NOV 19 1964

MASTER

83214

~~NOT FOR PUBLIC RELEASE - OFFICIAL
DISTRIBUTION MAY BE MADE - OFFICIAL
REQUESTS MAY BE FILLED - REPORT
CONTAINS NOTHING OF PATENT INTEREST
PROCEDURES ON FILE IN RECEIVING
SECTION.~~

Comments on a Floating-Point Version of Nordsieck's Scheme for the
Numerical Integration of Differential Equations*

H. R. Lewis, Jr. and E. J. Stovall, Jr.

Los Alamos Scientific Laboratory, University of California

Los Alamos, New Mexico

ABSTRACT

Modifications in Nordsieck's scheme for numerical integration of differential equations are described which have permitted satisfactory operation of the scheme in floating-point arithmetic. The major modification, recommended for both fixed-point and floating-point operation, is a reformulation of the test for numerical stability which corrects a flaw in the original formulation. Also discussed, in relation to the present floating-point scheme, are Nordsieck's use of guard digits and his novel rounding techniques.

* Work performed under the auspices of the U. S. Atomic Energy Commission.

This paper was submitted for publication in the open literature at least 9 months prior to the issuance date of this Microcard. Since the U.S.A.E.C. has no evidence that it has been published, the paper is being distributed in Microcard form as a preprint.

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency Thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

One of the important problems in numerical analysis which arises in scientific and engineering research is that of numerical integration of a system of differential equations. Because of the frequency with which this problem occurs, it is valuable to have a general purpose numerical scheme with which the integration of a large class of differential equations can be reliably performed. One numerical scheme, which is useful as such a general purpose method, and well adapted for efficient use with digital computers, has been proposed by A. Nordsieck [1]. This scheme is designed to solve a system of first order equations,

$$\frac{dy_i}{dt} = f_i(t, y_1, y_2, \dots, y_n), \quad i = 1, 2, \dots, n, \quad (1)$$

with given initial conditions, whenever the f_i are such that a unique solution exists. The basic formulas of the method are equivalent to finding the fifth degree polynomial approximation to the desired solution of the differential equations which is determined from the values of y_i and f_i at the current value of the independent variable t , and from the values of f_i at the four preceding values of t . The effective approximating polynomial is identical with that of the Adams method of integration [2,3]. However, Nordsieck has reformulated and modified the Adams method in a way which is of interest for practical application.

An important practical feature of Nordsieck's scheme is the automatic increase and decrease of the elementary interval size during

the course of the integration. This is accomplished by means of two tests which are performed at each elementary integration step. One test determines an approximate bound of the truncation error in the solution. The other test is intended to guarantee that the integration scheme be numerically stable throughout the integration; that is, those solutions of the equations of the numerical method which are not related to the differential equations are supposed to be damped out if this test is always satisfied.

Nordsieck formulated the integration scheme for fixed-point arithmetic. The purpose of this note is to explain modifications of the original formulation which allow satisfactory operation with floating-point arithmetic [4]. The major modification, discussed in the first section, is a reformulation of the stability test itself, in order to correct a flaw in the original formulation, and this modification is recommended for both fixed-point and floating-point operation. The special procedures proposed by Nordsieck (novel rounding techniques and the use of guard digits), which were helpful in avoiding malfunctions of the test as it was originally formulated, are discussed in the second section.

Modification of the stability test came as a result of difficulties which were sometimes encountered with earlier floating-point versions of the integration scheme; these were 1) reduction of the elementary integration interval to unnecessarily or absurdly small values, and 2) unstable "blow-up" of the solution. Problems which have been done with

floating-point versions of the integration scheme include all of those reported by Nordsieck [1] and, also, integration of the equations of motion of a charged particle in a magnetic field. The occurrence of difficulties was very much more frequent with the latter problem. With the current floating-point version of the integration scheme all of the problems are done satisfactorily.

I. THE STABILITY TEST

In his original paper [1] Nordsieck proposes a sufficient condition for insuring stability of the numerical method with a comfortable margin of safety. In terms of the elementary integration interval h and the eigenvalues of the matrix $\frac{\partial f}{\partial y}$, whose elements are $\frac{\partial f_i}{\partial y_j}$, the condition is [5]

$$\frac{95}{288} |h\lambda| \leq 1/8, \quad (2)$$

for each eigenvalue λ . Nordsieck does not directly require satisfaction of this inequality. Instead, he proposes a test which is intended to insure satisfaction of the inequality, and which is more easily applied. However, his test does not guarantee the validity of the stability condition (2), except in the special case that only a single differential equation is to be solved. Let y and f denote the column matrices whose elements are y_i and f_i respectively. In the course of the iterative solution of the implicit equations of the scheme, three approximations to y are computed — first $y^{(1)}$, then $y^{(2)}$, and finally $y^{(3)}$. These column matrices are related to the square matrix $\frac{\partial f}{\partial y}$ through the

approximate equality [6]

$$(y^{(3)} - y^{(2)}) \approx \frac{95}{288} h \frac{\partial f}{\partial y} (y^{(2)} - y^{(1)}). \quad (3)$$

If there is only one differential equation to be solved ($n = 1$), then $y^{(1)}$, $y^{(2)}$, $y^{(3)}$, and $\frac{\partial f}{\partial y}$ are all single numbers, as opposed to true matrix quantities, and we have

$$|y^{(3)} - y^{(2)}| \approx \frac{95}{288} \left| h \left(\frac{\partial f}{\partial y} \right) \right| |y^{(2)} - y^{(1)}|. \quad (n = 1) \quad (4)$$

The test to insure satisfaction of the stability condition (2) proposed by Nordsieck for the case $n = 1$ is [7]

$$|y^{(3)} - y^{(2)}| \leq \frac{1}{8} |y^{(2)} - y^{(1)}|. \quad (5)$$

Indeed, for $n = 1$, the inequalities (2) and (5) are identical to within the approximation that Eq. (3) is an exact equality. (If $n = 1$, then $\lambda = \frac{\partial f}{\partial y}$.)

However, if there are two or more equations to be solved simultaneously ($n > 1$), then the situation is different. Let a norm of a matrix be denoted by enclosing the matrix symbol between double vertical bars; then, with suitably chosen norms, the relation corresponding to Eq. (4) is [8]

$$\|y^{(3)} - y^{(2)}\| \leq \frac{95}{288} \left\| h \frac{\partial f}{\partial y} \right\| \|y^{(2)} - y^{(1)}\|. \quad (n > 1) \quad (6)$$

Let $\lambda_{\max}$ be that eigenvalue of $\frac{\partial f}{\partial y}$ which has the largest magnitude. A standard inequality relating $\|\frac{\partial f}{\partial y}\|$ and $|\lambda_{\max}|$ is [9]

$$|\lambda_{\max}| \leq \|\frac{\partial f}{\partial y}\|. \quad (7)$$

The stability test proposed by Nordsieck for the case $n > 1$ may be written as [7]

$$\|y^{(3)} - y^{(2)}\| \leq \frac{1}{8} \|y^{(2)} - y^{(1)}\|, \quad (8)$$

where the column matrix norm used is either the largest magnitude of any element, or the euclidean norm. Combining relations (6), (7) and (8), we have

$$\frac{95}{288} |h\lambda_{\max}| \leq \frac{95}{288} \|h \frac{\partial f}{\partial y}\| \geq \frac{\|y^{(3)} - y^{(2)}\|}{\|y^{(2)} - y^{(1)}\|} \leq \frac{1}{8}. \quad (9)$$

It is to be noted that no bound whatsoever is obtained on $\|h \frac{\partial f}{\partial y}\|$ or on $|h\lambda_{\max}|$. In fact, there are numerical examples for which the inequality (8) which represents the test is satisfied while the inequality (2) which represents the stability condition is violated. A two-dimensional case illustrating this point is the following. Let

$$\frac{95}{288} h \frac{\partial f}{\partial y} = \frac{1}{4} \begin{pmatrix} -1 & 2 \\ -3 & 4 \end{pmatrix} \quad \text{and}$$

$$(y^{(2)} - y^{(1)}) = \begin{pmatrix} 3 \\ 2 \end{pmatrix}; \quad \text{then, using (3),}$$

$$(y^{(3)} - y^{(2)}) = \frac{1}{4} \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

With these choices, we have

$$\frac{|y^{(3)} - y^{(2)}|}{|y^{(2)} - y^{(1)}|} = \begin{cases} \frac{1}{12}, & \text{if the "maximum element" norm is used.} \\ \frac{1}{\sqrt{104}}, & \text{if the euclidean norm is used.} \end{cases}$$

With either type of norm, we have

$$\frac{|y^{(3)} - y^{(2)}|}{|y^{(2)} - y^{(1)}|} < \frac{1}{8}, \text{ and the test (8) is satisfied.}$$

On the other hand, the eigenvalues of $\frac{95}{288} h \frac{\partial f}{\partial y}$ are $\frac{1}{4}$ and $\frac{1}{8}$, so that $\frac{95}{288} h \lambda_{\max} = \frac{1}{2} > \frac{1}{8}$, which does not satisfy the stability condition (2). Thus, it is evident that the original formulation of a test to insure satisfaction of the stability condition (2), and therefore to insure stability, is incorrect except for the special case $n = 1$ [10].

The stability test which we propose, and which we have used in a floating-point version of the scheme [4], is precisely the basic stability condition given by expression (2). That is, the elements of the matrix $\frac{\partial f}{\partial y}$ are evaluated — analytically if possible, numerically otherwise — and either an upper bound of the magnitudes of the

eigenvalues, or the largest eigenvalue itself, is computed.

The original stability test was subject to malfunction because of round-off noise. This difficulty was alleviated, in the fixed-point version, by use of novel rounding techniques and by the use of guard digits. Such round-off noise problems do not interfere with the new stability test.

II. GUARD DIGITS AND SPECIAL ROUNDING PROCEDURES

In the fixed-point version of this integration scheme, with the stability test in the form given by expression (9), Nordsieck found it desirable to carry $\log_{\beta}(|h|^{-1})$ more digits in y than in f — so-called guard digits [11]. (β is the base of the number system with which computations are performed. For example, $\beta = 2$ with binary arithmetic.) The reason given for this is to minimize the accumulation of round-off error in y when the number of elementary steps is large. A different reason for keeping the guard digits is that a certain form of round-off noise then tends not to interfere with the functioning of the original stability test (8). This can be seen in the following way. From the equations of the integration scheme [1], it is easily derived that the differences, $(y^{(3)} - y^{(2)})$ and $(y^{(2)} - y^{(1)})$, can be expressed as

$$y^{(3)} - y^{(2)} = \frac{95}{288} h \left\{ f[x + h, y^{(2)}(x + h)] - f[x + h, y^{(1)}(x + h)] \right\} \quad (10a)$$

and

$$y^{(2)} - y^{(1)} = \frac{95}{288} h \left\{ f[x + h, y^{(1)}(x + h)] - f^P \right\}. \quad (10b)$$

The $y^{(i)}$ are approximations to y for the independent variable equal to $x + h$; f^P , the "predicted" value of f , is a first approximation to the value of f at $x + h$. Using Eqs. (10a) and (10b), the inequality (8) can be rewritten as

$$\|f[x+h, y^{(2)}(x+h)] - f[x+h, y^{(1)}(x+h)]\| \leq \frac{1}{8} \|f[x+h, y^{(1)}(x+h)] - f^P\|. \quad (11)$$

When the test is written in this form, it is evident that round-off noise in the computed values of the derivatives can interfere with the functioning of the test. This is accentuated by the fact that both of the differences in the inequality can be quite small. We can estimate the amount of error in $f(x, y)$ due to an error in y , for the case $n = 1$, as follows. Letting Δy be the error in y and Δf the corresponding error in f , we have, in first approximation,

$$|\Delta f| \approx \left| \frac{\partial f}{\partial y} \right| |\Delta y|. \quad (12)$$

However, $\left| \frac{\partial f}{\partial y} \right|$ is bounded by the stability condition

$$\left| \frac{\partial f}{\partial y} \right| \leq \frac{1}{8 \left(\frac{95}{288} \right) |h|} \quad (13)$$

Combining (12) and (13), we have

$$|\Delta f| \lesssim \frac{36}{95} \left| \frac{1}{h} \right| |\Delta y| = \beta^{\log_{\beta} \left(\frac{36}{95} \left| \frac{1}{h} \right| \right)} \Delta y. \quad (14)$$

Taking the case of binary arithmetic ($\beta = 2$), we see that, for $n = 1$, an error equal to the least count in y will usually give rise to an error less than the least count in f if $\log_{\beta} \left(\frac{1}{|h|} \right)$ more digits are carried in y than in f . Thus, round-off noise of the magnitude of the least count in y would tend not to interfere with application of a stability test in the form of expression (11).

Because the number of figures in a floating-point number in a digital computer does not change, the use of guard digits could not be easily employed for stopping the propagation of round-off noise from y into f , even in the case $n = 1$. (An alternative would be to use double-precision arithmetic for y .) It is, therefore, fortunate that the modified stability test, which is identical with the stability condition (2), is not influenced by such noise. Neither guard digits nor double-precision arithmetic have been used in the floating-point version of Nordsieck's scheme.

A novel way of rounding certain quantities which appear in this integration scheme was introduced in the original fixed-point version; this type of rounding was called "rounding away from zero" [12]. The purpose of this rounding was to eliminate a type of noise which

sometimes interfered with the proper operation of the tests which control the size of the elementary interval. A floating-point procedure can be devised which is analogous to rounding away from zero in fixed-point. We have tried this procedure in the floating-point version of the integration scheme which incorporates the new stability test. However, with that version of the integration scheme, we have not observed any over-all improvement in the operation of the interval control logic when the rounding procedures are included; nor have we observed any malfunction of the interval control logic when these special rounding techniques are omitted. The floating-point "rounding away from zero", at least when done with the Fortran computer language, is rather time consuming, with the result that the computer time necessary for a particular problem can be substantially longer with the special rounding techniques than without. For these reasons, "rounding away from zero" has been omitted from the floating-point version of the integration scheme.

We express our appreciation to R. M. Brown and P. Ponzo of the Coordinated Science Laboratory of the University of Illinois, and to E. P. Gray and J. G. Monteabaro of the Applied Physics Laboratory of the Johns Hopkins University, for kindly providing us with their floating-point versions of Nordsieck's integration scheme. The present work began as a result of using their computer programs.

FOOTNOTES AND REFERENCES

1. A. Nordsieck, "On Numerical Integration of Ordinary Differential Equations," Math. Comp. 16, 22 (1962). (See also A. Nordsieck, "Automatic Numerical Integration of Ordinary Differential Equations," Proc. Sympos. Appl. Math. Amer. Math. Soc., Vol. XV, pp. 241-250 (1963).)
2. W. E. Milne, Numerical Solution of Differential Equations (John Wiley and Sons, Inc., New York, 1953). pp. 53-55.
3. L. Collatz, The Numerical Treatment of Differential Equations (Springer-Verlag, Berlin, 1960), pp. 83-86.
4. A computer program, incorporating these modifications and using floating-point arithmetic, has been written in the Fortran-II and Fortran-IV languages and has been found to operate very satisfactorily with the IBM 7094 and 7030 computers. Details of this program will be presented in a Los Alamos Scientific Laboratory report which is in preparation. The programming details are in large measure based on the computer program of an earlier floating-point version of Nordsieck's integration scheme which was kindly made available to us by R. M. Brown P. Ponzo of the Coordinated Science Laboratory of the University of Illinois. That program, dated October 1, 1963 and designated D2 (F) UOFI DEQ, was part of the 1604 computer library of the Coordinated Science Laboratory. The program of another earlier version, dated October 26, 1962 and designated RCC Library Routine No. 5.02.06, was made available to us by E. P. Gray and J. G. Monteabaro of the Applied Physics Laboratory of the Johns Hopkins University.

5. Reference 1, Eq. (8) and pp. 29, 30, and 36.
6. Reference 1, Eqs. (8) and (20).
7. Reference 1, Eq. (22a) and p. 36.
8. If the square matrix and column matrix norms are so chosen that (6) is valid, then these norms are said to be consistent with one another. Norms are usually chosen in this way. See, for example, the reference given in footnote 9.
9. J. H. Wilkinson, Rounding Errors in Algebraic Processes (Her Majesty's Stationery Office, London, 1963), pp. 79-82.
10. A. Nordsieck (private communication) agrees that the stability test as originally formulated does not insure satisfaction of the stability condition (2).
11. Reference 1, pp. 25, 35, and 46.
12. Reference 1, pp. 38-39 and Appendix C.