Journal of Applied Mathematics and Physics (ZAMP)
Vol. 30, 1979 Birkhäuser Verlag Basel

Isotropic Distributions of Test Matrices[1]

By Garrett Birkhoff and Surender Gulati, Dept. of Mathematics, Harvard University, Cambridge, Mass., USA

Dedicated to Eduard Stiefel

1. Introduction

During the last 30 years, there has been enormous progress in our ability to solve large linear systems. Although this progress has been mainly due to hardware improvements, it has been helped substantially by advances in software design and theoretical understanding, by Wilkinson ([9], [10], [11]) and others.

However, it is still difficult to estimate reliably and accurately the *error* $e = x_c - x$ and the *relative error* (R.E.) $\|e\|/\|x\|$ of the computed solution x_c of $Ax = b$, where A is a given nonsingular matrix and b a given vector. It is easy to compute the *residual* $r = Ax_c - b$ and the *relative residual* (R.R.) $\|r\|/\|b\|$. But unfortunately, the R.R. is very weakly correlated with the R.E.; in general, we only know that, in any norm

$$1/\kappa(A) \leq (\text{R.E.})/(\text{R.R.}) \leq \kappa(A), \quad \text{where} \quad \kappa(A) = \|A\| \cdot \|A^{-1}\|. \tag{1.1}$$

This is true because

$$\frac{\text{R.E.}}{\text{R.R.}} = \left(\frac{\|e\|}{\|x\|}\right) \Big/ \left(\frac{\|r\|}{\|b\|}\right) = \left(\frac{\|b\|}{\|x\|}\right) \Big/ \left(\frac{\|r\|}{\|e\|}\right), \tag{1.2}$$

where $b = Ax$ and $r = Ax - Ax_c = A(x - x_c)$, which implies $1/\|A^{-1}\| \leq \|b\|/\|x\|$, $\|e\|/\|r\| \leq \|A\|$; this makes the R.R. unreliable as an error diagnostic.

The IRE. In [3], we proposed the following, relatively inexpensive and (in our experience) very reliable way of estimating the R.E. First form the residual $r = Ax_c - b$ *in the same precision used to compute* x_c. Next, define the *indicated error* e_c as the computed solution of $Ay = r$; this costs little if the factors of the LU-decomposition used to 'solve' $Ax = b$ are stored. Finally, compute the *indicated relative error* (IRE) of the computed solution, defined as the ratio

$$\text{IRE} = \|e_c\|/\|x_c\|. \tag{1.3}$$

In our experience, this has the same order of magnitude as the true relative error.

[1]) The main results of this paper were reported in [3], and communicated in 1976 to those working on the LINPACK project.

An alternative *a priori* bound on the residual can be based on *Wilkinson's inequality* [10, (64.3)]:

$$\|r\| \leq f(n)\|A\| \cdot \|x_c\|\epsilon, \quad (A: \quad n \times n), \tag{1.4}$$

Here $\epsilon = 2^{-t}$ is the working precision ('macheps'), and $f(n)$ is a *slowly growing* function of n, depending on the norm used. Wilkinson states that $f(n) \leq 3n$ if the max norm $\| \ \|_\infty$ is used. This *a priori* bound on the residual implies

$$\text{R.R.} \leq 3n\|A\| \cdot \|x_c\|/\|b\| \tag{1.4'}$$

as a corollary. Since $Ae = A(x_c - x) = b - Ax = -r$, (1.4) also implies as another corollary

$$\|e\| \leq \|A^{-1}\| \cdot \|r\| = f(n)\kappa(A)\epsilon\|x_c\|, \tag{1.5}$$

Hence (1.4) has the third corollary

$$\text{R.E.} \simeq \|e\|/\|x_c\| \leq f(n)\kappa(A)\epsilon. \tag{1.5'}$$

Clearly, in order to use (1.5') as a R.E. diagnostic, one needs an estimate of $\kappa(A)$.

RCOND.[2]) A new algorithm called RCOND gives a reliable way of estimating $\kappa(A)$ in $O(n^2)$ operations. The method involves first constructing from A an appropriate vector b with entries ± 1, then solving $U_c^T(L_c^T z) = b$ and $L_c U_c = z$, and finally setting $\text{RCOND} = \|A\| \cdot \|y\|/\|z\|$. Finally, combining (1.5') with RCOND, we obtain the order-of-magnitude estimate

$$\text{R.E.} = O(\epsilon \cdot \text{RCOND}). \tag{1.6}$$

Supporting (1.6) is the fact that all numerical experiments to date are compatible with the inequality $\text{R.E.} \leq 10\epsilon \cdot \text{RCOND}$.

RCOND is being incorporated into LINPACK as its main error diagnostic (for LU-decomposition).

Singular value decomposition. To understand error diagnostics in depth, and in particular the dependence of R.E./R.R. on b, one must have clearly in mind the *singular value decomposition* (SVD) of A, which we briefly recall next.[3])

The SVD factors any nonsingular A as $A = PDQ$, where P and Q are orthogonal and D is positive diagonal. The diagonal entries of D, which we can suppose arranged in descending order $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n > 0$, are the *singular values* of A. They are also the stationary values of the ratio $\|Ax\|/\|x\|$, and so (equivalently) of its square

$$\|Ax\|^2/\|x\|^2 = (x^T A^T Ax)/(x^T x). \tag{1.7}$$

It follows that the σ_k^2 are the *eigenvalues* of $A^T A$ (and of AA^T).

[2]) A. K. Cline, C. B. Moler, G. W. Stewart, and J. H. Wilkinson, Argonne Internal Report TM-310, July, 1977.

[3]) See Contribution I/10 in [11], by G. H. Golub and C. Reinsch; also [10, §§52–56].

Again, for any vector norm $\| \ \|_v$, the *condition number* of A is by definition

$$\begin{aligned}\kappa_v(A) &= \|A\|_v \cdot \|A^{-1}\|_v \\ &\equiv \max_{x \neq 0} (\|Ax\|_v / \|x\|_v) / \min_{y \neq 0} (\|Ay\|_v / \|y_v\|) \\ &= \max_{x,y \neq 0} (\|Ax\|_v \cdot \|y\|_v / \|Ay\|_v \cdot \|x\|_v). \end{aligned} \tag{1.8}$$

In particular, for $\|x\| = (\sum x_k^2)^{1/2}$ the *Euclidean* norm, therefore, $\kappa_v(A) = \kappa(A) = \sigma_1/\sigma_n$. For this and other reasons, we will use the Euclidean norm and condition number below unless otherwise specified.

The singular values of A being the stationary values of $\|Ax\|/\|x\|$, the corresponding singular domain vectors $x = \varphi_j$ are the rows of Q (columns $\varphi_j = Q^T e_j$ of Q^T) since $PDQ(Q^T e_j) = P\sigma_j e_j$; their images $\psi_j = P\sigma_j e_j$ (the 'singular' *range* vectors) are also orthogonal. Conversely, we have

Proposition 1: If A carries an orthogonal basis of vectors φ_j into an orthogonal basis of vectors ψ_j, then the φ_j are a basis of singular domain vectors of A, and the ψ_j a basis of singular range vectors.

2. The BMvN Distribution

Studies of *individual* linear systems can only lead to *qualitative* empirical conclusions about the comparative efficiency and reliability of numerical methods and error diagnostics. This is because their applicability to other systems is largely conjectural.

Quantitative conclusions must necessarily be based on *probabilistic* estimates and statistical analyses referring to large families of matrices. Hopefully, these can be related through *a priori* estimates to *performance profiles* of methods in actual tests covering a wide range of sample matrices. And indeed, such statistical analyses have recently begun to replace discussions of individual cases in the technical literature on the subject.[4])

Actually, this trend was foreshadowed in a fascinating 1946 report to the U.S. Navy, by von Neumann, V. Bargmann, and Deane Montgomery, reprinted in [7, pp. 421–478]. Here an attempt was made to predict the 'number of extra digits' needed to 'invert matrices' associated with linear systems of orders '$n = 20, 50, 100$ or even larger'. At the end of § 13 of this report, the authors consider *random* matrices whose entries have *identical* independent normal random distributions, and refer (see [7, p. 460]) to 'a statistical analysis' published later in the Proc. AMS 2 (1951), 188–202. We will call the *a priori* distribution of random matrices so defined the *BMvN distribution.*

The authors named above were interested in this distribution 'primarily for estimating probabilistically the bounds of matrices and norms of vectors whose

[4]) See for example J. T. Goodman and C. B. Moler, 'LINPACK working note #8,' Argonne AMD TM-311, July, 1977.

elements are sums of one or more rounding errors'. However, it also has some very attractive, purely mathematical properties, which we will now try to summarize.

Theorem 1. *For any n, the* BMvN *probability measure μ_B on the space of all $n \times n$ matrices A is invariant under the biorthogonal group $A \mapsto PAQ$, P, Q orthogonal $n \times n$ matrices.*

Proof: The probability density of each column vector of A is

$$2^n \pi^{-n/2} e^{-r^2} = \prod_{i=1}^{n} (2/\sqrt{\pi})\, e^{-x_i^2}.$$

Since this is spherically symmetric, and the independence of the random column vectors is preserved under orthogonal transformation, μ_B is invariant under $A \mapsto AQ$. But μ_B is obviously invariant under $A \mapsto A^T$, since its definition is; hence μ_B is also invariant under orthogonal transformations $A \mapsto PA$ of the random row vectors, completing the proof.

But now, the singular value decomposition selects from each equivalence class of $n \times n$ matrices under the group $A \mapsto PAQ$ a unique *diagonal* matrix D with nonincreasing nonnegative values. Hence standard arguments about measures on compact groups and their products give

Theorem 2. *The most general biorthogonally invariant measure on the space GL_n of nonsingular $n \times n$ matrices is obtained by setting up a measure μ_d on the space of sequences $(d_1, \ldots, d_n)$ of nonincreasing nonnegative numbers d_i, and forming its product $\mu_0 \times \mu_d \times \mu_0$ with two copies of the group measure μ_0 of the orthogonal group, on the space of matrices $A = PDQ$.*

Similar results hold for *symmetric* matrices whose diagonal and lower triangular entries have IID (independent identically distributed) normal distributions (these matrices, of course, form a Jordan algebra but not a group). The set S_n of all such matrices is invariant under $A \mapsto PAQ$ if and only if $Q = P^T = P^{-1}$; hence the appropriate symmetry condition is *orthogonal* invariance under $A \mapsto PAP^{-1}$, and not biorthogonal invariance. We have

Theorem 3. *The most general orthogonally invariant measure on the space S_n of* $n \times n$ symmetric *matrices is obtained by setting up measures μ_0 and μ_d as in Theorem 2, and forming their product $\mu_0 \times \mu_d$ on the space of symmetric matrices PDP^{-1}.*

Corollary. *The* BMvN *distribution is characterized by its biorthogonal invariance and the probability distribution μ_d of its spectrum of singular values $\sigma_j = d_j$; the analogous distribution of random symmetric matrices is characterized by its orthogonal invariance and the probability distribution μ_d' of its spectrum of eigenvalues.*[5])

[5]) We recall that the singular values of a symmetric matrix are its eigenvalues.

3. First Numerical Experiments

Since von Neumann and his collaborators had already expended considerable effort in deducing the properties of random matrices having the BMvN distribution from first principles, we decided to make extensive experimental tests before doing anything else.

With advice from David Hoaglin about the best choice of random number generator, and extensive programming help from W. G. Ledsham, we solved $Ax = b$ for *general real* and for *symmetric indefinite random matrices* A having normal IID entries, and entries of b constructed in the same way. In both cases, we tested one thousand 20×20 matrices, two hundred 40×40 matrices, and fifty 80×80 matrices. These tests were made in Nov.–Dec., 1974; they were performed in some haste, because we wanted to report our findings at the 'Gatlinburg' VI meeting to be held near Munich that December.

Specifically, we made statistical tabulations of the R.E. made in solving $Ax = A\mathbf{1}$, where $A = A(\omega)$ was a random matrix[6]) taken from a BMvN distribution and $\mathbf{1} = (1, \ldots, 1)^T$. We produced random a_{ij} using the pseudo-random number generator LLRANDOM, a package produced at the Naval Postgraduate School by Drs. Learmouth and Lewis. Its basic element is a pure congruential generator with

$$c_n \equiv 7^5 c_{n-1} \pmod{2^{31} - 1}. \tag{3.1}$$

The output is further randomized by using the seven least significant bits of the integer produced by (3.1) to access a table, and then setting $a_{ij} = f(c_n)$. We used the initial seed 216002115.

Our initial objective was in large part to compare the accuracies and execution times of solutions of linear systems obtained using programs from the carefully designed SL-MATH, IMSL, and LINSYS[7]) packages. We found only minor differences, which we reported at the Gatlinburg VI meeting.

Most striking was the fact that *the R.E. was very small* for all of the matrices tested, exceeding 1000ϵ in only one case in 1000. It was very much less than what had been estimated *a priori* by von Neumann and his co-workers in [7, (7.16)], where the estimate $2000n^4\epsilon$ was given. (We tried $n = 20$, 40, 80, and 160.)

Very reassuring was the good correlation (to within a factor between 0.1 and 10) of the R.E. with $\kappa(A)$,[8]) regardless of n over the range $20 \leq n \leq 160$. This pretty much laid to rest the idea that the random rounding errors would be greatly amplified during the elimination process. Correlation of the R.E. with the first Hadamard number $\Pi \|A_i\| / |A|$, where A_i is the ith column of A, was much less good.

[6]) We use here the standard notation: ω denotes a sample element from a probability space Ω endowed with a measure μ. Besides the BMvN probability measure $R^{n \times n}$, we let the a_{ij} be IID variables having a uniform distribution on $[-1, 1]$.

[7]) The LINSYS package was developed at the University of Victoria Computing Center, using Fortran versions of the Algol programs in [11, Part I].

[8]) $\kappa(A)$ was computed using EISPACK.

Finally, we found that the distribution of $\kappa(A)$ was largely independent of n, and that the occurrence of very large $\kappa(A)$ was quite sporadic, the frequency being roughly proportional to $1/\kappa(A)$.

In the spring of 1976 we made a second series of tests, this time on $Ax = b$, where b was a random vector having an 'isotropic' (i.e., spherically symmetric) normal distribution, produced by LLRANDOM with the seed 145365181. We determined the R.E. by iterative refinement (using IBM quadruple precision in accumulating inner products). We found that the R.E. correlated quite well (to within a factor ordinarily between 0.3 and 3.0) with both the product $\kappa(A) \times$ R.R. and with the IRE. As we anticipated that more careful and systematic experiments would be reported in [4], we did not pursue these experiments further.

4. Theoretical Analysis

Having satisfied ourselves that the errors were primarily associated with the size of $\kappa(A)$, we set ourselves the task of *predicting*[9]) the distribution of $\kappa(A)$—and more generally of the singular values of A—for a BMvN distribution of random matrices. Our most important observations turned out to be fairly easy to explain.

First, we note two results which we state as lemmas without proof.

Lemma 1. *If A is an $m \times n$ matrix with $m \leq n$, and the spectrum of AA^T is $\sigma_1^2, \ldots, \sigma_m^2 = \lambda_1, \ldots, \lambda_m$, then the spectrum of A^TA is $\lambda_1, \ldots, \lambda_m, 0, \ldots, 0$.*

Lemma 2. (*Monotonicity*). *If $\overline{A}_m$ is the $m \times n$ matrix consisting of the first m columns of an $n \times n$ matrix A, and $\sigma_1' \geq \cdots \geq \sigma_m'$ are the singular values of $\overline{A}_m$, then the singular values $\sigma_1'' \geq \cdots \geq \sigma_{m+1}''$ of $\overline{A}_{m+1}$ satisfy*

$$\sigma_1'' \geq \sigma_1' \geq \sigma_2'' \geq \sigma_2' \geq \cdots \geq \sigma_m'' \geq \sigma_m' \geq \sigma_{m+1}''. \tag{4.1}$$

We then apply these lemmas to the QR factorization of A into an orthonormal matrix Q and upper ('right') triangular matrix R with positive diagonal entries d_m [8, p. 214]. Geometrically, if S_m denotes the subspace spanned by the first m columns $A_1, \ldots, A_m$ of A, then d_m is the length of the component of A_m orthogonal to S_m, while each r_{km} $(k < m)$ is the component of A_m in the direction of the kth component of the Gram–Schmidt orthogonalization of the sequence of column vectors $A_1, \ldots, A_k$.

Theorem 4. *The factors Q and R of the QR factorization of a random matrix $A = QR$ having the* BMvN *distribution have independent probability distributions μ_Q and μ_R. Here μ_Q is the orthogonal group measure on the space of orthogonal matrices, while μ_R assigns independent probability measures to the upper triangular entries of R.*

Proof: The BMvN distribution constructs[10]) a square matrix A in which each column (or row!) vector A_m has an isotropic (= spherically symmetric) normal

[9]) For *a priori* error estimates on solutions of $Ax = b$ for A having a BMvN distribution, see also H. P. Mulholland, Proc. AMS *3* (1952), 310–321.

[10]) To within the usual approximation of rounding off each 'random number' to make it machine representable.

distribution, independent of the other A_k, $k \neq m$. Hence its projection onto S_m has an isotropic normal distribution 'circularly' invariant under orthogonal transformations of S_m. Hence the lengths r_{km} ($k < m$) of its projections onto the first $m - 1$ orthogonalized axes are independent IID normal variables with mean 0 and standard deviation 1.

Moreover, since *orthogonal* components of isotropic normal distributions are also independent, the length $r_{mm} = d_m$ of the component of A_m in $S_m^{\perp}$ is that of a random vector with normally distributed IID components in $S_m^{\perp}$; hence, d_m is a positive random number with probability density proportional to $r^{n-m}e^{-r^2}$ at $d_m = r$, and $\overline{d_m{}^2} = n - m + 1$, on normalized scales.

It follows from the monotonicity lemmas that the largest singular value σ_1 of R is at least d_1, while the smallest singular value σ_n is at most d_n. Since $\kappa(A) = \kappa(R)$, it follows that $\kappa(A) \geq d_1/d_n$.

We now come to the all-important question: what is the probability that $\kappa(A) > M = 1/\eta$, a preassigned constant? Since $\sigma_i \simeq \sqrt{n}$ with a very large probability for large n, this is the probability that $\sigma_n < \eta/\sqrt{n}$. It is $O(\eta)$ and exceeds

$$\text{Prob}\{d_n < \eta/\sqrt{n}\} \simeq (2\pi/n)^{1/2}\eta.$$

Also, the 2×2 submatrix at the lower right corner of A is just

$$\begin{pmatrix} d_{n-1} & c_{n-1,n} \\ 0 & d_n \end{pmatrix},$$

where: (i) The nonzero entries are *independent* normal random variables, (ii) $c_{n-1,n}$ is normal with mean 0 and S.D. 1, while (iii) d_{n-1} and d_n are positive with probability densities proportional to $re^{-r^2}dr$ and e^{-r^2}/r, respectively. More specifically, it is the matrix C_2 obtained by Gram–Schmidt triangularization from the 2×2 matrix of components of A_{n-1} and A_n orthogonal to the subspace S_{n-2} spanned by $A_1, \ldots, A_{n-1}$. This has the same probability distribution of singular values as the 2×2 random matrix whose entries are independently normally distributed with mean 0 and standard deviation 1. Most important, the probability that both $d_{n-1} < \eta$ and $d_n < \eta$ is very small, being $O(\eta^3)$. Hence the probability that $d_{n-2} < \epsilon$ is $O(n^{-3/2})$, and the inverse of a large nearly singular matrix from a BMvN distribution will almost certainly be nearly of rank one.

5. Isotropic Distributions of Ill-Conditioned Matrices

As a test of the robustness of programs for solving $Ax = b$, the BMvN distribution turned out to be wasteful because its random matrices are so unlikely to be ill-conditioned. However, Theorems 2 and 3 above make it easy to generate both random matrices and random symmetric matrices whose probability distributions have the same 'isotropic' character, yet whose matrices A are as ill-conditioned as is desired. One can do this for general real matrices by the following procedure.

Step 1. Use random number generators to construct sequences $d = (d_1, \ldots, d_n)$ of positive nonincreasing real numbers $d_1 \geq d_2 \geq d_3 \geq \cdots \geq d_n$, $d_n > 0$, having any desired distribution. For example, one can take $d_1 = 1$ and $d_{k+1} = r_k d_k$, where r_k is independently uniformly distributed between 0 and 1. With $\epsilon \simeq 10^{-15}$, as with CDC single or IBM double precision, the range $10^{10} < \kappa(A) = d_1/d_n < 10^{15}$ is most interesting. One can achieve this range by giving the d_k a roughly uniform logarithmic distribution over the range $0 > \log_{10} d > -n$, where n is (say) 10, 12, or 14.

Step 2. Form independent *random orthogonal matrices* P and Q by taking two random matrices B and C from a BMvN distribution, and letting $B = PR_1$ and $C = QR_2$ be their QR factorizations, as obtained by Householder transformations using the Businger–Golub procedure [11, I/8].[11])

Step 3. Let $A = PDQ$, where D is the random diagonal matrix with the diagonal entries $d_1, \ldots, d_n$ computed in Step 1.

To construct random *symmetric* matrices having an 'isotropic' distribution, one proceeds similarly, but only constructs $B = PR_1$ and forms $A = PDP^{-1}$.

An attractive feature of the above procedure is that it can be easily modified so as to give random samples from *any* biorthogonally invariant distribution of general real or orthogonally invariant distribution of symmetric random matrices. Of course, it is quite expensive.

To reduce the cost, one can construct clumps of 8 random orthogonal matrices with probability distribution $\mu_0 \times \cdots \times \mu_0 = \mu_0^8$, as follows. First, construct 8 random matrices A_k from the BMvN distribution, next factor each $A_k = Q_k R_k$ by a QR-transformation. Then form the eight $Q_k D Q_k^T$; they will be eight random symmetric matrices having independent isotropic distributions.

As far as *general real* matrices are concerned, define $l(1) = 5$, $l(2) = 6$, $l(3) = 8$, $l(4) = 2$, $l(5) = 7$, $l(6) = 1$, $l(7) = 3$, $l(8) = 4$, thus making the $(k, l(k))$ locations of eight queens on a chessboard so that none can take any other. Although pairs of matrices like $Q_1 D Q_4^T$ and $Q_4 D Q_8^T$ do not have independent probability distributions, none is the transpose of any other and no two have a common left- or right-factor. Hence it would seem to be worth testing all eight.

For given $A = PDQ = Q_i D Q_j$, one can of course let b be an independent random isotropic vector. Another interesting choice is to let $b_1 = Q^T e_1$ be in the direction of the least singular and $b_2 = Q^T e_n$ in the direction of the most singular range vector, to see the extreme values assumed by the ratio R.E./R.R. in these two cases.

Another interesting series of tests consists in *inverting* A_j. The contrast between the ratio R.E./R.R. for the *matrix* equations $AX = I$ (with error $(A^{-1})_c - A^{-1}$) and $AX = A$ (with error $X_c - I$) when A is ill-conditioned is interesting. So is the distribution of the differences $((A^{-1})_c^{-1})_c - A$, as regards R.E. and R.R. In this connection, we note that A (when ill-conditioned) has a *very* different distribution of

[11]) Alternatively, one can use the LINPACK subroutine SQRSL (which invokes STRSL).

singular values in the BMvN distribution than A^{-1} does. This is because A has (with *very* high probability) just *one* small singular value, whence A^{-1} has just one *large* singular value.

Note that biorthogonally invariant probability distributions of matrices, also invariant under $A \mapsto A^{-1}$, can be constructed by random number generators which give the log d_i IID values *symmetrically* distributed with mean 0. For example, these could be uniformly distributed on $[-l_i, l_i]$ *or* normally distributed with mean square l_i.

Hadamard matrices. Finally, it would be interesting to see whether, for a given D, the process of *random biorthogonal shuffling* (by setting $A = PDQ$) proposed above gives noticeably different results than the (much cheaper) process of scrambling by a symmetric *Hadamard matrix*, i.e., setting $A = HDH^T$. For example, if $n = 20$, we can define H_{20} by the following condition:[12])

$$\left.\begin{aligned} &h_{ij} = h_{ji} = 1 \quad \text{for all} \quad i > 1 \quad \text{and} \quad j > 1 \\ &\text{such that} \quad i + j = 4, 5, 8, 9, 10, 11, 13, 15, 20, 21, 23, 24, 27, 28, 29, \\ &\qquad\qquad 30, 32, 34, 39, 40. \\ &h_{ij} = h_{ji} = 0 \quad \text{otherwise.} \end{aligned}\right\} \tag{5.1}$$

Having constructed H_{20}, one can construct Hadamard matrices of orders 40, 80, and 160 by repeated application of the prescription

$$H_{2n} = H_2 \otimes H_n = \begin{bmatrix} H_n & H_n \\ H_n & -H_n \end{bmatrix}, \tag{5.2}$$

where $\otimes$ stands for the Kronecker product.

6. Third Set of Tests

A third set of tests, based on an implementation of the scheme described in §5, was carried out and reported at length in [4]. We will here only summarize very briefly some of the main conclusions from these tests.

The ratio of the R.E./R.R. was found to be indeed dependent on b. For b a random isotropic vector, the R.E. had typically the same order of magnitude as the R.R. in every case tried. For b a singular range vector, the ratio depended on the associated singular value σ_i, *qualitatively* in essentially the way predicted in §2. However, whereas $(\text{R.E./R.R.})_{\max}$ was essentially proportional to $\kappa^{0.9}$, $(\text{R.E./R.R.})_{\min}$ was only proportional to $\kappa^{-0.15}$ or $\kappa^{-0.2}$.

Both the R.E. and the R.R. grew nearly linearly with n, although the R.R.

[12]) S. W. Golomb and L. D. Baumert, Am. Math. Monthly *70*, 12–17 (1963).

grew somewhat less rapidly. Perhaps unfortunately, no attempt was made to find the α_j, β_j, and γ_j which gave a least squares fit of the data to

$$\log \text{R.R.} = A_1 \log \kappa + B_1 \log n + \log \epsilon + \gamma_1,$$

$$\log \text{R.E.} = A_2 \log \kappa + B_2 \log n + \log \epsilon + \gamma_2.$$

Tests were made of the reliability of COND as a method of estimating $\kappa(A)$, and of the reliability of IRE as a way of estimating the R.E. The IRE was an especially good diagnostic when $b = \varphi_n$, the most singular range vector. In general, it was found that both were reliable up to a factor of 2 or 3 in most cases.[13]) A new QRPiv scheme (first proposed in [4]) seemed to correlate with $\kappa(A)$ slightly better than RCOND (in the tests run). Unfortunately, no effort was made in [4] to correlate RCOND with the R.E.

The operation counts and execution times were roughly proportional to $0.6n^3$ for LU, n^3 for QR, and $9n^3$ for SVD. The errors of LU were consistently only about half as big as the errors of QR.

Finally, confirming the statement in Forsythe [5, §10], it was found that scaling almost never reduced the error by more than 25%, and in fact sometimes increased it.

References

[1] T. J. Aird and R. E. Lynch, *Computable Accurate ... Error Bounds for ... Solutions of Linear Algebraic Systems*, ACM Trans. Math. Software *1*, 217–231 (1975).

[2] G. Birkhoff, *Two Hadamard Numbers for Matrices*, Comm. ACM *18*, 25–29 (1975).

[3] G. Birkhoff and S. Gulati, Preliminary Tests for a LINPACK, Harvard NSF Report dated June, 1976.

[4] D. E. Coleman, 'Solving Linear Systems: a Statistical Approach,' Harvard A.B. Thesis, 1977.

[5] G. E. Forsythe, *Today's Computational Methods of Linear Algebra*, SIAM Rev. *9*, 489–515, (1967).

[6] G. E. Forsythe and Cleve B. Moler, *Computer Solution of Linear Algebraic Systems*, Prentice-Hall, Englewood Cliffs (1967).

[7] John von Neumann, *Collected Works*, vol. V, Pergamon, Oxford (1963).

[8] G. W. Stewart, *Introduction to Matrix Computations*, Academic Press, New York (1973).

[9] J. W. Wilkinson, *Rounding Errors in Algebraic Processes*, Prentice-Hall, Englewood Cliffs (1963).

[10] J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, Oxford University Press (1965).

[11] J. H. Wilkinson and C. Reinsch, *Linear Algebra*, Springer, Berlin (1971).

Abstract

In 1946, von Neumann and his collaborators used a special distribution of random matrices as a model for estimating *a priori* the machine precision needed to solve large linear systems. The present paper identifies *isotropy* as a group-theoretic property of this distribution, shows that its matrices are almost never ill-conditioned, and explains how to use other isotropically distributed random matrices for testing the accuracy of numerical methods for solving linear systems and associated error diagnostics.

[13]) Similar results about RCOND (which estimates the 'invertibility' $1/\kappa(A)$ of A), were reported in internal reports TM-310 and TM-311 (using DGECO) by J. T. Goodman and Cleve Moler op. cit.

Zusammenfassung

Die zur Lösung linearer Gleichungssysteme benötigte Genauigkeit wurde schon 1946 durch von Neumann und seine Mitarbeiter mittels speziell verteilter Zufallsmatrizen geschätzt. In der vorliegenden Arbeit erscheint die *Isotropie* als gruppentheoretische Eigenschaft dieser Verteilung. Ferner wird gezeigt, dass die Zufallsmatrizen fast nie schlecht konditioniert sind. Schliesslich diskutieren die Autoren die Verwendung anderer isotrop verteilter Zufallsmatrizen zur Prüfung von Genauigkeit und a-priori-Fehlerschranken bei Algorithmen zur Lösung linearer Gleichungssysteme.

(Received: October 18, 1978)