SOME REMARKS ON THE CONDITION NUMBER OF A REAL RANDOM SQUARE MATRIX

J.A. Cuesta-Albertos*
Departamento de Matemáticas, Estadística y Computación,
Universidad de Cantabria, Spain
E-mail: cuestaj@unican.es

M. Wschebor
Centro de Matemática
Facultad de Ciencias
Universidad de la República, Montevideo, Uruguay
E-mail: wschebor@cmat.edu.uy

Abstract

In this paper we obtain some bounds for the expectation of the logarithm of the condition number of a random matrix whose elements are independent and identically distributed random variables.

We also include some examples and extensions to cover the smoothed analysis as well higher order moments.

Key words and phrases: Condition number, random matrices.
A.M.S. 1980 subject classification: 15A52, 65F35.

1 Introduction.

Let $A = (a_{i,j})_{i,j=1,\dots,n}$ be a real $n \times n$ non singular matrix. We denote by

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

the operator norm of A, where in the right-hand member of (1), "$\|\cdot\|$" is the euclidean norm in $I\!R^n$.

Define the *condition number of A* by

*This author has been partially supported by the Spanish Ministerio de Ciencia y Tecnología, grant BFM2002-04430-C02-02.

$$\mathcal{K}(A) = \|A\| \times \|A^{-1}\|. \tag{2}$$

The role of $\mathcal{K}(A)$ in numerical analysis -and specially in numerical linear algebra- has been recognized since a long time ([6], [7]. See also [3] and references therein).

In practice, it is useful to consider that the matrices we have to deal with are obtained in a random way (see [1], [2], [4] and [5]). We will assume that $\{a_{i,j}\}_{i,j=1,\ldots,n}$ are independent and identically distributed (i.i.d.) random variables (r.v.'s) defined on the same probability space (Ω, σ, ν) where the common distribution of $\{a_{i,j}\}_{i,j=1,\ldots,n}$ depends on the kind of problem under consideration.

When the matrix A is random, a certain number of natural questions about complexity of algorithms and effects of round-off errors leads to the study of the probability distribution of the random variable $\log \mathcal{K}(A)$ (Note that $\mathcal{K}(A) \geq 1$ for any matrix A) .

In what follows, we state and prove some elementary inequalities for the expectation and higher order moments of $\log \mathcal{K}(A)$, under quite general conditions on the randomness of A.

In the next section (Theorem 2.2) we prove the inequality for $E[\log \mathcal{K}(A)]$ which constitutes the main result of this paper. In Section 2 we present some applications and possible extensions of Theorem 2.2 including the bounds for higher order moments of $\log \mathcal{K}(A)$.

2 Main result.

The proof employs the following simple a very well known identity.

Lemma 2.1 *If X is a positive random variable defined on the probability space (Ω, σ, ν), then*

$$E[X] = \int_0^\infty \nu[X > t]dt.$$

Theorem 2.2 *Assume that $a_{i,j}, i, j = 1, \ldots n$ are independent and identically distributed random variables and that their common probability distribution P satisfies the following conditions*

1. *For any pair α, β of real numbers, $\alpha < \beta$, one has*

$$P\left([\alpha, \beta]\right) \leq P\left(\left[-\frac{\beta - \alpha}{2}, \frac{\beta - \alpha}{2}\right]\right) . \tag{3}$$

2. $E[|a_{1,1}|^r] = \int_{-\infty}^{\infty} |x|^r P(dx) = 1$, *for some $r > 0$.*

3. *There exist positive numbers C, γ such that*

$$P([-\alpha, \alpha]) \leq C\alpha^\gamma, \ \text{for all } \alpha > 0.$$

Then

$$E[\log \mathcal{K}(A)] \le \left(1+\frac{2}{r}\right)\log n + \frac{1}{r} + \frac{1}{\gamma}\left\{\left[(2+\gamma)\log n + \log C\right]^{+} + 1\right\}, \qquad (4)$$

where $x^{+} = \max(x,0)$ *for real* x.

PROOF.- Note that $\|A\| \le \left(\sum_{i,j=1}^{n} a_{i,j}^2\right)^{1/2}$. So, with the only assumption that the r.v.'s are i.d., for $t>0$

$$\begin{aligned} \nu\left[\|A\| > t\right] &\le \nu\left[n^2 \sup_{i,j=1,\dots n} a_{i,j}^2 > t^2\right] \\ &\le \nu\left[\bigcup_{i,j=1}^{n}\left\{|a_{i,j}| > \frac{t}{n}\right\}\right] \le n^2\nu\left[|a_{1,1}| > \frac{t}{n}\right]. \end{aligned} \qquad (5)$$

Hence, applying Lemma 2.1, for $\alpha_n \ge 0$

$$\begin{aligned} E[\log\|A\|] &\le \alpha_n + \int_{\alpha_n}^{\infty} \nu[\log\|A\| > x]dx \\ &= \alpha_n + \int_{\alpha_n}^{\infty} \nu[\|A\| > e^x]dx \\ &\le \alpha_n + \int_{\alpha_n}^{\infty} n^2\nu\left[|a_{1,1}| > \frac{e^x}{n}\right]dx \\ &\le \alpha_n + \int_{\alpha_n}^{\infty} n^2\left(\frac{n}{e^x}\right)^r dx = \alpha_n + n^{2+r}\frac{1}{r}e^{-r\alpha_n}, \end{aligned} \qquad (6)$$

where the last inequality follows from Markov's inequality and assumption 2.

Now choose $\alpha_n \ge 0$ to minimize the right-hand member of (6), i.e.

$$\alpha_n = \left(1+\frac{2}{r}\right)\log n$$

and it follows that

$$E[\log\|A\|] \le \left(1+\frac{2}{r}\right)\log n + \frac{1}{r}. \qquad (7)$$

Let us now consider the term $\|A^{-1}\|$. Denote $A^{-1} = (b_{i,j})_{i,j=1,\dots,n}$. Thus

$$b_{i,j} = \frac{a^{i,j}}{\det(\mathrm{A})}, i,j = 1,\dots,n,$$

where $a^{i,j}$ is the adjoint of the position (i,j) in the matrix A.

Clearly the r.v.'s $|b_{i,j}|, i,j = 1,..,n$ are i.d. and, so, we may apply (5) to the matrix A^{-1} instead of A thus obtaining

$$
\begin{aligned}
\nu\left[\|A^{-1}\| > t\right] &\le n^2\nu\left[|b_{1,1}| > \frac{t}{n}\right] = n^2\nu\left[\left|\frac{a^{1,1}}{\sum_{j=1}^n a_{1,j}a^{1,j}}\right| > \frac{t}{n}\right] \\
&= n^2\nu\left[\left|a_{1,1} + \sum_{j=2}^n a_{1,j}\frac{a^{1,j}}{a^{1,1}}\right| < \frac{n}{t}\right].
\end{aligned}
$$

The r.v.'s

$$
a_{1,1} \text{ and } \eta = \sum_{j=2}^n a_{1,j}\frac{a^{1,j}}{a^{1,1}}
$$

are independent, so that, for each $\alpha > 0$, denoting by P_η the probability distribution of η, and using Fubini's theorem and assumption 1, we have

$$
\begin{aligned}
\nu\left[|a_{1,1} + \eta| < \alpha\right] &= \int_{-\infty}^{\infty} P[(-\alpha - y, \alpha - y)]P_\eta(dy) \\
&\le \int_{-\infty}^{\infty} P[(-\alpha, \alpha)]P_\eta(dy) = P[(-\alpha, \alpha)].
\end{aligned}
$$

Hence, by assumption 3,

$$
\nu\left[\|A^{-1}\| > t\right] \le n^2 P\left(\left[-\frac{n}{t}, \frac{n}{t}\right]\right) \le n^2 C\left(\frac{n}{t}\right)^\gamma, \tag{8}
$$

and, with $\beta_n \ge 0$:

$$
\begin{aligned}
E[\log\|A^{-1}\|] &\le \beta_n + \int_{\beta_n}^{\infty} \nu\left[\|A^{-1}\| > e^x\right] dx \\
&\le \beta_n + \int_{\beta_n}^{\infty} Cn^{2+\gamma}e^{-\gamma x}dx = \beta_n + C\frac{n^{2+\gamma}}{\gamma}e^{-\gamma\beta_n}.
\end{aligned}
$$

Choosing

$$
\beta_n = \frac{1}{\gamma}[(2+\gamma)\log n + \log C]^+,
$$

one obtains

$$
E[\log\|A^{-1}\|] \le \frac{1}{\gamma}\left\{ [(2+\gamma)\log n + \log C]^+ + 1\right\}, \tag{9}
$$

and putting together (7) and (9), (4) is obtained.

•

Next we discuss briefly the assumptions in Theorem 2.2

Remark 2.2.1 It is not too difficult to show that assumption 1 implies that P is symmetric around 0. Therefore, if the r.v.'s $a_{i,j}, i, j = 1, .., n$ are integrable, their common expectation must be 0.

Remark 2.2.2 Since $\mathcal{K}(\lambda A) = \mathcal{K}(A)$ for any real number λ and any nonsingular matrix A, in case $m_r = \int_{-\infty}^{\infty} |x|^r P(dx) < \infty$ it is possible to replace A by $m_r^{-1/r} A$ so that assumption 2 holds without modifying the condition number. Of course in this case one must change accordingly the constant C in assumption 3.

In this sense, assumption 2 is not more restrictive than the finiteness of the r-th moment of the probability measure P.

3 Examples and extensions.

2.1 Assume that P has a density function f, that f is even and non-increasing on $[0, \infty)$ and that $m_r = \int_{-\infty}^{\infty} |x|^r f(x)dx < \infty$ for some $r > 0$.

We replace the original density f by $m_r^{1/r} f(m_r^{1/r} x)$ so that assumption 2 is satisfied without changing $\mathcal{K}(A)$; assumption 3 is verified with $\gamma = 1$ and $C = 2m_r^{1/r} f(0)$. Inequality (4) becomes

$$E[\log \mathcal{K}(A)] \le \left(1 + \frac{2}{r}\right) \log n + \frac{1}{r} + \left[3 \log n + \frac{1}{r} \log m_r + \log(2f(0))\right]^+ + 1. \qquad (10)$$

•

2.2 Consider the example in which P is the uniform distribution on $[-H, H]$, $H > 0$. In this case, $m_r = H^r (r + 1)^{-1}$ and (10) holds true for any $r > 0$. Letting $r \to +\infty$, we obtain

$$E[\log \mathcal{K}(A)] \le 4 \log n + 1. \qquad (11)$$

•

2.3 (Strong concentration near the mean) Here we analyze a family of distributions supported by $[-1, 1]$ but more concentrated around 0 as the uniform is. Assume that the density has the form

$$\frac{1}{2} \frac{\gamma}{|x|^{1-\gamma}} \mathbf{1}_{[-1,1]}(x),$$

for some γ, $0 < \gamma < 1$.

One has $m_r = \frac{\gamma}{r+\gamma}$ for each $r > 0$ and easily checks that introducing the modification suggested in 2.1 above, assumptions 1, 2 and 3 are satisfied with $C = m_r^{\gamma/r}$. Hence, Theorem 2.2 implies that for any $r > 0$

$$E[\log \mathcal{K}(A)] \le \left(1 + \frac{2}{r}\right) \log n + \frac{1}{r} + \frac{1}{\gamma} \left\{ \left[(2 + \gamma) \log n + \frac{\gamma}{r} \log \frac{\gamma}{r + \gamma}\right]^+ + 1 \right\},$$

and, letting $r \to \infty$ it follows that

$$E[\log \mathcal{K}(A)] \leq \left(2 + \frac{2}{\gamma}\right) \log n + \frac{1}{\gamma}.$$

Notice that in this case, as $\gamma < 1$ the bound we obtain is worse than (11).

•

2.4 The bound in Theorem 2.2 can be improved by using the actual distribution P instead of the Markov inequality in (6) or the bound in (8). This is, for example, the case for symmetric exponential or standard Gaussian distributions but we will not pursue the subject here. In the later case, the precise behavior of $E[\log \mathcal{K}(A)]$ as $n \to \infty$ is given in [5] as

$$E[\log \mathcal{K}(A)] = \log n + C_0 + o(1),$$

where C_0 is a known constant.

•

2.5 ("Smoothed analysis") We consider now the condition number when the r.v.'s in the matrix $A = (a_{i,j})_{i,j=1,\ldots,n}$ have the form

$$a_{i,j} = \mu_{i,j} + \psi_{i,j},\ i,j = 1, ..., n,$$

where $M = (\mu_{i,j})_{i,j=1,\ldots,n}$ is non-random and $(\psi_{i,j})_{i,j=1,\ldots,n}$ are i.i.d. r.v.'s with common distribution P satisfying assumptions 1, 2 and 3 in Theorem 2.2. This -and other similar studies- have recently been called "smoothed analysis" (see [1] and [4]).

Theorem 3.1 *Under the conditions stated above, if*

$$\mu_n = \sup_{i,j=1,\ldots,n} |\mu_{i,j}| \leq n^{2/r},$$

then

$$E[\log \mathcal{K}(A)] \leq \left(1 + \frac{2}{r}\right) \log n + \log 2 + \frac{1}{r} + \frac{1}{\gamma} \left\{ \left[(2+\gamma) \log n + \log C \right]^{+} + 1 \right\} \qquad (12)$$

PROOF.- The proof (as well as the result) is very similar to that of Theorem 2.2. For $t > 0$ one has:

$$\begin{aligned} \nu\left[\|A\| > t\right] &\leq \nu\left[\sum_{i,j=1}^{n} a_{i,j}^2 > t^2\right] \leq \sum_{i,j=1}^{n} \nu\left[a_{i,j}^2 > \frac{t^2}{n^2}\right] \\ &= \sum_{i,j=1}^{n} \nu\left[|\mu_{i,j} + \psi_{i,j}| > \frac{t}{n}\right] \leq n^2 \nu\left[|\psi_{1,1}| > \frac{t}{n} - \mu_n\right]. \end{aligned}$$

Now choose $\alpha_n = \left(1 + \frac{2}{r}\right) \log n + \log 2$ and check that if $x > \alpha_n$, then

$$\frac{e^x}{n} - \mu_n > \frac{1}{2n} e^x.$$

Thus,

$$\begin{aligned} E[\log \|A\|] &\leq \alpha_n + \int_{\alpha_n}^{\infty} \nu \left[\|A\| > e^x\right] dx \\ &\leq \alpha_n + n^2 \int_{\alpha_n}^{\infty} \nu \left[|\psi_{1,1}| > \frac{1}{2n} e^x\right] dx \\ &\leq \alpha_n + n^2 \int_{\alpha_n}^{\infty} \frac{1}{\left(\frac{1}{2n} e^x\right)^r} dx \\ &= \left(1 + \frac{2}{r}\right) \log n + \log 2 + \frac{1}{r}, \end{aligned}$$

where last equality follows from a simple computation.

On the other hand, with the same notation as in the proof of Theorem 2.2, $A^{-1} = (b_{i,j})_{i,j=1,\ldots,n}$ and

$$\nu \left[\|A^{-1}\| > t\right] \leq \sum_{i,j=1}^{n} \nu \left[|b_{i,j}| > \frac{t}{n}\right].$$

For each term in this sum it is possible to repeat exactly the same computations as in the proof of Theorem 2.2 to bound $\nu \left[|b_{1,1}| > \frac{t}{n}\right]$ and obtain the same bound as there for $E[\log \|A^{-1}\|]$. This finishes the proof.

•

2.6 (Higher order moments) Is it possible to obtain upper bounds for $E\left[(\log \mathcal{K}(A))^k\right], k = 2, 3, \ldots$ much in the same way as we did for $k = 1$. We consider here the centered case, for smoothed analysis, the situation is similar.

Since $\log \mathcal{K}(A) \geq 0$ we have that

$$E\left[(\log \mathcal{K}(A))^k\right] \leq 2^k \left[E\left\{\left(\log^+ \|A\|\right)^k\right\} + E\left\{\left(\log^+ \|A^{-1}\|\right)^k\right\}\right].$$

Using the same estimates as in the case $k = 1$ for the tails of the probability distributions of $\|A\|$ and $\|A^{-1}\|$, after an elementary computation, it is possible to obtain that if $k \in I\!N$ satisfies that $2 \leq k \leq 1 + (2 + \gamma \wedge r) \log n$, then

$$E\left[(\log \mathcal{K}(A))^k\right] \leq (2 \log n)^k \left[\left(1 + \frac{2}{r}\right)^k (1 + k) + \left(1 + \frac{2}{\gamma}\right)^k (1 + Ck)\right]$$

•

References

[1] Blum, A. and Dunagan, J. (2002). Smoothed Analysis of the Perceptron Algorithm for Linear Programming. *S.O.D.A.*

[2] Castro, D.; Montaña, J.L.; Pardo, L.M. and San Martín, J. (2002), The Distribution of Condition Numbers of Rational Data of Bounded Bit Length. *Foundat. of Computat. Math.,* **2**, 1-52.

[3] Cucker, F. (2002). Real Computations with Fake Numbers. *J. of Complexity,* **18**, 104-134.

[4] Dunagan, J,; Spielman, D.A. and Teng, S. (2002). Smoothed Analysis of Renegar's Condition Number of Linear Programming. *Preprint.*

[5] Edelman, A. (1988). Eigenvalues and condition numbers of random matrices. *SIAM J. of Matrix Anal. and Applicat.*, **9**, 543-556.

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

[7] von Neuman, J. and Goldstine, H. (1947). Numerical inverting matrices of high order. *Bull. Amer. Math. Soc..* **53**, 1021-1099.