MATHEMATICS

1961

SovietRxiv

Abstract

Full Text

MATHEMATICS

TAN CHZHEN

ON ROUND-OFF ERRORS IN THE NUMERICAL SOLUTION OF SYSTEMS OF CONSISTENT LINEAR ALGEBRAIC EQUATIONS BY THE CHOLESKY METHOD AND THE METHOD OF PIVOTAL ELEMENTS

(Presented by Academician M. V. Keldysh on XI 11, 1960)

1. In the paper of Neumann and Goldstein ([1]), round-off errors in the method of pivotal elements for fixed-point machines are investigated in greatest detail. We have obtained estimates of round-off errors in the Cholesky method for machines operating in the binary system with λ binary digits in floating-point mode. It is assumed here that for any two real numbers c and d the exact algebraic operations (addition "+", subtraction "−", multiplication "·", division "/", extraction of the square root "$(\ )^{1/2}$") and the corresponding pseudo-operations ("$+^*$", "$-^*$", "$\times$", ":", "$\sqrt{\ }$") satisfy the following conditions:

$$(c+d)-(c+^* d)=(c+d)\varepsilon_1,$$

$$(c-d)-(c-^* d)=(c-d)\varepsilon_2,$$

$$c\cdot d-c\times d=(c\cdot d)\varepsilon_3,$$

$$c/d-c:d=(c/d)\varepsilon_4 \qquad (\text{for } d\neq 0),$$

$$(c)^{1/2}-\sqrt{c}=(c)^{1/2}\varepsilon_5 \qquad (\text{for } c\geqslant 0),$$

where $|\varepsilon_i|\leqslant 2^{-\lambda}$ $(i=1,2,\ldots,5)$.

Let m systems with one and the same symmetric matrix A of order n be written in the form

$$AX = B,$$

where B is the matrix of the right-hand sides of the systems, of size $n \times m$; X is the matrix of unknowns of size $n \times m$. Suppose that $\overline{L}'\overline{L}$ is an approximate decomposition of the symmetric matrix A, where $\overline{L}$ and $\overline{L}'$ are, respectively, an upper triangular matrix and its transposed matrix. We shall have

$$M(A - \overline{L}'\overline{L}) \leqslant \left[\frac{1}{2}(n^2 + n + 6)M^2(\overline{L}) + (n-2)M(A)\right] 2^{-\lambda},$$

where $M(A)$ denotes the maximum of the moduli of the elements of the matrix A. If we denote by $\overline{Y}$ the approximate solution of the equation $\overline{L}'Y = B$, then we obtain

$$M(B - \overline{L}'\overline{Y}) \leqslant \left[\frac{1}{2}(n^2 + 5n - 8)M(\overline{L})M(\overline{Y}) + 2M(B)\right] 2^{-\lambda}.$$

Moreover, the inequality

$$M\left[\overline{L}'(\overline{Y} - \overline{LX})\right] \leqslant \left[\left(\frac{1}{6}n^3 + \frac{3}{2}n^2 - \frac{8}{3}n\right) M^2(\overline{L})M(\overline{X}) + 2nM(\overline{L})M(\overline{Y})\right] 2^{-\lambda}.$$

Finally, we obtain the estimate

$$M(B - A\bar{X}) \leq \left[\left({}^3/_2 n^3 + 2n^2 + {}^1/_3 n\right) M^2(\bar{L})M(\bar{X}) + \right.$$
$$\left. + \left({}^1/_2 n^2 + \frac{9}{2}n - 4\right) M(\bar{L})M(\bar{Y}) + (n^2 - 2n)M(A)M(\bar{X}) + 2M(B)\right] 2^{-\lambda}. \quad (1)$$

As an example, we consider the matrix $A = SS'$ of order 15, where

$$S = \begin{bmatrix} 1 & 2 & \ldots & 15 \\ 15 & 1 & \ldots & 14 \\ \ldots & \ldots & \ldots & \ldots \\ 2 & 3 & \ldots & 1 \end{bmatrix}.$$

For $\lambda = 35$, in inverting the matrix A, the right-hand side of estimate (1) gives $0.95 \cdot 10^{-6}$, which is close to the actual error on the left-hand side, equal to $0.14 \cdot 10^{-8}$.

In practice, for large n, one may consider only the principal term in estimate (1), namely ${}^2/_3 n^3 M^2(\bar{L})M(\bar{X})2^{-\lambda}$. Then in the present case, instead of $0.95 \cdot 10^{-6}$ we obtain $0.72 \cdot 10^{-6}$, a result differing only slightly from the preceding one.

For large n, the principal part of the error $N(B - A\bar{X})$ does not exceed $0.58\, n^3 N^2(\bar{L}) N(\bar{X}) 2^{-\lambda}$, where $N(A) = (\mathrm{Sp}\, A'A)^{1/2}$, and the principal part of the error $|I - A\bar{X}|$ does not exceed $0.58\, n^3 |\bar{L}|^2 |\bar{X}| 2^{-\lambda}$, or $0.58\, n^3 |A\bar{X}| P(A) 2^{-\lambda}$. Here

$$|A| = \max_{z \neq 0} \left| \frac{Az}{z} \right|,$$

where z denotes an n-dimensional vector, and

$$P(A) = \frac{\lambda_1}{\lambda_n},$$

where λ_1, λ_n are, respectively, the largest and smallest eigenvalues of the symmetric matrix A.

2. It is generally accepted ([2,3]) that, in application to the solution of systems of compatible linear algebraic equations with symmetric matrices, Cholesky' s method is more accurate than other methods. The question arises whether it can be proved that Cholesky' s method is more accurate than the method of principal elements. Our theoretical investigations and attempts to solve this problem led to an example for which Cholesky' s method gives a less accurate result than the method of principal elements.

We consider a system of equations in the form

$$Ax = b,$$

where

$$A = \begin{bmatrix} 200 & -100 & 400 & -310 & 100 \\ -100 & 1 & 2 & 1 & 3 \\ 400 & 2 & 3 & 3 & -1 \\ -300 & 1 & 3 & 2 & 4 \\ 100 & 3 & -1 & 4 & 4 \end{bmatrix}, \qquad b = \begin{bmatrix} 11 \\ 14 \\ 4 \\ 16 \\ 18 \end{bmatrix},$$

and x is the column vector of unknowns. For this system, by Cholesky's method, the Strela machine gave the solution

$$x = \begin{bmatrix} -0.773529482 \cdot 10^{-1} \\ 0.282058854 \cdot 10^{-2} \\ 0.100000004 \cdot 10 \\ -0.973529539 \cdot 10 \\ -0.473529496 \cdot 10 \end{bmatrix},$$

where

$$r = Ax - b = \begin{bmatrix} 0.72 \cdot 10^{-5} \\ 0 \\ 0.43 \cdot 10^{-6} \\ -0.70 \cdot 10^{-6} \\ -0.10 \cdot 10^{-6} \end{bmatrix}.$$

so that

$$\|r\|_1 = 0.72 \cdot 10^{-5}, \qquad \|r\|_2 = 0.84 \cdot 10^{-5}, \qquad \|r\|_3 = 0.72 \cdot 10^{-5},$$

while by the method of principal elements the solution is

$$x = \begin{bmatrix} -0.773529411 \cdot 10^{-1} \\ 0.282058823 \cdot 10^{-2} \\ 0.100000000 \cdot 10 \\ -0.973529411 \cdot 10 \\ -0.473529411 \cdot 10 \end{bmatrix},$$

where

$$r = \begin{bmatrix} 0.35 \cdot 10^{-5} \\ -0.10 \cdot 10^{-6} \\ -0.60 \cdot 10^{-7} \\ -0.10 \cdot 10^{-6} \\ -0.80 \cdot 10^{-6} \end{bmatrix},$$

so that

$$\|r\|_1 = 0.35 \cdot 10^{-5}, \qquad \|r\|_2 = 0.46 \cdot 10^{-5}, \qquad \|r\|_3 = 0.34 \cdot 10^{-5},$$

where $\|r\|_i$ $(i = 1, 2, 3)$ are the norms of the vectors (4).

These results mean that the Cholesky method is not in all cases more accurate than the method of principal elements, as is commonly assumed. Thus, the loss of accuracy in solving systems of equations by either method is approximately the same.

In conclusion, I express my sincere gratitude to M. R. Shura-Bura for assistance in carrying out this work.

Moscow State University
named after M. V. Lomonosov

Received
4 XI 1960

References

[1] J. Neumann, H. H. Goldstine, Bull. Am. Math. Soc., 53 (1947); M. R. Shura-Bura, UMN, 6, no. 4 (1951). [2] L. Fox, H. D. Huskey, J. H. Wilkinson, Quart. J. Mech. and Appl. Math., 1, 149 (1948); UMN, 5, no. 3 (1950). [3] A. M. Turing, Quart. J. Mech. and Appl. Math., 1, 287 (1948); UMN, 6, no. 1 (1951). [4] V. N. Faddeeva, *Computational Methods of Linear Algebra*, Moscow–Leningrad, 1950.

Note: Figure translations are in progress. See original paper for figures.