Journal of Research of the National Bureau of Standards Vol. 60, No. 1, January 1958 Research Paper 2815

The Condition of Certain Matrices, III [1]

John Todd [2]

The condition-numbers of certain matrices associated with various discretizations of the two-dimensional Laplacian operator are estimated. The condition-number gives an estimate of the error obtained by solving the corresponding systems of simultaneous linear equations.

1. *Introduction.* In a previous paper [Todd, 8a] [3] the condition of a matrix associated with a particular method of solving a simple partial differential equation was discussed. Certain experimental computations of D. M. Young and his collaborators [Young, 13] suggested the discussion of other methods of handling the same equation. The P-condition number of a matrix M is defined as λ/μ where λ is the maximum and μ is the minimum of the absolute values of the eigenvalues of M; it gives a measure of the difficulty in the numerical inversion of M, or, more precisely, of an error in the inverse computed by an elimination method [von Neumann and Goldstine, 11; Todd, 8a, 9].

In [8a] we considered the solution of the Laplace equation in a unit square, with given boundary values. We then used the following five-point approximation to the Laplace operator:

$$z_{xx}+z_{yy} \doteq h^{-2}[z(r-1,s)+z(r,s-1)-4z(r,s)+z(r,s+1)+z(r+1,s)], \tag{1.1}$$

where $z(r,s)=z(rh,sh)$ and $h=1/(n+1)$, n a positive integer. Following Milne [4, p. 131] we indicate this approximation by the "stencil"

$$\left\{\begin{matrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{matrix}\right\}.$$

(This has been called a "computational molecule" by Bickley [2]). The $n^2\times n^2$ matrix which corresponds to the equations for the $z(r,s)$ can be indicated as

$$Z_1=\begin{pmatrix} & \cdots & \\ \cdots, & I,X,I, & \cdots \\ & \cdots & \end{pmatrix} \quad \text{where} \quad X=\begin{pmatrix} & \cdots & \\ \cdots, & 1,-4,1, & \cdots \\ & \cdots & \end{pmatrix}$$

is an $n\times n$ matrix and I is the $n\times n$ unit matrix. Here, and elsewhere, for brevity, we have indicated only the general row in the triple diagonal matrices. For clarity we describe Z_1 and X in words. The matrix Z_1 is an $n^2\times n^2$ matrix partitioned into $n\times n$ blocks; of these the diagonal blocks are all X and the blocks adjacent to the diagonal are unit matrices. The matrix X is an $n\times n$ matrix with diagonal elements all -4 and the elements adjacent to the diagonal all 1: In symbols $X=(x_{ij})$ where $x_{ii}=-4$, $i=1,2, \ldots n$, and for $i\neq j$, $x_{ij}=0$ unless $|i-j|=1$, when $x_{ij}=1$.

The equations corresponding to the problem

$$z_{xx}+z_{yy}=0 \text{ in interior,} \qquad z=f \text{ on boundary}$$

can be written as

$$Z_1\mathbf{z}=\mathbf{f}$$

---

[1] The problems considered here were suggested by lectures of D. M. Young, Jr., in the National Bureau of Standards-National Science Foundation Training Program in Numerical Analysis held in 1957.

[2] Present address, California Institute of Technology, Pasadena 4, Calif.

[3] Figures in brackets indicate the literature references at the end of this paper.

where the variables $z(r,s)$ are combined as the n^2-dimensional vector:

$$\mathbf{z}=(z(1,n),\ z(2,n),\ \ldots,\ z(n,n);\ z(1,n-1),\ z(2,n-1),\ \ldots,\ z(n,n-1);\ \ldots;\ z(1,1),\ z(2,1),\ \ldots,\ z(n,1))$$

and where

$$-\mathbf{f}=(f(0,n)+f(1,n+1), f(2,n+1),\ \ldots,\ f(n,n+1)+f(n+1,n);\ f(0,n-1), 0, \ldots, f(n+1,n-1);\ \ldots;\ f(0,1)+f(1,0), f(2,0),\ \ldots,\ f(n,0)+f(n+1,1)).$$

The essential result established in [8] was that

$$P(Z_1) \doteq 4\pi^{-2}n^2.$$

We shall now obtain the corresponding results for other discretizations of the Laplacian [Panow 5, Bickley 2].

2. *Auxiliary results.* We require the following results:

(2.1) *The characteristic values of the triple diagonal matrix of order n:*

$$C=\begin{pmatrix} & \cdots & \\ \ldots, & c,\ a,\ b, & \ldots \\ & \cdots & \end{pmatrix}$$

are

$$\lambda_k=a-2\sqrt{(bc)}\cos k\theta, \qquad k=1,2,\ \ldots,n$$

where

$$\theta=\pi/(n+1).$$

This is well known; it is easily proved by solving the difference equation

$$\Delta_n(\lambda)=(a-\lambda)\Delta_{n-1}(\lambda)-bc\Delta_{n-2}(\lambda)$$

for the characteristic polynomial $\Delta_n(\lambda)=|C-\lambda I|$.

We shall use the notation $\theta=\pi/(n+1)$ throughout this paper.

(2.2) *The characteristic roots of the quintuple diagonal matrix of order n:*

$$A=\begin{pmatrix} z-\beta, & 2\alpha, & \beta & & & & & \\ 2\alpha, & z, & 2\alpha, & \beta & & & & \\ \beta, & 2\alpha, & z, & 2\alpha, & \beta & & & \\ & \beta, & 2\alpha, & z, & 2\alpha, & \beta & & \\ & & & \ddots & & & & \\ & & & \beta, & 2\alpha, & z, & 2\alpha, & \beta \\ & & & & \beta, & 2\alpha, & z, & 2\alpha \\ & & & & & \beta, & 2\alpha, & z-\beta \end{pmatrix}$$

are

$$\lambda_k=z-2\beta-\beta^{-1}\{\alpha^2-(\alpha-2\beta\cos k\theta)^2\}, \qquad k=1,2,\ \ldots,n.$$

This result has been given by Rutherford [6, 7]. It can be verified as follows: If in the matrix C we put $c=b$ and square we obtain essentially the matrix A. In fact if we choose $b=\sqrt{\beta}$, $a=\alpha/\sqrt{\beta}$, then

$$A=C^2+(z-\alpha^2\beta^{-1}-2\beta)I.$$

(2.3) *Let F be a matrix of order n^2, partitioned into an array of n^2 submatrices f_{ij}, each of order n, such that each f_{ij} is a rational function f_{ij} ($\mathbf{a}$) of a fixed matrix $\mathbf{a}_2$ of order n. If the characteristic values of $\mathbf{a}$ are $\alpha_1, \alpha_2, \ldots, \alpha_n$ then those of F are given by the characteristic values of the n matrices*

$$(f_{ij}(\alpha_k)), \qquad k=1,2, \ldots, n,$$

each of order n.

This has been established by Williamson [12]; for extensions see Afriat [1].

3. We begin by discussing another five-point approximation which is represented by the stencil

$$\frac{1}{2}\begin{Bmatrix} 1 & 0 & 1 \\ 0 & -4 & 0 \\ 1 & 0 & 1 \end{Bmatrix}.$$

Omitting the factor $\frac{1}{2}$, we see that the corresponding $n^2\times n^2$ matrix is

$$Z_2=\begin{pmatrix} & \cdots & \\ \ldots, X, & -4I, & X, \ldots \\ & \cdots & \end{pmatrix},$$

where

$$X=\begin{pmatrix} & \cdots & \\ \ldots, 1, & 0, & 1, \ldots \\ & \cdots & \end{pmatrix}$$

is an $n\times n$ matrix and I the $n\times n$ unit matrix. The characteristic values of X are, from (2.1),

$$-2\cos k\theta, \qquad k=1,2, \ldots, n.$$

It follows that those of Z_2 are those of the n triple diagonal matrices

$$Z_2^{(k)}=\begin{pmatrix} & \cdots & \\ \ldots, -2\cos k\theta, & -4, & -2\cos k\theta, \ldots \\ & \cdots & \end{pmatrix}, \qquad k=1,2, \ldots, n,$$

which are, from (2.1),

$$\nu_{k,l}=-4[1+\cos k\theta\cos l\theta], \qquad k=1,2, \ldots, n, \qquad l=1,2, \ldots, n.$$

We have

$$\lambda(Z_2)=4[1+\cos^2\theta]$$

$$\mu(Z_2)=4\sin^2\theta$$

so that

$$P(Z_2)\doteq 2\pi^{-2}n^2.$$

4. We next discuss the nine-point approximation represented by the stencil

$$\begin{Bmatrix} 1 & 4 & 1 \\ 4 & -20 & 4 \\ 1 & 4 & 1 \end{Bmatrix}.$$

Omitting the factor $\frac{1}{6}$, we see that the corresponding $n^2 \times n^2$ matrix is

$$Z_3 = \begin{pmatrix} & \cdots & \\ \cdots, & Y, \; X, \; Y, & \cdots \\ & \cdots & \end{pmatrix}$$

where

$$X = \begin{pmatrix} & \cdots & \\ \cdots, & 4, \; -20, \; 4, & \cdots \\ & \cdots & \end{pmatrix} \quad \text{and} \quad Y = \begin{pmatrix} & \cdots & \\ \cdots, & 1, \; 4, \; 1, & \cdots \\ & \cdots & \end{pmatrix}.$$

We note that (2.3) will apply since

$$Y = \frac{1}{4} X + 9I.$$

The characteristic values of X are

$$-20 - 8 \cos k\theta, \qquad k = 1, 2, \ldots, n.$$

It follows that the characteristic values of Z_3 are those of the n matrices

$$Z_3^{(k)} = \begin{pmatrix} & \cdots & \\ \cdots, & 4 - 2\cos k\theta, \; -20 - 8 \cos k\theta, \; 4 - 2 \cos k\theta, & \cdots \\ & \cdots & \end{pmatrix}$$

which are

$$\nu_{k,l} = -20 - 8 \cos k\theta - 8 \cos l\theta + 4 \cos k\theta \cos l\theta, \qquad k = 1, 2, \ldots, n, \qquad l = 1, 2, \ldots, n.$$

Hence

$$\lambda(Z_3) = \nu_{1,1} \doteq 32$$
$$\mu(Z_3) = \nu_{n,n} \doteq 12 \pi^2 n^{-2}$$

which gives

$$P(Z_3) \doteq \frac{8}{3} \pi^{-2} n^2.$$

5. We conclude by discussing the nine-point approximation represented by the stencil

$$\frac{1}{12} \begin{Bmatrix} 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 16 & 0 & 0 \\ -1 & 16 & -60 & 16 & -1 \\ 0 & 0 & 16 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{Bmatrix}.$$

This stencil must be modified for points near the boundary. Thus, at the corner $(1,n)$, we use

$$\frac{1}{12}\left\{\begin{matrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 14 & 0 & 0 \\ 0 & 14 & -58 & 16 & -1 \\ 0 & 0 & 16 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{matrix}\right\};$$

this means that we have estimated the outside value by linear interpolation on the two nearest values on the same line. For points (r,s) at the edge, i. e., $r=1$ or $r=n$ or $s=1$, $s=n$ which are not corners, e. g., $r=1$, $s=n-1$, we have to use a stencil of the form:

$$\frac{1}{12}\left\{\begin{matrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 14 & 0 & 0 \\ -1 & 16 & -59 & 16 & -1 \\ 0 & 0 & 16 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{matrix}\right\}.$$

Omitting the factor $1/12$ we see that the corresponding $n^2 \times n^2$ matrix is

$$Z_4=\left\{\begin{matrix} Z+I, & 16I, & -I & & & & \\ 16I, & Z, & 16I, & -I & & & \\ -I, & 16I, & Z, & 16I, & -I & & \\ & & \cdot & & & & \\ & & & \cdot & & & \\ & & & & \cdot & & \\ & & -I, & 16I, & Z, & 16I, & -I \\ & & & -I, & 16I, & Z, & 16I \\ & & & & -I, & 16I, & Z+I \end{matrix}\right\}$$

where Z is the $n \times n$ matrix:

$$Z=\left\{\begin{matrix} -59, & 16, & -1 & & & & \\ 16, & -60, & 16, & -1 & & & \\ -1, & 16, & -60, & 16, & -1 & & \\ & & \cdot & & & & \\ & & & \cdot & & & \\ & & & & \cdot & & \\ & & -1, & 16, & -60, & 16, & -1 \\ & & & -1, & 16, & -60, & 16 \\ & & & & -1, & 16, & -59 \end{matrix}\right\}.$$

From (2.2) the characteristic values of Z are

$$\lambda_k = -60+2+\{64-(8+2\cos k\theta)^2\}, \qquad k=1,2,\ldots,n,$$

and by (2.3) the characteristic roots of Z_4 are those of the n matrices

$$Z_4^{(k)} = \begin{bmatrix} \lambda_k+1, 16, -1 & & & & & & \\ & 16, \lambda_k, 16, -1 & & & & & \\ & -1, 16, \lambda_k, 16, -1 & & & & & \\ & & \cdot & & & & \\ & & & \cdot & & & \\ & & & & \cdot & & \\ & & & -1, 16, \lambda_k, 16, -1 & & & \\ & & & & -1, 16, \lambda_k, 16 & & \\ & & & & & -1, 16, \lambda_k+1 & \end{bmatrix}, \qquad k=1,2,\ldots,n,$$

which are, again using (2.2),

$$\nu_{k,l} = \lambda_k+2+\{64-(8+2\cos l\theta)^2\}, \qquad k=1,2,\ldots,n, \qquad l=1,2,\ldots,n,$$
$$=\lambda_k+\lambda_l+60.$$

It follows that

$$\lambda(Z_4)=\nu_{1,1}\doteq 128$$
$$\mu(Z_4)=\nu_{n,n}\doteq 24\pi^2 n^{-2}$$

so that

$$P(Z_4)=\frac{16}{3}\,\pi^{2-}n^2.$$

6. An evaluation of the various methods could now be made. In addition to the inversion error, which is given in terms of the P-condition number by von Neumann and Goldstine, the truncation error must be considered. The local truncation errors for the various stencils have been given by Bickley [2]; however, it seems that the global truncation errors have been studied only in the case of the approximation (1.1) by S. A. Gerschgorin, P. C. Rosenbloom, J. L. Walsh, W. Wasow, and others. We shall not, however, carry out an evaluation, for it seems clear that the elimination method is not the most suitable one for problems involving sparse matrices; numerical experience, nevertheless, does show that the P-condition number of a matrix is a good indication of the intrinsic difficulty of the inversion problem.

7. *Remarks.*

(1) Dr. Young pointed out that our results could be obtained by solving the difference equations by separating the variables (cf., e. g., B. Friedman [3]).

(2) As in [8] the estimates of the N-condition of Turing [10] can be obtained, since we know *all* the eigenvalues of the matrices in question.

(3) Similar results can be obtained for the biharmonic equation (cf. [9]).

(4) I take this opportunity to correct slips in two papers in this series.

In p. 117 of [8] the expression for γ_{kj} for $j\leq k\leq n$ should read $(n+1)\gamma_{kj}=-k(n-j+1)+(n+1)(k-j)$. This was pointed out by D. E. Rutherford [7].

In p. 471 of [8a], line 3, the adjective "symmetric" (or "normal") should be inserted to qualify "matrix". This was pointed out in conversation by Martin Pearl.

References

[1] S. N. Afriat, Composite matrices, Quart. J. Math., Oxford 2d Ser. **5,** 81–98 (1954).

[2] W. G. Bickley, Finite difference formulae for the square lattice, Quart. J. Mech. Appl. Math. **1,** 35–42 (1948).

[3] B. Friedman, Eigenvalues of compound matrices, New York Univ. Rept. TW–16 (1951).

[4] W. E. Milne, Numerical solution of differential equations (John Wiley & Sons, Inc., New York, N. Y., 1953).

[5] D. J. Panow, Formelsammlung zur numerische Behandlung partieller Differentialgleichungen nach dem Differenzenverfahren (translated from 5th Russian ed.) (Akad. Verlag, Berlin, 1955).

[6] D. E. Rutherford, Some continuant determinants arising in physics and chemistry, Proc. Roy. Soc. Edinburgh **62A,** 229–236 (1947).

[7] D. E. Rutherford, Some continuant determinants arising in physics and chemistry, II, Proc. Roy. Soc. Edinburgh **63A,** 232–241 (1952).

[8] J. Todd, The condition of a certain matrix, Proc. Cambridge Phil. Soc. **46,** 116–118 (1949).

[8a] J. Todd, The condition of certain matrices, I, Quart. J. Mech. Appl. Math. **2,** 469–472 (1949).

[9] J. Todd, The condition of certain matrices, II, Arch. Math. **5,** 249–257 (1954).

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

[11] H. H. Goldstine and J. von Neumann, Numerical inverting of matrices of high order, Bul. Am. Math. Soc. **53,** 1021–1097 (1947); Proc. Am. Math. Soc. **2,** 188–202 (1951).

[12] J. Williamson, The latent roots of a matrix of special type, Bul. Am. Math. Soc. **37,** 585–590 (1931).

[13] D. M. Young, Jr., Ordvac solutions of the Dirichlet problem, J. Assoc. Computing Mach. **2,** 137–161 (1955).

[14] H. I. Meyer and B. J. Hollingsworth, A method of inverting large matrices of special form, Math. Tables and Other Aids to Computation **11,** 94–97 (1957).

[15] P. Stein and J. E. L. Peck, On the numerical solution of Poisson's equation over a rectangle, Pacific J. Math. **5,** 999–1011 (1955).

WASHINGTON, June 17, 1957.