectrum (for example, $(\lambda_1 + \cdots + \lambda_m)/\lambda_1$ as in Edelman [8]), one can get upper and lower bounds for the tails of the distribution of $\kappa(A)$ which again do not reach the precise behavior $(\mathrm{const})/x$.

2. Proof of Theorem 1.1. It is easy to see that, almost surely, the eigenvalues of A^TA are pairwise different. We introduce the following additional notation:

- $\langle .,. \rangle$ is usual scalar product in $\mathbb{R}^m$ and $\{e_1, \dots, e_m\}$ the canonical basis.
- I_k denotes the $k \times k$ identity matrix.
- $B = A^TA = ((b_{ij}))_{i,j=1,\dots,m}$.
- For $s \neq 0$ in $\mathbb{R}^m$, $\pi_s : \mathbb{R}^m \to \mathbb{R}^m$ denotes the orthogonal projection onto $\{s\}^\perp$, the orthogonal complement of s in $\mathbb{R}^m$.
- $M \succ 0$ (resp., $M \prec 0$) means that the symmetric matrix M is positive definite (resp., negative definite).

- If ξ is a random vector, $p_\xi(.)$ is the density of its distribution whenever it exists.
- For a differentiable function F defined on a smooth manifold M embedded in some Euclidean space, $F'(s)$ and $F''(s)$ are the first and the second derivative of F that we will represent, in each case, with respect to an appropriate orthonormal basis of the tangent space.

Instead of (1.1) we prove the equivalent statement: for $x > m$,

$$\frac{cm}{x} < \mathrm{P}\{\kappa(A) > x\} < \frac{Cm}{x}. \tag{2.1}$$

We break the proof into several steps. Our main task is to estimate the joint density of the pair (λ_m, λ_1); this will be done in Step 4.

Step 1. For $a, b \in \mathbb{R}$, $a > b$, one has almost surely

$$\begin{aligned} &\{\lambda_m \in (a, a+da)\,, \lambda_1 \in (b, b+db)\} \\ &= \left\{ \begin{array}{c} \exists\, s,t \in S^{m-1}, \langle s,t\rangle = 0, X(s) \in (a,a+da)\,, X(t) \in (b,b+db)\,, \\ \pi_s(Bs) = 0, \pi_t(Bt) = 0, X''(s) \prec 0, X''(t) \succ 0 \end{array} \right\}. \end{aligned} \tag{2.2}$$

An instant reflection shows that almost surely the number

$$N_{a,b,da,db}$$

of pairs (s,t) belonging to the right-hand side of (2.2) is equal to 0 or to 4, so that

$$\mathrm{P}\{\lambda_m \in (a,a+da), \lambda_1 \in (b,b+db)\} = \frac{1}{4}\mathrm{E}(N_{a,b,da,db}). \tag{2.3}$$

Step 2. In this step we will give a bound for $\mathrm{E}\left(N_{a,b,da,db}\right)$ using what we call a Rice-type formula (see Azaïs and Wschebor [3] for some related problems and general tools). Let

$$V = \left\{(s,t) : s,t \in S^{m-1}, \langle s,t\rangle = 0\right\}.$$

V is a C^∞-differentiable manifold without boundary, embedded in $\mathbb{R}^{2m}$, $\dim(V) = 2m-3$. We will denote by $\tau = (s,t)$ a generic point in V and by $\sigma_V(d\tau)$ the geometric measure on V.

It is easy to see that $\sigma_V(V) = \sqrt{2}\sigma_{m-1}.\sigma_{m-2}$, where σ_{m-1} denotes the surface area of $S^{m-1} \subset \mathbb{R}^m$, that is, $\sigma_{m-1} = \frac{2\pi^{m/2}}{\Gamma(m/2)}$. On V we define the random field

$$Y : V \to \mathbb{R}^{2m}$$

by means of

$$Y(s,t) = \begin{pmatrix} \pi_s(Bs) \\ \pi_t(Bt) \end{pmatrix}.$$

For $\tau = (s,t)$ a given point in V, we have that

$$Y(\tau) \in \{(t,-s)\}^\perp \cap \left\{\{s\}^\perp \times \{t\}^\perp\right\} = W_\tau$$

for any value of the matrix B, where $\{(t,-s)\}^\perp$ is the orthogonal complement of the point $(t,-s)$ in $\mathbb{R}^{2m}$.

In fact, $(t,-s) \in \{s\}^{\perp} \times \{t\}^{\perp}$ and

$$\begin{aligned}\langle Y(s,t),(t,-s)\rangle_{\mathbb{R}^{2m}} &= \langle \pi_s(Bs),t\rangle - \langle \pi_t(Bt),s\rangle \\ &= \langle Bs - \langle s,Bs\rangle s,t\rangle - \langle Bt - \langle t,Bt\rangle t,s\rangle = 0\end{aligned}$$

since $\langle s,t\rangle = 0$ and B is symmetric. Notice that $\dim(W_\tau) = 2m-3$.

We also set

$$\Delta(\tau) = \left[\det\left[\left(Y'(\tau)\right)^T Y'(\tau)\right]\right]^{\frac{1}{2}},$$

$$N = \sharp\left\{\tau : \tau \in V, Y(\tau) = 0\right\}.$$

For $\tau = (s,t) \in V$, F_τ denotes the event

$$F_\tau = \left\{X(s) \in (a,a+da), X(t) \in (b,b+db), X''(s) \prec 0, X''(t) \succ 0\right\},$$

and $p_{Y(\tau)}(.)$ is the density of the random vector $Y(\tau)$ in the $(2m-3)$-dimensional subspace W_τ of $\mathbb{R}^{2m}$.

Assume that 0 is not a critical value of Y, that is, if $Y(\tau) = 0$, then $\Delta(\tau) \neq 0$. This holds true with probability 1. By compactness of V, this implies $N < \infty$. Assume that $N \neq 0$ and denote by $\tau_1,\dots,\tau_N$ the roots of the equation $Y(\tau) = 0$.

Because of the implicit function theorem, if $\delta > 0$ is small enough, one can find in V open neighborhoods $U_1,\dots,U_N$ of the points $\tau_1,\dots,\tau_N$, respectively, so that the following hold:

- Y is a diffeomorphism between U_j and $Y(V) \cap B_{2m}(0,\delta)$ ($B_{2m}(0,\delta)$ is the Euclidean ball of radius δ centered at the origin, in $\mathbb{R}^{2m}$).
- $U_1,\dots,U_N$ are pairwise disjoint.
- If $\tau \notin \bigcup_{j=1}^N U_j$, then $Y(\tau) \notin B_{2m}(0,\delta)$.

Using the change of variable formula, it follows that

$$\int_V \Delta(\tau)\ 1_{\{\|Y(\tau)\|<\delta\}}\sigma_V(d\tau) = \sum_{j=1}^N \int_{U_j} \Delta(\tau)\ \sigma_V(d\tau) = \sum_{j=1}^N \mu\left(Y(U_j)\right), \tag{2.4}$$

where $\mu\left(Y(U_j)\right)$ denotes the—$(2m-3)$-dimensional—geometric measure of $Y(U_j)$. As $\delta \downarrow 0$, $\mu\left(Y(U_j)\right) \sim |B_{2m-3}(\delta)|$, where $|B_{2m-3}(\delta)|$ is the $(2m-3)$-dimensional Lebesgue measure of a ball of radius δ in $\mathbb{R}^{2m-3}$. It follows from (2.4) that, almost surely,

$$N = \lim_{\delta\downarrow 0} \frac{1}{|B_{2m-3}(\delta)|} \int_V \Delta(\tau)\mathbb{1}_{\{\|Y(\tau)\|<\delta\}}\sigma_V(d\tau).$$

In exactly the same way, one can prove that

$$N_{a,b,da,db} = \lim_{\delta\downarrow 0} \frac{1}{|B_{2m-3}(\delta)|} \int_V \Delta(\tau)\mathbb{1}_{F_\tau}\mathbb{1}_{\{\|Y(\tau)\|<\delta\}}\sigma_V(d\tau).$$

Applying Fatou's lemma and Fubini's theorem,

$$\begin{aligned}\mathrm{E}(N_{a,b,da,db}) &\leq \liminf_{\delta\downarrow 0} \frac{1}{|B_{2m-3}(\delta)|} \int_V \mathrm{E}\left(\Delta(\tau)\ \mathbb{1}_{F_\tau}\mathbb{1}_{\{\|Y(\tau)\|<\delta\}}\right)\sigma_V(d\tau) \\ &= \liminf_{\delta\downarrow 0} \int_V \sigma_V(d\tau) \int_{B_{m,\delta,\tau}} \mathrm{E}\left(\Delta(\tau)\mathbb{1}_{F_\tau}/Y(\tau) = y\right) p_{Y(\tau)}(y)\ \frac{dy}{|B_{2m-3}(\delta)|} \\ &= \int_V \mathrm{E}\left(\Delta(\tau)\mathbb{1}_{F_\tau}/Y(\tau) = 0\right) p_{Y(\tau)}(0)\ \sigma_V(d\tau),\end{aligned}$$

where $B_{m,\delta,\tau} = B_{2m}(0,\delta) \cap W_\tau$. The validity of the last passage to the limit will become clear below, since it will follow from the calculations we will perform that the integrand in the inner integral is a continuous function of the pair (τ, y). Hence,

$$\text{(2.5)} \quad \mathrm{E}\left(N_{a,b,da,db}\right) \leq \int_a^{a+da} dx \int_b^{b+db} dy \int_V \mathrm{E}\left(\Delta(s,t)\mathbb{1}_{\{X''(s)\prec 0, X''(t)\succ 0\}}/\mathcal{C}_{s,t,x,y}\right) \times p_{X(s),X(t),Y(s,t)}(x,y,0)\ \sigma_V(d(s,t)),$$

where $\mathcal{C}_{s,t,x,y}$ is the condition $\{X(s) = x, X(t) = y, Y(s,t) = 0\}$. The invariance of the law of A with respect to isometries of $\mathbb{R}^m$ implies that the integrand in (2.5) does not depend on $(s,t) \in V$. Hence, we have proved that the joint law of λ_m and λ_1 has a density $g(a,b)$, $a > b$, and

$$\text{(2.6)} \quad g(a,b) \leq \frac{\sqrt{2}}{4}\sigma_{m-1}\sigma_{m-2}\mathrm{E}\left(\Delta(e_1,e_2)\mathbb{1}_{\{X''(e_1)\prec 0, X''(e_2)\succ 0\}}/\mathcal{C}_{e_1,e_2,a,b}\right) \times p_{X(e_1),X(e_2),Y(e_1,e_2)}(a,b,0).$$

In fact, using the method of Azaïs and Wschebor [3], it could be proved that (2.6) is an equality, but we do not need such a precise result here.

Step 3. Next, we compute the ingredients in the right-hand member of (2.6). We take as orthonormal basis for the subspace $W_{(e_1,e_2)}$

$$\left\{(e_3,0),\ldots,(e_m,0),(0,e_3),\ldots,(0,e_m),\frac{1}{\sqrt{2}}(e_2,e_1)\right\} = L_1.$$

We have

$$\begin{aligned} X(e_1) &= b_{11}, \\ X(e_2) &= b_{22}, \\ X''(e_1) &= B_1 - b_{11}I_{m-1}, \\ X''(e_2) &= B_2 - b_{22}I_{m-1}, \end{aligned}$$

where B_1 (resp., B_2) is the $(m-1)\times(m-1)$ matrix obtained by suppressing the first (resp., the second) row and column in B,

$$Y(e_1,e_2) = (0, b_{21}, b_{31}, \ldots, b_{m1}, b_{12}, 0, b_{32}, \ldots, b_{m2})^T,$$

so that it has the following expression in the orthonormal basis L_1:

$$Y(e_1,e_2) = \sum_{i=3}^{m}\left(b_{i1}(e_i,0) + b_{i2}(0,e_i)\right) + \sqrt{2}b_{12}\left(\frac{1}{\sqrt{2}}(e_2,e_1)\right).$$

It follows that the joint density of $X(e_1), X(e_2), Y(e_1,e_2)$ appearing in (2.6) in the space $\mathbb{R}\times\mathbb{R}\times W_{(e_1,e_2)}$ is the joint density of the random variables

$$b_{11}, b_{22}, \sqrt{2}b_{12}, b_{31}, \ldots, b_{m1}, b_{32}, \ldots, b_{m2}$$

at the point $(a,b,0,\ldots,0)$. To compute this density, first compute the joint density q of

$$b_{31}, \ldots, b_{m1}, b_{32}, \ldots, b_{m2},$$

given a_1, a_2, where a_j denotes the jth column of A which is Gaussian standard in $\mathbb{R}^m$. q is the normal density in $\mathbb{R}^{2(m-2)}$, centered with variance matrix

$$\begin{pmatrix} \|a_1\|^2 I_{m-2} & \langle a_1, a_2\rangle I_{m-2} \\ \langle a_1, a_2\rangle I_{m-2} & \|a_2\|^2 I_{m-2} \end{pmatrix}.$$

Set

$$a'_j = \frac{a_j}{\|a_j\|}, \ j = 1, 2.$$

The density of the triplet

$$(b_{11}, b_{22}, b_{12}) = (\|a_1\|^2, \|a_2\|^2, \|a_1\|\|a_2\|\langle a'_1, a'_2\rangle)$$

at the point $(a, b, 0)$ can be computed as follows.

Since $\langle a'_1, a'_2\rangle$ and $(\|a_1\|, \|a_2\|)$ are independent, the density of the triplet at $(a, b, 0)$ is equal to

$$\chi^2_m(a)\chi^2_m(b)(ab)^{-1/2} p_{<a'_1, a'_2>}(0),$$

where $\chi^2_m(.)$ denotes the χ^2 density with m degrees of freedom.

Let $\xi = (\xi_1, \dots, \xi_m)^T$ be Gaussian standard in $\mathbb{R}^m$. Clearly, $\langle a'_1, a'_2\rangle$ has the same distribution as $\frac{\xi_1}{\|\xi\|}$, because of the invariance under rotations.

$$\begin{aligned}\frac{1}{2t}\mathrm{P}\{|\langle a'_1, a'_2\rangle| \le t\} = \frac{1}{2t}\mathrm{P}\left\{\frac{\xi_1^2}{\chi^2_{m-1}} \le \frac{t^2}{1-t^2}\right\} &= \frac{1}{2t}\mathrm{P}\left\{F_{1,m-1} \le \frac{t^2(m-1)}{1-t^2}\right\} \\ &= \frac{1}{2t}\int_0^{\frac{t^2(m-1)}{1-t^2}} f_{1,m-1}(x)dx,\end{aligned}$$

where $\chi^2_{m-1} = \xi_2^2 + \cdots + \xi_m^2$ and $F_{1,m-1}$ has the Fisher distribution with $(1, m-1)$ degrees of freedom and density $f_{1,m-1}$. Letting $t \to 0$, we obtain

$$p_{\langle a'_1, a'_2\rangle}(0) = \frac{1}{\sqrt{\pi}}\frac{\Gamma(m/2)}{\Gamma\big((m-1)/2\big)}.$$

Summing up, the density in (2.6) is equal to

$$\frac{1}{\sqrt{2}}(2\pi)^{2-m}\pi^{-\frac{1}{2}}\frac{1}{\Gamma(m/2)\Gamma\big((m-1)/2\big)}2^{-m}\frac{1}{\sqrt{ab}}\exp\Big(-\frac{a+b}{2}\Big). \tag{2.7}$$

We now consider the conditional expectation in (2.6). First, observe that the $(2m-3)$-dimensional tangent space to V at the point (s, t) is parallel to the orthogonal complement in $\mathbb{R}^m \times \mathbb{R}^m$ of the triplet of vectors $(s, 0); (0, t); (t, s)$. This is immediate from the definition of V.

To compute the associated matrix for $Y'(e_1, e_2)$ take the set

$$\Big\{(e_3, 0), \dots, (e_m, 0), (0, e_3), \dots, (0, e_m), \frac{1}{\sqrt{2}}(e_2, -e_1)\Big\} = L_2$$

as orthonormal basis in the tangent space and the canonical basis in $\mathbb{R}^{2m}$. A direct calculation gives

$$Y'(e_1,e_2)=\begin{pmatrix} -v^T & 0_{1,m-2} & -\frac{1}{\sqrt{2}}b_{21} \\ w^T & 0_{1,m-2} & \frac{1}{\sqrt{2}}(-b_{11}+b_{22}) \\ B_{12}-b_{11}I_{m-2} & 0_{m-2,m-2} & \frac{1}{\sqrt{2}}w \\ 0_{1,m-2} & -w^T & \frac{1}{\sqrt{2}}(-b_{11}+b_{22}) \\ 0_{1,m-2} & v^T & \frac{1}{\sqrt{2}}b_{21} \\ 0_{m-2,m-2} & B_{12}-b_{22}I_{m-2} & -\frac{1}{\sqrt{2}}v \end{pmatrix},$$

where $v^T=(b_{31},\dots,b_{m1}), w^T=(b_{32},\dots,b_{m2}), 0_{i,j}$ is a null matrix with i rows and j columns, and B_{12} is obtained from B by suppressing the first and second rows and columns. The columns represent the derivatives in the directions of L_2 at the point (e_1,e_2). The first m rows correspond to the components of $\pi_s(Bs)$, the last m ones to those of $\pi_t(Bt)$. Thus, under the condition $\mathcal{C}_{e_1,e_2,a,b}$ that is used in (2.6),

$$Y'(e_1,e_2)=\begin{pmatrix} 0_{1,m-2} & 0_{1,m-2} & 0 \\ 0_{1,m-2} & 0_{1,m-2} & \frac{1}{\sqrt{2}}(b-a) \\ B_{12}-aI_{m-2} & 0_{m-2,m-2} & 0_{m-2,1} \\ 0_{1,m-2} & 0_{1,m-2} & \frac{1}{\sqrt{2}}(b-a) \\ 0_{1,m-2} & 0_{1,m-2} & 0 \\ 0_{m-2,m-2} & B_{12}-bI_{m-2} & 0_{m-2,1} \end{pmatrix}$$

and

$$\left[\det\left[\big(Y'(e_1,e_2)\big)^T Y'(e_1,e_2)\right]\right]^{\frac{1}{2}}=|\det(B_{12}-aI_{m-2})||\det(B_{12}-bI_{m-2})|(a-b).$$

Step 4. Note that $B_1-aI_{m-1}\prec 0\Rightarrow B_{12}-aI_{m-2}\prec 0$, and similarly, $B_2-bI_{m-1}\succ 0\Rightarrow B_{12}-bI_{m-2}\succ 0$, and that for $a>b$, under $\mathcal{C}_{e_1,e_2,a,b}$, there is equivalence in these relations.

It is also clear that, since $B_{12}\succ 0$, one has

$$|\det(B_{12}-aI_{m-2})|\mathbb{1}_{B_{12}-aI_{m-2}\prec 0}\le a^{m-2},$$

and it follows that the conditional expectation in (2.6) is bounded by

$$a^{m-1}\mathrm{E}\big(|\det(B_{12}-bI_{m-2})|\mathbb{1}_{B_{12}-bI_{m-2}\succ 0}/\mathcal{C}\big), \tag{2.8}$$

where $\mathcal{C}$ is the condition $\{b_{11}=a, b_{22}=b, b_{12}=0, b_{i1}=b_{i2}=0\ (i=3,\dots,m)\}$.

To compute the conditional expectation in (2.8) we further condition on the value of the random vectors a_1 and a_2. Since unconditionally $a_3,\dots,a_m$ are i.i.d. standard Gaussian vectors in $\mathbb{R}^m$, under this new conditioning, their joint law becomes the law of i.i.d. standard Gaussian vectors in $\mathbb{R}^{m-2}$ and independent of the condition. That is, (2.8) is equal to

$$a^{m-1}\mathrm{E}\big(|\det(M-bI_{m-2})|\mathbb{1}_{M-bI_{m-2}\succ 0}\big), \tag{2.9}$$

where M is an $(m-2)\times(m-2)$ random matrix with entries $M_{ij}=\langle v_i,v_j\rangle$ $(i,j=1,\dots,m-2)$ and the vectors $v_1,\dots,v_{m-2}$ are i.i.d. Gaussian standard in $\mathbb{R}^{m-2}$. The expression in (2.9) is bounded by

$$a^{m-1}\mathrm{E}\big(\det(M)\big)=a^{m-1}(m-2)!.$$

The last equality is contained in the following lemma, which is well known; see, for example, Edelman [7].

LEMMA 2.1. *Let $\xi_1, \dots, \xi_m$ be i.i.d. random vectors in $\mathbb{R}^p$, $p \geq m$, their common distribution being Gaussian centered with variance I_p.*

Denote by $W_{m,p}$ the matrix

$$W_{m,p} = ((\langle \xi_i, \xi_j \rangle))_{i,j=1,\dots,m},$$

and by

$$D(\lambda) = \det(W_{m,p} - \lambda I_m)$$

its characteristic polynomial.

Then

(i)

$$\mathrm{E}(\det(W_{m,p})) = p(p-1)\dots(p-m+1), \tag{2.10}$$

(ii)

$$\mathrm{E}(D(\lambda)) = \sum_{k=0}^{m} (-1)^k \binom{m}{k} \frac{p!}{(p-m+k)!} \lambda^k. \tag{2.11}$$

Returning to the proof of the theorem and summing up this part, after substituting in (2.6), we get

$$g(a,b) \leq C_m \frac{\exp\big(-(a+b)/2\big)}{\sqrt{ab}} a^{m-1}, \tag{2.12}$$

where $C_m = \frac{1}{4(m-2)!}$.

Step 5. Now we prove the upper-bound part in (2.1). One has, for $x > 1$,

(2.13)

$$\mathrm{P}\{\kappa(A) > x\} = \mathrm{P}\left\{\frac{\lambda_m}{\lambda_1} > x^2\right\} \leq \mathrm{P}\left\{\lambda_1 < \frac{L^2 m}{x^2}\right\} + \mathrm{P}\left\{\frac{\lambda_m}{\lambda_1} > x^2, \lambda_1 \geq \frac{L^2 m}{x^2}\right\},$$

where L is a positive number to be chosen later on. For the first term in (2.13), we use Proposition 9 in Cuesta-Albertos and Wschebor [4], which is a slight modification of Theorem 3.2 in Sankar, Spielman, and Teng [13]:

$$\mathrm{P}\left\{\lambda_1 < \frac{L^2 m}{x^2}\right\} = \mathrm{P}\left\{\|A^{-1}\| > \frac{x}{L\sqrt{m}}\right\} \leq C_2(m) \frac{Lm}{x}.$$

Here,

$$C_2(m) = \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \left[\sup_{0<c<m} \sqrt{c}\,\mathrm{P}\left\{t_{m-1}^2 > \frac{(m-1)c}{m-c}\right\}\right]^{-1} \leq C_2(+\infty) \approx 2.3473,$$

where t_{m-1} is a random variable having Student's distribution with $m-1$ degrees of freedom.

For the second term in (2.13),

$$\mathrm{P}\left\{\frac{\lambda_m}{\lambda_1} > x^2, \lambda_1 \geq \frac{L^2 m}{x^2}\right\} = \int_{L^2 m x^{-2}}^{+\infty} db \int_{bx^2}^{+\infty} g(a,b)da \leq G_m(x^2)$$

with

$$G_m(y) = C_m \int_{L^2 m y^{-1}}^{+\infty} db \int_{by}^{+\infty} \frac{\exp\big(-(a+b)/2\big)}{\sqrt{ab}} a^{m-1} da,$$

using (2.12). We have

$$\text{(2.14)} \quad G'_m(y) = C_m \Bigg[-\int_{L^2 m y^{-1}}^{+\infty} \exp(-b/2)\sqrt{b}\exp\big(-(by)/2\big)(by)^{m-3/2} db \\ + L^2 m y^{-2} \int_{L^2 m}^{+\infty} \exp\left(-\frac{1}{2}\left(a + \frac{L^2 m}{y}\right)\right) a^{m-3/2} L^{-1} m^{-\frac{1}{2}} y^{\frac{1}{2}} da \Bigg],$$

which implies

$$\begin{aligned} -G'_m(y) &\leq C_m y^{m-3/2} \int_{L^2 m y^{-1}}^{+\infty} \exp\left(-\frac{b(1+y)}{2}\right) b^{m-1} db \\ &= \frac{y^{-3/2}}{4(m-2)!} \left(\frac{y}{1+y}\right)^m 2^m \int_{\frac{L^2 m}{2y}(1+y)}^{+\infty} e^{-z} z^{m-1} dz \\ &\leq \frac{y^{-3/2}}{4(m-2)!} 2^m \int_{\frac{L^2 m}{2}}^{+\infty} e^{-z} z^{m-1} dz. \end{aligned}$$

Put $I_m(a) = \int_a^{+\infty} e^{-z} z^{m-1} dz$. Integrating by parts,

$$I_m(a) = e^{-a}\big[a^{m-1} + (m-1)a^{m-2} + (m-1)(m-2)a^{m-3} + \cdots + (m-1)!\big],$$

so that for $a > 2.5m$

$$I_m(a) \leq \frac{5}{3} e^{-a} a^{m-1}.$$

If $L^2 > 5$, we obtain the bound

$$-G'_m(y) \leq D_m y^{-3/2} \quad \text{with} \quad D_m = \frac{5}{6} \frac{m^{m-1}}{(m-2)!} L^{2(m-1)} \exp\left(-\frac{L^2 m}{2}\right).$$

We now apply Stirling's formula (Abramowitz and Stegun [1, sect. 6.1.38]), i.e., for all $x > 0$

$$\Gamma(x+1)\exp\left(-\frac{1}{12x}\right) \leq \left(\frac{x}{e}\right)^x \sqrt{2\pi x} \leq \Gamma(x+1),$$

to get

$$D_m \leq \frac{5\sqrt{2}}{12\sqrt{\pi}L^2} \frac{m}{\sqrt{m-2}} \exp\left(-m\frac{L^2 - 4log(L) - 2}{2}\right) \leq \frac{5\sqrt{2}}{12\sqrt{\pi}L^2} m,$$

if we choose for L the only root larger than 1 of the equation $L^2 - 4\log(L) - 2 = 0$ (check that $L \approx 2.3145$). To finish,

$$0 \leq G_m(y) = \int_y^{+\infty} -G'_m(t)dt < D_m \int_y^{+\infty} \frac{dt}{t^{3/2}} = 2D_m y^{-\frac{1}{2}}.$$

Replacing y by x^2 and performing the numerical evaluations, the upper bound in (2.1) follows, and we get for the constant C the value 5.60.

Step 6. We consider now the lower bound in (2.1). For $\gamma > 0$ and $x > 1$, we have

$$\text{(2.15)}\quad \begin{aligned} \mathrm{P}\{\kappa(A) > x\} &= \mathrm{P}\left\{\frac{\lambda_m}{\lambda_1} > x^2\right\} \geq \mathrm{P}\left\{\frac{\lambda_m}{\lambda_1} > x^2, \lambda_1 < \frac{\gamma^2 m}{x^2}\right\} \\ &= \mathrm{P}\left\{\lambda_1 < \frac{\gamma^2 m}{x^2}\right\} - \mathrm{P}\left\{\frac{\lambda_m}{\lambda_1} \leq x^2, \lambda_1 < \frac{\gamma^2 m}{x^2}\right\}. \end{aligned}$$

A lower bound for the first term in the right-hand member of (2.15) is obtained using the following inequality, which we state as a separate lemma. In fact, this result is known; see, for example, Szarek [15, Theorem 1.2], where it is proved without giving an explicit value for the constant. See also Edelman [7, Corollary 3.1], for a related result.

LEMMA 2.2. *If* $0 < a < 1/m$, *then*

$$\mathrm{P}\{\lambda_1 < a\} \geq \beta\sqrt{am},$$

where we can choose $\beta = \left(\frac{2}{3}\right)^{3/2} e^{-1/3}$.

Proof. Define the index $i_X(t)$ of a critical point $t \in S^{m-1}$ of the function X as the number of negative eigenvalues of X"(t). For each $a > 0$ put

$$N_i(a) = \sharp\{t \in S^{m-1} : X(t) = t^T Bt < a, X'(t) = 0, i_X(t) = i\}$$

for $i = 0, 1, \ldots, m-1$. One easily checks that if the eigenvalues of B are $\lambda_1, \ldots, \lambda_m$, $0 < \lambda_1 < \cdots < \lambda_m$, then

- if $a \leq \lambda_1$, then $N_i(a) = 0$ for $i = 0, 1, \ldots, m-1$;
- if $\lambda_i < a \leq \lambda_{i+1}$, then $N_k(a) = 2$ for $k = 0, \ldots, i-1$, $N_k(a) = 0$ for $k = i, \ldots, m-1$; for some $i = 0, 1, \ldots, m_1$
- if $\lambda_m < a$, then $N_i(a) = 2$ for $i = 0, 1, \ldots, m-1$.

Now consider

$$M(a) = \sum_{i=0}^{m-1} (-1)^i N_i(a).$$

$M(a)$ is the Euler characteristic of the set $S = \{t \in S^{m-1} : X(t) < a\}$; see Adler [2]. It follows from the relations above that

- if $N_0(a) = 0$, then $N_i(a) = 0$ for $i = 1, \ldots, m-1$, and hence $M(a) = 0$;
- if $N_0(a) = 2$, then $M(a) = 0$ or 2,

so that in any case

$$M(a) \leq N_0(a).$$

Hence,

$$\mathrm{P}\{\lambda_1 < a\} = \mathrm{P}\{N_0(a) = 2\} = \frac{1}{2}\mathrm{E}\big(N_0(a)\big) \geq \frac{1}{2}\mathrm{E}\big(M(a)\big). \tag{2.16}$$

The expectation of $M(a)$ can be written using the Rice-type formula (see Azaïs and Wschebor [3] or Taylor and Adler [16])

$$\mathrm{E}\big(M(a)\big) = \int_0^a dy \int_{S^{m-1}} \mathrm{E}\big[\det\big(X"(t)\big)/X(t) = y, X'(t) = 0\big] p_{X(t),X'(t)}(y,0)\sigma_{m-1}(dt)$$
$$= \int_0^a \sigma_{m-1}(S^{m-1})\mathrm{E}\big[\det\big(X"(e_1)\big)/X(e_1) = y, X'(e_1) = 0\big] p_{X(e_1),X'(e_1)}(y,0)dy,$$

where we have used again invariance under isometries. Applying a similar Gaussian regression—as we did in Step 4 to get rid of the conditioning—we obtain

$$\mathrm{E}\big(M(a)\big) = \int_0^a \mathrm{E}\big[\det\big(Q - yI_{m-1}\big)\big] \frac{\sqrt{2\pi}}{2^{m-1}}\Gamma^{-2}\left(\frac{m}{2}\right)\frac{\exp(-y/2)}{\sqrt{y}}dy, \tag{2.17}$$

where Q is an $(m-1)\times(m-1)$ random matrix with entry i,j equal to $(\langle v_i, v_j\rangle)$ and $v_1,\dots,v_{m-1}$ are i.i.d. Gaussian standard in $\mathbb{R}^{m-1}$. We now use part (ii) of Lemma 2.1:

$$\mathrm{E}\big[\det\big(Q - yI_{m-1}\big)\big] = (m-1)!\sum_{k=0}^{m-1}\binom{m-1}{k}\frac{(-y)^k}{k!}. \tag{2.18}$$

Under condition $0 < a < m^{-1}$, since $0 < y < a$, as k increases, the terms of the sum in the right-hand member of (2.18) have decreasing absolute value, so that

$$\mathrm{E}\big[\det\big(Q - yI_{m-1}\big)\big] \geq (m-1)![1-(m-1)y].$$

Substituting into the right-hand member of (2.17), we get

$$\mathrm{E}\big[M(a)\big] \geq \frac{\sqrt{2\pi}}{2^{m-1}}\frac{(m-1)!}{\Gamma^2(m/2)}J_m(a),$$

where, using again $0 < a < m^{-1}$,

$$J_m(a) = \int_0^a \big(1-(m-1)y\big)\frac{\exp(-y/2)}{\sqrt{y}}dy \geq \int_0^a \frac{\big(1-(m-1)y\big)}{\sqrt{y}}(1-y/2)dy \geq \frac{4}{3}\sqrt{a}$$

by an elementary computation. Going back to (2.17), applying Stirling's formula, and remarking that $\big(1+1/n\big)^{n+1} \geq e$, we get

$$\mathrm{P}\{\lambda_1 < a\} \geq \left(\frac{2}{3}\right)^{3/2} e^{-1/3}\sqrt{am}.$$

This proves the lemma. $\square$

End of the proof of Theorem 1.1. Using Lemma 2.2, the first term on the right-hand side of (2.15) is bounded below by

$$\beta\gamma\frac{m}{x}.$$

TABLE 1
Values of the estimations $\mathrm{P}\{\kappa(A) > mx\}$ *for* $x = 1, 2, 3, 5, 10, 15, 30, 50, 100$ *and* $m = 3, 5, 10, 30, 100, 300, 500$ *by Monte Carlo method over* 40,000 *simulations.*

	Value of x								
Probability	1	2	3	5	10	20	30	50	100
Lower bound: $.13/x$	.13	.065	.043	.026	.013	.007	.004	.003	.001
Upper bound: $5.6/x$	1	1	1	1	.56	.28	.187	.112	.056
$m = 3$	.881	.57	.41	.26	.13	.067	.044	.027	.013
$m = 5$	.931	.66	.48	.30	.16	.079	.053	.033	.016
$m = 10$	.959	.71	.52	.34	.17	.088	.059	.035	.017
$m = 30$	.974	.75	.56	.36	.19	.096	.063	.038	.019
$m = 100$	.978	.77	.58	.38	.20	.098	.066	.040	.019
$m = 300$	.982	.77	.58	.38	.20	.101	.069	.041	.022
$m = 500$	.980	.77	.59	.38	.20	.100	.066	.039	.020

To obtain a bound for the second term, we use again our upper bound (2.12) on the joint density $g(a,b)$, so that we obtain

$$
\begin{aligned}
(2.19)\quad \mathrm{P}\Big\{\frac{\lambda_m}{\lambda_1} \le x^2, \lambda_1 < \frac{\gamma^2 m}{x^2}\Big\} &= \int_0^{\frac{\gamma^2 m}{x^2}} db \int_b^{bx^2} g(a,b)da \\
&\le C_m \int_0^{\frac{\gamma^2 m}{x^2}} db \int_b^{bx^2} \frac{\exp\big(-(a+b)/2\big)}{\sqrt{ab}} a^{m-1} da \\
&\le C_m \int_0^{\frac{\gamma^2 m}{x^2}} b(x^2-1) b^{-\frac{1}{2}} (bx^2)^{\frac{m-3}{2}} db \\
&\le \frac{1}{4(m-2)!} \frac{x^2-1}{x^3} \gamma^{2m} m^{m-1} \le \frac{\sqrt{2}}{8\sqrt{\pi}} e^m \gamma^{2m} \frac{m}{x}
\end{aligned}
$$

on applying Stirling's formula. Now choosing $\gamma = 1/e$, we see that the hypothesis of Lemma 2.2 is satisfied and also

$$\mathrm{P}\Big\{\frac{\lambda_m}{\lambda_1} \le x^2, \lambda_1 < \frac{\gamma^2 m}{x^2}\Big\} \le \frac{\sqrt{2}}{8\sqrt{\pi}} e^{-3} \frac{m}{x}.$$

Substituting into (2.15), we obtain the lower bound in (1.1) with

$$c = \left(\frac{2}{3}\right)^{3/2} e^{-4/3} - \frac{\sqrt{2}}{8\sqrt{\pi}} e^{-3} \approx 0.138. \qquad \square$$

3. Monte Carlo experiment. To study the tail of the distribution of the condition number of Gaussian matrices of various size, we used the following Matlab functions:

- `normrnd`, to simulate normal variables;
- `cond`, to compute the condition number of matrix A.

The results of over 40,000 simulations using Matlab are given in Table 1 and in Figure 1.

The table suggests, taking into account the simulation variability, that the constants c and C should take values smaller than 0.88 and bigger than 2.00, respectively.

Acknowledgments. The authors thank Professors G. Letac and F. Cucker for valuable discussions. They also thank the associate editor and two anonymous referees for helpful comments that contributed to improving this paper.

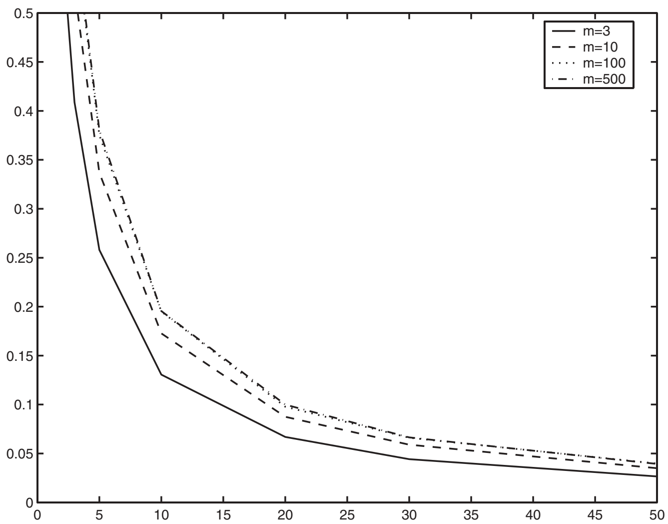

Fig. 1. *Values of* $\mathrm{P}\{\kappa(A) > mx\}$ *as a function of* x *for* $m = 3, 10, 100,$ *and* 500.

REFERENCES

[1] M. Abramowitz and I. A. Stegun, eds., *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, John Wiley & Sons, New York, 1984

[2] R. J. Adler, *The Geometry of Random Fields*, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, Chichester, UK, 1981.

[3] J.-M. Azaïs and M. Wschebor, *On the distribution of the maximum of a Gaussian field with d parameters*, Ann. Appl. Probab., to appear.

[4] J. Cuesta-Albertos and M. Wschebor, *Condition numbers and extrema of random fields*, in Seminar on Stochastic Analysis, Random Fields and Applications IV, Progr. Probab. 58, Birkhäuser, Basel, 2004, pp. 69–82.

[5] K. R. Davidson and S. J. Szarek, *Local operator theory, random matrices and Banach spaces*, in Handbook of the Geometry of Banach Spaces, Vol. I, North-Holland, Amsterdam, 2001, pp. 317–366.

[6] J. W. Demmel, *Applied Numerical Linear Algebra*, SIAM, Philadelphia, 1997.

[7] A. Edelman, *Eigenvalues and condition numbers of random matrices*, SIAM J. Matrix Anal. Appl., 9 (1988), pp. 543–560.

[8] A. Edelman, *On the distribution of a scaled condition number*, Math. Comp., 58 (1992), pp. 185–190.

[9] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*, 2nd ed., SIAM, Philadelphia, 2002.

[10] M. Kendall, A. Stuart, and J. K. Ord, *The Advanced Theory of Statistics. Vol.* 3, 4th ed., Macmillan, New York, 1983.

[11] P. R. Krishnaiah and T. C. Chang, *On the exact distributions of the extreme roots of the Wishart and* MANOVA *matrices*, J. Multivariate Anal., 1 (1971), pp. 108–117.

[12] M. Ledoux, *A remark on hypercontractivity and tail inequalities for the largest eigenvalues of random matrices*, Séminaire de Probabilités XXXVII, Lecture Notes in Math. 1832, Springer-Verlag, Berlin, 2003, pp. 360–369.

[13] A. Sankar, D. A. Spielman, and S. H. Teng, *Smoothed Analysis of the Condition Number and Growth Factors of Matrices*, preprint, 2002.

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

[15] S. J. Szarek, *Condition numbers of random matrices*, J. Complexity, 7 (1991), pp. 131–149.

[16] J. E. Taylor and R. J. Adler, *Euler characteristics for Gaussian fields on manifolds*, Ann. Probab., 31 (2003), pp. 533–563.

[17] A. M. Turing, *Rounding-off errors in matrix processes*, Quart. J. Mech. Appl. Math., 1 (1948), pp. 287–308.

[18] J. von Neumann and H. H. Goldstine, *Numerical inverting of matrices of high order*, Bull. Amer. Math. Soc., 53 (1947), pp. 1021–1099.
[19] E. P. Wigner, *Random matrices in physics*, SIAM Rev., 9 (1967), pp. 1–23.
[20] J. H. Wilkinson, *Rounding Errors in Algebraic Processes*, Prentice-Hall, Englewood Cliffs, NJ, 1963.
[21] S. S. Wilks, *Mathematical Statistics*, John Wiley & Sons, New York, 1962.
[22] M. Wschebor, *Smoothed analysis of $\kappa(a)$*, J. Complexity, 20 (2004), pp. 97–107.