NASA TECHNICAL REPORT

NASA TR R-262

c. 1

NASA TR R-262

AN OPERATIONAL UNIFICATION OF FINITE DIFFERENCE METHODS FOR THE NUMERICAL INTEGRATION OF ORDINARY DIFFERENTIAL EQUATIONS

by Harvard Lomax

Ames Research Center
Moffett Field, Calif.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • WASHINGTON, D. C. • MAY 1967

NASA TR R-262

AN OPERATIONAL UNIFICATION OF FINITE DIFFERENCE METHODS FOR THE NUMERICAL INTEGRATION OF ORDINARY DIFFERENTIAL EQUATIONS

By Harvard Lomax

Ames Research Center
Moffett Field, Calif.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

For sale by the Clearinghouse for Federal Scientific and Technical Information
Springfield, Virginia 22151 – CFSTI price $3.00

TABLE OF CONTENTS

Page

AN OPERATIONAL UNIFICATION OF FINITE DIFFERENCE METHODS FOR THE NUMERICAL INTEGRATION OF ORDINARY DIFFERENTIAL EQUATIONS

By Harvard Lomax
Ames Research Center

SUMMARY

One purpose of this report is to present a mathematical procedure which can be used to study and compare various numerical methods for integrating ordinary differential equations. This procedure is relatively simple, mathematically rigorous, and of such a nature that matters of interest in digital computations, such as machine memory and running time, can be weighed against the accuracy and stability provided by the method under consideration. Briefly, the procedure is as follows:

(1) Find a single differential equation that is sufficiently representative (this is fully defined in the report) of an arbitrary number of nonhomogeneous, linear, ordinary differential equations with constant coefficients.

(2) Solve this differential equation exactly.

(3) Choose any given numerical method, use it -- in its entirety -- to reduce the differential equation to difference equations, and, by means of operational techniques, solve the latter exactly.

(4) Study and compare the results of (2) and (3).

Conceptually there is nothing new in this procedure, but the particular development presented in this report does not appear to have been carried out before.

Another purpose is to use the procedure just described to analyze a variety of numerical methods, ranging from classical, predictor-corrector systems to Runge-Kutta techniques and including various combinations of the two.

INTRODUCTION

At present a large body of literature is devoted to the development and presentation of methods for integrating ordinary differential equations with given initial conditions. These methods are based on local polynomial approximation and are commonly divided into two classes, predictor-corrector methods and Runge-Kutta methods. The former are, as generally presented, not

self-starting and use a fixed interval, or step, at which the function and its derivative are evaluated as the integration proceeds. The latter are self-starting and the interval of evaluation may vary from step to step. A current trend is to combine these two classes. The resulting methods are variously referred to as hybrid, generalized predictor-corrector, and combined. The latter designation is used herein.

In this report a mathematical procedure, outlined in the summary, is presented which provides us with the capability of comparing these methods as they apply to simultaneous, linear, ordinary differential equations with constant coefficients. It is quite true that linear equations with constant coefficients are an extremely special set of all possible differential equations, and, in fact, the numerical methods being discussed here are rarely used to solve them. However, such equations can be solved analytically both as differential equations, and as difference equations when transformed to the latter by a linear numerical scheme. The conclusion regarding the accuracy and stability of a numerical method when studied in this way is, therefore, precise. We need then only to defend the reasonable hypothesis that a numerical method which, on some given basis, is unquestionably inferior in solving linear cases, is, on the same basis, also inferior, in general, for use in solving nonlinear ones.

When studied by the above procedure, all polynomial methods (known to the author) proposed for integrating ordinary differential equations fall into a smoothly connected system. By "smoothly connected," we mean, for example, that there is no sharp dividing line between predictor-corrector and Runge-Kutta methods. In fact, the standard, fourth-order, Runge-Kutta method is, in predictor-corrector terminology, a method composed of the successive application of an Euler predictor, an Euler corrector, a Nystrom predictor, and a Milne corrector. As such statements indicate, one of the principal difficulties that can arise when different schools of thought are brought together is the construction of a consistent and precise terminology. And the most troublesome problem in this area is to guard against conclusions based on implication. In particular, such a difficulty arises in the use of the term "step number" when combined methods are discussed. This is examined in the next paragraph.

All numerical methods of the type being considered are cyclic in application; that is, for a fixed reference value of the independent variable, a pattern of calculations is performed (solving equations, evaluating derivatives, estimating errors, etc.). At the end of these calculations the value of the function has been determined at a point advanced by some interval. The independent variable is re-referenced ahead by another interval and the identical pattern of calculations is repeated. These cycles are continued indefinitely. The interval involved is referred to as the step size. The number of locations, spaced by a fixed value of this interval, at which the function and/or its derivative are retained for use in the next cycle or pattern of computations, corresponds to the step number of the given method. The definitions of step size and step number when made in this way hold for these terms as they are generally used in the literature for both combined and uncombined methods. In predictor-corrector schemes this step number is a fundamental parameter that can be, and is, used to connect the stability and

accuracy of a given method. In fact, in a well known theorem, Dahlquist states that no stable, predictor-corrector method with a step number, k, can give a local polynomial approximation of order $k+2$ if k is odd, or of order $k+3$ if k is even. However, in Runge-Kutta methods, or methods that combine the Runge-Kutta and predictor-corrector concepts, this step number is not connected in any way with stability. Thus, stable combined methods having any value for the step number (including one) can be constructed that will fit local polynomials of any order. This implies that combined methods are greatly superior to uncombined ones. But, in fact, for fixed values of machine memory and running time, the maximum order[1] of a local polynomial fit appears to be the same for stable methods combined or uncombined.

At the beginning of the report certain basic terms are defined so as to make the subsequent discussion more precise. Then the approach to be used in the analysis is described and it is shown that a single representative differential equation can be used to study the accuracy and stability of difference-differential approximations as they apply to the analysis of simultaneous differential equations. An attempt is made to classify various classical and modern numerical methods according to three categories:

1) The number of iterations per cycle of computation

2) Whether they are complete or incomplete

3) Whether they are combined or uncombined.

Some general procedures falling into certain combinations of these categories are analyzed in detail. Finally, the operational form of a difference-differential equation is defined and its implications with regard to the study of numerical methods is discussed.

SYMBOLS

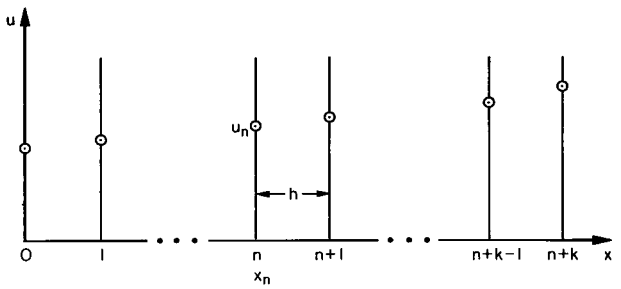
A	constant in representative equation (See eq. (37).)
$DE(E)$	see equation (52)
er_p	error of a numerical method in terms of local polynomial approximation
er_μ	error of a numerical method in calculating the particular solution of the representative equation (37) (See eqs. (57) and (63).)
er_λ	error of a numerical method in calculating the complementary solution of the representative equation (37) (See eqs. (68) and (71).)

[1]The magnitude of the leading error term found by means of a Taylor series expansion is lowest for the combined methods, however.

E	difference operator (See eq. (9).)
h	computational step size (See eq. (123).)
H	representational step size (See definition (2).)
j	index used in expressing difference-differential equations (See eq. (3).)
J	$k + 1 - j$
k	step number in a predictor or corrector
L_{ij}	coefficients in the operational form (See, e.g., eqs. (51).)
n	reference step location
NU	see equation (52)
R_{ij}	coefficients in the operational form (See, e.g., eqs. (51).)
u,w	dependent variables
u',w'	$\frac{du}{dx}$, $\frac{dw}{dx}$
x	independent variable
$\alpha,\beta,\gamma,\delta,\ldots$	coefficients of dependent variable in difference-differential equations
$\alpha',\beta',\gamma',\delta',\ldots$	coefficients of derivative of dependent variable in differential equations
λ	representative eigenvalue, that is, coefficient of u in representative differential equation (37)
$\lambda_i, i>1$	spurious roots of difference equation
λ_1	principal root of difference equation
$\lvert\lambda h\rvert_c$	induced stability boundary (See eq. (73).)
μ	representative maximum frequency (See eq. (37).)
σ_m	eigenvalues of simultaneous ordinary differential equations (11)

DEFINITION OF TERMS

Some of the following expressions are in common usage but vary slightly in meaning with different authors. The definitions given below are intended for this report to simplify and make more precise the subsequent discussion.



Sketch (a)

Difference-differential equations: Let the dependent variable u be a function of the independent variable x. Let u' represent the derivative of u with respect to x and designate x_n by nh and $u(x_n)$ by u_n where n is an integer and h is a constant. Then equations which relate $u_{n+k+1-j}$, $u'_{n+k+1-j}$ and $x_{n+k+1-j}$ where $j = 1, 2, \ldots, k+1$ are called difference-differential equations with step number k.

Predictor: Any difference-differential equation relating u_{n+k} to values of u and u' at previous steps. Thus, for a k-step predictor

$$u_{n+k} = f(u_{n+k+1-j},\ u'_{n+k+1-j},\ x_{n+k+1-j}), \qquad j = 2, 3, \ldots, k+1$$

A predictor is an explicit formula that extrapolates given data.

Corrector: Any difference-differential equation relating u_{n+k} to the values of u and u' at $n+k$ as well as to those at previous steps. Thus, for a k-step corrector

$$u_{n+k} = f(u_{n+k+1-j},\ u'_{n+k+1-j},\ x_{n+k+1-j}), \qquad j = 1, 2, \ldots, k+1$$

In this form the corrector is an implicit formula. In practice the values of u and u' in the arguments of f are generally those determined by predictors or previous correctors.

Iteration: In the numerical solution of ordinary differential equations the repeated calculation of the right-hand side of equations having the form

$$u' = F(x, u) \tag{1a}$$

or for multiple equations

$$\left.\begin{aligned} u'_1 &= F_1(x, u_1, u_2, \ldots) \\ u'_2 &= F_2(x, u_1, u_2, \ldots) \\ &\vdots \end{aligned}\right\} \tag{1b}$$

is necessitated. In this report we refer to every such evaluation (i.e., explicit calculation of the derivatives using the differential equations) as an iteration. By this definition, methods composed only of predictors require one iteration per step. Methods using one predictor followed by one corrector require two iterations per step, etc.

Reference step: We will inspect a wide variety of methods in which the words "step size," by common usage, have different implications. In order to have a parameter by means of which all methods can be compared on a common basis, the term "reference step" is introduced and designated by the symbol H.

$$H \equiv \text{the increment in } x \text{ that a solution is advanced by two iterations} \tag{2}$$

In many applications the numerical calculations necessary to evaluate the derivatives, $F_j(x,u)$ in equations (1), are extremely complicated and time consuming. In such applications, if errors are referenced to H, the accuracy of various numerical methods can be compared with the assurance that the total machine running time will be very nearly the same. Since most methods in practical use employ a predictor followed by just one corrector, two iterations were chosen for a base (rather than one) so that H would coincide with the most commonly used error reference. Both[2] h, the computational step size, and H, the reference step size, are used in error terms in the following analysis.

Cycle of computation: All the calculations and logic required to advance the data while n refers to the same location. A cycle is completed when all the dependent variables and their derivatives at $n + k$ have been calculated as accurately as the chosen method permits and preparation for stepping ahead commences.

Family: Any combination of values of u, u' and other families at $n + k$, $n + k - 1$, . . ., n that is formulated and used in a cycle of computation. A family may or may not be saved for future cycles of computation. In this report a family is usually designated by a superscript, and a predictor always generates the first family. A derivative belongs to that family of u used in its calculation; that is,

$$u^{(i)'} = F\left(x, u^{(i)}\right)$$

Final family: The new values of u and u' last evaluated in a cycle of computation. The superscript is always omitted from the final family of u (its distinguishing feature) and sometimes from the final family of u' (see the definitions below of complete and incomplete methods).

---

[2]When corrector methods are analyzed as such, without regard to how their equality is brought about, the reference step H is undefined.

Memory in a k-step method: All those values of u, u' and other families (if there are any) that are used but do not change during one cycle of computation.

Incomplete methods: Methods for which the dependent variable and its derivative are members of the same final family. That is, after the dependent variable is evaluated for the last time at a given point, it is used to calculate the derivative at the same point. Most "conventional" methods (Hammings's, Milne's, etc.) are incomplete. In this case the superscript is omitted from the final family representing the derivative.

Complete methods: Methods in which the derivative of at least one final family is never evaluated. They are referred to as complete because they most completely fill the matrix which determines the operational form.

Combined methods: Methods that combine the concepts usually separately designated as predictor-corrector and Runge-Kutta. A combined method can be thought of either as a predictor-corrector method without equal spacing, or a Runge-Kutta method with memory. Combined methods can be either complete or incomplete.

Fundamental family: One that is computed using a memory composed only of final families.

Embedded polynomial: The highest order polynomial which is an exact solution to a given set of difference-differential equations.

FUNDAMENTALS

Difference-Differential Equations

Two of the simplest difference-differential equations are

$$u_{n+1} - u_n - hu'_n = 0$$

and

$$u_{n+1} - u_n - \frac{1}{2} h\left(u'_{n+1} + u'_n\right) = 0$$

and are referred to as the Euler and modified Euler equations, respectively. These and all such formulas presented in books on numerical analysis are special forms of the general, linearized, k-step, difference-differential equation with constant coefficients which can be written

$$u_{n+k} - \beta_1 u_{n+k} - h\beta'_1 u'_{n+k} - \cdots \beta_j u_{n+k-j} - h\beta'_j u'_{n+k-j} - \cdots - \beta_{k+1} u_n - h\beta'_{k+1} u'_n = 0 \quad (3)$$

Nearly always, equation (3) represents formulas based on polynomial approximation. This simply means that if each u and u' is expanded in a Taylor series about x_n, the coefficients of the powers of h in equation (3) will vanish up through some integer L. The number L is then the order of the polynomial approximating the function in the interval $x_n \leq x \leq x_{n+k}$ (the embedded polynomial) and the product of h^{L+1}, and its coefficient is the first term of the truncation error. For example, since

$$u_{n+1} = u(nh + h) = u(x_n + h) = u_n + hu'_n + \frac{1}{2} h^2 u''_n + \frac{1}{6} h^3 u'''_n + \ldots$$

$$hu'_{n+1} = hu'(x_n + h) = hu'_n + h^2 u''_n + \frac{1}{2} h^3 u'''_n + \ldots$$

we can construct for the modified Euler method the simple table:

From	u_n	hu'_n	$h^2u''_n$	$h^3u'''_n$	. . .
u_{n+1}	1	1	1/2	1/6	. . .
$-u_n$	-1	0	0	0	. . .
$-1/2\ hu'_{n+1}$	0	-1/2	-1/2	-1/4	. . .
$-1/2\ hu'_n$	0	-1/2	0	0	. . .
Sums to	0	0	0	-1/12	. . .

Clearly, the order of the polynomial embedded in the modified Euler method is 2 (even though only one step is used) and the truncation error is predominantly $-u''' h^3/12$. A similar tabulation for equation (3) is shown below.

From	u_n	hu'_n	$h^2u''_n$	$h^3u'''_n$	$h^4u'''_n$	. . .
u_{n+k}	1	k	$1/2\ k^2$	$1/6\ k^3$	$1/24\ k^4$	. . .
$-\beta_1 u_{n+k}$	$-\beta_1$	$-\beta_1 k$	$-1/2\ \beta_1 k^2$	$-1/6\ \beta_1 k^3$	$-1/24\ \beta_1 k^4$	. . .
$-h\beta'_1 u'_{n+k}$	0	$-\beta'_1$	$-\beta'_1 k$	$-1/2\ \beta'_1 k^2$	$-1/6\ \beta'_1 k^3$	. . .
$-\beta_2 u_{n+k-1}$	$-\beta_2$	$-\beta_2(k-1)$	$-1/2\ \beta_2(k-1)^2$	$-1/6\ \beta_2(k-1)^3$	$-1/24\ \beta_2(k-1)^4$	. . .
$-h\beta'_2 u'_{n+k-1}$	0	$-\beta'_2$	$-\beta'_2(k-1)$	$-1/2\ \beta'_2(k-1)^2$	$-1/6\ \beta'_2(k-1)^3$	. . .
.	.	.	.	.	.	. . .
.	.	.	.	.	.	. . .
.	.	.	.	.	.	. . .
$-\beta_{k+1} u_n$	$-\beta_{k+1}$	0	0	0	0	. . .
$-h\beta'_{k+1} u'_n$	0	$-\beta'_{k+1}$	0	0	0	. . .
Sums to	0	0	0	0	0	. . .

Equating the sum of the first L columns to zero gives the conditions on the

β_j and β'_j required if equation (3) is to represent a polynomial of order L through k + 1 points. One can show that the product of the sum of the lth column and its heading is

$$er_p(l) = -\frac{h^l}{l!} u_n^{(l)} \left\{ \sum_{j=1}^{k+1} \left[l(k+1-j)^{l-1}\beta'_j + \beta_j(k+1-j)^l \right] - k^l \right\} \qquad (4)$$

Therefore, $er_p(L+1)$ is the first term in the truncation error of a Taylor series expansion for any function $u(x)$ represented by the difference-differential equation (3). The total truncation error is given by

$$\sum_{l=L+1}^{\infty} er_p(l)$$

Until recent years[3] equation (3) was used as the sole basis for determining the accuracy and stability of a numerical method. As is now well recognized, this is rarely a correct procedure. Let us suppose, for example, we are using equation (3) to find the value of u at x_{n+k}. Then the only time it describes the total numerical method is when both β_1 and β'_1 are identically zero. This is the case when a predicted value is calculated but no correction is made. Equation (3) also represents the numerical result of an implicit method where the terms multiplying β_1 and β'_1 might have been calculated by some iterative procedure. This is the assumption under which it is usually applied. Almost all practical methods use at least one corrector and when such is the case the accuracy and stability of the actual results are affected by the mutual interaction of the predictor and all of the subsequent correctors. This will be fully developed in the following sections.

At this point we wish merely to define a notation for a true predictor-corrector process. Consider, as above, that the values of u_{n+k} and u'_{n+k} are unknown but all values with prior subscripts are known, being either given or obtained from previous calculations. Then for the predicted value at $n + k$ we can write

$$u_{n+k}^{(1)} = \sum_{j=2}^{k+1} (\alpha_j u_{n+k+1-j} + \alpha'_j h u'_{n+k+1-j}) \qquad (5)$$

or, to shorten the notation,

$$u_{n+k}^{(1)} = \sum_{j=2}^{k+1} (\alpha_j u_{n+J} + \alpha'_j h u'_{n+J}) \qquad (6a)$$

[3] See the next section for some historical discussion.

where $J = k + 1 - j$. This forms the first family[4] at the location $n + k$ which is designated by the superscript (1). If this is followed by a corrected value at $n + k$, we can write, where again $J = k + 1 - j$,

$$u_{n+k}^{(2)} = \beta_1' h u_{n+k}^{(1)'} + \sum_{j=2}^{k+1} (\beta_j u_{n+J} + \beta_j' h u_{n+J}') \tag{6b}$$

which defines a second family at $n + k$. (The possibility of including a β_1 term is discussed under equation (44).) Another corrector could be added with coefficients γ_j, γ_j' forming a third family, etc. If, however, we consider the cycle of computation complete after the evaluation of equation (6b), then the second family is the final family for the dependent variable u. Next, a decision must be made as to whether or not $u_{n+k}^{(2)}$ shall be used to evaluate another estimate of the derivative at $n + k$. (In this regard, see Definition - Family.) A derivative has already been calculated at $n + k$, namely $u_{n+k}^{(1)'}$, and it can be used to advance the solution. If this path is chosen, we (in this report) refer to the method as a complete method. The function and its derivative are members of different families (initially generating the various families presents the same difficulties that arise in starting multistep methods) and equations (6) can be written

$$\left.\begin{aligned} u_{n+k}^{(1)} &= \sum_{j=2}^{k+1} \left(\alpha_j u_{n+J} + \alpha_j' h u_{n+J}^{(1)'}\right) \\ u_{n+k} &= \beta_1' h u_{n+k}^{(1)'} + \sum_{j=2}^{k+1} \left(\beta_j u_{n+J} + \beta_j' h u_{n+J}^{(1)'}\right) \end{aligned}\right\} \tag{7a}$$

where the superscript (2) has been omitted from the final family for u. Choosing the other path provides (what is referred to herein as) an incomplete method. In this case $u_{n+k}^{(2)}$ is used to find $u_{n+k}^{(2)'}$ and the latter is placed in memory for use by subsequent predictors and correctors. Now, the function and its derivative are members of the same family and the superscript (2) is omitted from both; thus,

$$\left.\begin{aligned} u_{n+k}^{(1)} &= \sum_{j=2}^{k+1} (\alpha_j u_{n+J} + \alpha_j' h u_{n+J}') \\ u_{n+k} &= \beta_1' h u_{n+k}^{(1)'} + \sum_{j=2}^{k+1} (\beta_j u_{n+J} + \beta_j' h u_{n+J}') \end{aligned}\right\} \tag{7b}$$

Incomplete methods are the most common, but not necessarily the best.

[4]See Definition of Terms.

Table I lists the coefficients of a few of the commonly used difference-differential equations, together with the leading terms of their truncation errors as defined by equation (4). Identifying names are included, although these are not unique.

The Method Used to Measure Accuracy and Stability

Development.- In the application of equations (7) to equations (1), the matter of overall accuracy in the resulting numerical scheme depends not only on the truncation error but also on the stability of the numerical process as it is continued along a number of steps. Thus, as is well known, the modified Euler method, row 1 in table I(b), is stable; but the Nystrom equation, row 2 in table I(a), when used by itself, is unstable. The usefulness of any set of difference-differential equations depends upon a balance between accuracy and stability. Dahlquist (ref. 1) found the maximum order of a polynomial that could be embedded in equation (3) for a given k under the condition that the resulting predictor-corrector method would be stable as $h \to 0$. He concluded, for example, that a three-step method could never support a polynomial of order h^5 or higher and still be stable. Hamming (ref. 2), at about the same time, developed a stable three-step corrector formula having a truncation error led by a term of order h^5, the minimum possible according to the proof of Dahlquist. Hamming's stable corrector formula and the most accurate, but highly unstable, three-step corrector formula are shown in rows 6 and 9 in table I(b).

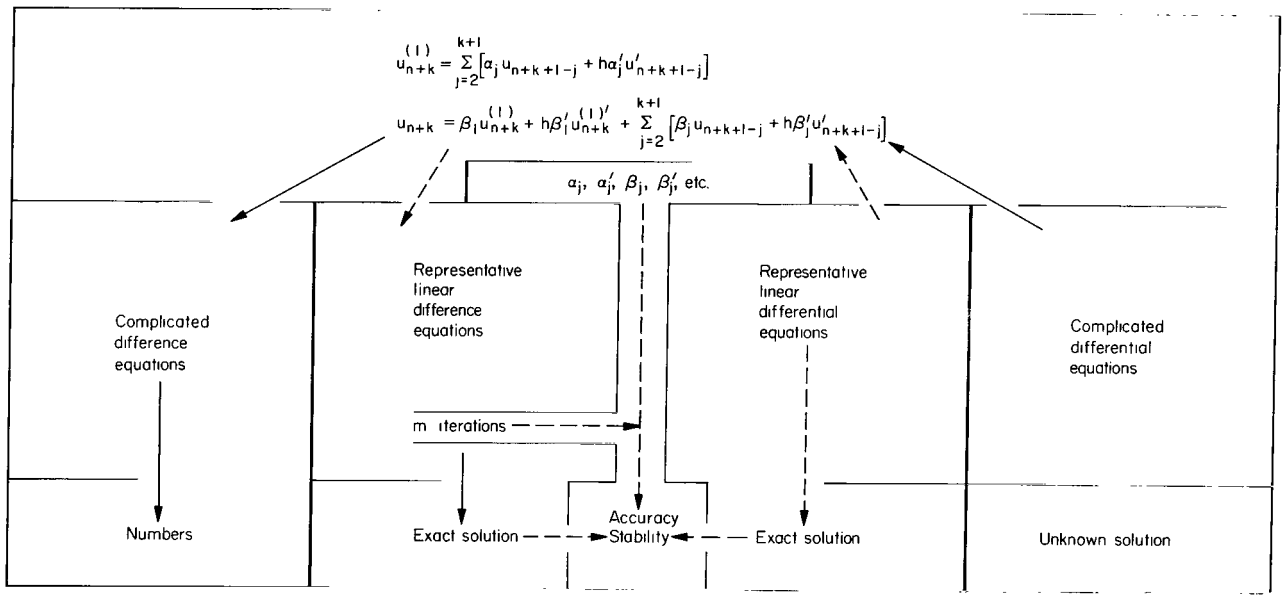
In a very interesting paper, Chase (ref. 3) put a new light on the developments mentioned above and brought out two important points. First (with a notable exception in Hamming's article) nearly all theorems and proofs regarding stability published prior to Chase's paper are based on the limiting case when $h \to 0$. Second, nearly all analyses of corrector formulas, including Hamming's, assume that all effects of the predictor have vanished, or, in other words, that the corrector equation is brought into complete balance. Chase showed that when the above conditions are not met (which is certainly the practical case, since step sizes cannot be zero and very often the corrector is used only once), the conclusions regarding stability of the various methods undergo startling changes. For example, Hamming's corrector formula - fully satisfied - is stable for values of $-|\lambda h|$ even less than -2. But the result of predicting with row 6 of table I(a) and correcting with row 6 of table I(b) (Hamming's method without the modifiers) is unstable for $-|\lambda h| < -0.5$. Even more dramatically, Chase showed that predicting with row 6 in table I(a) and correcting with row 5 of table I(b) (Milne's method without modifiers) is stable for $-0.3 > -|\lambda h| > -0.8$. In other words, a predictor and a corrector individually unstable for all $-|\lambda h| < 0$ combine to form a stable method for a certain applicable range of $-|\lambda h|$!

A final lesson to be learned from Chase's paper concerns the use of "modifiers." These are weighted combinations of the predicted and succesive corrected values. In the two cases mentioned above, Hamming's and Milne's, the use of modifiers in one case increased and in the other case decreased the range of stability. This suggested a study which would determine optimum

values for the modifiers. In the terminology of the present report, a modifier is a family that is not fundamental. By means of operational techniques, we shall see that a method with a given step number and modifiers can be identified with a method having a higher step number but composed only of fundamental families.

Description.- Consideration of the matters discussed above suggested an approach differing from the one presented in references 4 and 5. The approach in this report does not optimize the coefficients β_j and β'_j in equation (3), but rather the coefficients in an "operational form" as defined in the last section of this report. The coefficients defined are the fundamental parameters governing the accuracy and stability of linear numerical quadrature formulas and are functions of all the α_j, α'_j, β_j, and β'_j and their complete interaction. Furthermore, stability will not be studied in the limit as $h \to 0$, but rather over a finite range of h, a range which is to be made as large as possible for a given accuracy, and includes the entire complex plane.

Sketch (b) graphically presents the details of the proposed approach. A representative differential equation (or set of differential equations),



Sketch (b)

which is discussed fully in the next section, is chosen and solved exactly. Then a group of linear difference-differential equations with unspecified constant coefficients is introduced and combined with the differential equation to form a set of linear difference equations. Operational techniques are used to solve these difference equations exactly. The solution to the difference equations is then compared with the solution to the differential

equations. Eventually, the coefficients α_j, α'_j, etc., are chosen so that the two exact solutions match as closely as possible under the condition that the resulting process remain stable over a given range of step size.

Operational Solution of Difference Equations

In the presentation of the analysis in the following sections the reader is assumed to have some knowledge of the theory of ordinary difference equations. This theory is well developed but its publication is not nearly so widespread as that on the theory of differential equations. A brief review of a portion of an operational approach to difference equations is given below. For complete treatments see Boole (ref. 6) or Milne-Thomson (ref. 7).

Classically,

$$u_{n+k} + C_1 u_{n+k-1} + \ldots + C_k u_n = F(n) \tag{8}$$

is defined as an ordinary difference equation of order k. Unfortunately, the word order in most modern books and articles on numerical methods is used to designate the highest integer exponent in the polynomial embedded in equation (8). In this report we also refer to the order of a method in the latter fashion and refer to k in equation (8) as the step number. If the coefficients $C_1, C_2, \ldots, C_k$ in equation (8) are independent of n and u, the equation is an ordinary linear difference equation of step number k with constant coefficients. The solution to the equation when $F(n) = 0$ is the complementary solution which is added to the particular solution involving $F(n)$ when $F(n) \neq 0$.

Solutions to equation (8), when the coefficients are constant, are obtained by operational methods similar to the Laplace or Fourier transforms for differential equations. If E is defined as the operator

$$Eu_n \equiv e^{h(d/dx)} u_n = u_{n+1} \tag{9}$$

equation (8) can be written

$$\{E^k + C_1 E^{k-1} + \ldots + C_k\} u_n = F(n)$$

The complementary solution is determined by finding the roots to the characteristic equation

$$E^k + C_1 E^{k-1} + \ldots + C_k = \{E - \lambda_1\}\{E - \lambda_2\} \ldots \{E - \lambda_k\} = 0$$

and this solution is

$$u_n = \overline{C}_1 \lambda_1^n + \overline{C}_2 \lambda_2^n + \ldots + \overline{C}_k \lambda_k^n$$

where the $\overline{C}_j$ are constants determined by the initial conditions. If a root is repeated m times, its coefficient is a polynomial in n of order $m - 1$.

Thus, for

$$\{E - \lambda_1\}^{m+1}\{E - \lambda_2\}u_n = 0$$

the solution would be

$$u_n = (\overline{C}_{10} + \overline{C}_{11}n + \ldots + \overline{C}_{1m}n^m)\lambda_1^n + \overline{C}_2\lambda_2^n$$

If we consider forms of equation (8) for which the C_j are real, any complex root of the characteristic equation must have a conjugate. Such cases can be treated as in the following example. Let

$$\{E - \overline{\alpha} - i\overline{\beta}\}\{E - \overline{\alpha} + i\overline{\beta}\}\{E - \lambda_3\} = 0$$

be the characteristic equation. Then setting

$$r^2 = \overline{\alpha}^2 + \overline{\beta}^2$$

$$\theta = \tan(\overline{\beta}/\overline{\alpha})$$

we can write the equation in the form

$$u_n = \overline{C}_1 r^n \cos n\theta + \overline{C}_2 r^n \sin n\theta + \overline{C}_3\lambda_3^n$$

The particular solution to equation (8) is easily expressed when $F(n)$ is a polynomial in n or any sum of terms having the form $\overline{\gamma}^n$ where $\overline{\gamma}$ is any complex constant. The latter solution is given by Boole's first rule and is of particular value to us. Boole showed

$$\{G(E)\}\overline{\gamma}^n = \overline{\gamma}^n G(\overline{\gamma}) \tag{10}$$

where the notation reads $G(E)$ operating on $\overline{\gamma}^n$ equals the product of $\overline{\gamma}^n$ and $G(\overline{\gamma})$. Thus, the <u>particular</u> solution of

$$\{E^k + C_1E^{k-1} + \ldots + C_k\}u_n = Ae^{\mu hn}$$

is

$$(u_n)_p = \left\{\frac{1}{E^k + C_1E^{k-1} + \ldots + C_k}\right\}Ae^{\mu hn}$$

or

$$(u_n)_p = Ae^{\mu hn}/(e^{\mu hk} + C_1e^{\mu h(k-1)} + \ldots + C_k)$$

It is valid for complex μ and its evaluation does <u>not</u> require that the roots to the characteristic equation be known.

The solution of simultaneous linear difference equations with constant coefficients offers no particular difficulty. For example, the two equations

$$u_{n+1} + 2v_n = 2^{-n}$$

$$u_{n+1} - \frac{1}{2} u_n + v_{n+1} = 0$$

in operational form become

$$Eu_n + 2v_n = 2^{-n}$$

$$\left\{E - \frac{1}{2}\right\} u_n + Ev_n = 0$$

or in matrix notation

$$\begin{bmatrix} E & 2 \\ E - \frac{1}{2} & E \end{bmatrix} \begin{bmatrix} u_n \\ v_n \end{bmatrix} = \begin{bmatrix} 2^{-n} \\ 0 \end{bmatrix}$$

The characteristic equation for the set is

$$\begin{bmatrix} E & 2 \\ E - \frac{1}{2} & E \end{bmatrix} = E^2 - 2E + 1 = (E - 1)^2 = 0$$

and the general solution for u_n is

$$u_n = (\bar{C}_o + \bar{C}_1 n)(+1)^n + \left\{\frac{E}{(E - 1)^2}\right\} 2^{-n}$$

or

$$u_n = \bar{C}_o + \bar{C}_1 n + 2^{(1-n)}$$

THE REPRESENTATIVE DIFFERENTIAL EQUATIONS

Development

Fundamentally, the representative differential equations, for which the conclusions throughout this report exactly apply, are the set of simultaneous, linear, first-order, differential equations with constant coefficients ($w' \equiv dw/dx$)

$$\left.\begin{aligned} w_1' &= a_{11}w_1 + a_{12}w_2 \ldots + f_1(x) \\ w_2' &= a_{21}w_1 + a_{22}w_2 \quad\quad + f_2(x) \\ w_3' &= a_{31}w_1 + a_{32}w_2 \quad\quad + f_3(x) \\ &\cdot \quad\quad\quad\quad\quad\quad\quad\quad \cdot \\ &\cdot \quad\quad\quad\quad\quad\quad\quad\quad \cdot \\ &\cdot \quad\quad\quad\quad\quad\quad\quad\quad \cdot \end{aligned}\right\} \tag{11}$$

or any group of higher order differential equations which reduce to such a set. Although the subsequent analysis rigorously applies to equations (11), it is unnecessary, for the purpose of studying the accuracy and stability of a numerical method, to consider them in such a general form. In fact, we will see in the next few sections that the stability and accuracy that result from integrating equations (11) by any of the numerical methods considered herein are completely independent of the elements a_{ij} except as these elements determine the roots of the characteristic equation (the eigenvalues). In other words, if equations (11) are integrated by some polynomial numerical method for any number of steps and then uncoupled (put in the form of eigenvectors), or if equations (11) are first uncoupled and then integrated using the same method and step location, the results will (except for roundoff error) be identical regardless of whether or not they are correct or the numerical method is stable.

Exact solution using the Laplace transform.- To begin with, let us consider the exact solution to equations (11) as it is derived by means of the Laplace transform. If $\overline{w}(s)$ is the Laplace transform of $w(x)$

$$\overline{w}(s) = \int_0^\infty e^{-sx}w(x)dx \tag{12}$$

and

$$s\overline{w}(s) = w(0) + \int_0^\infty e^{-sx}w'(x)dx \tag{13}$$

Multiplying both sides of equations (11) by e^{-sx} and integrating with respect to x from 0 to ∞ gives

$$\begin{aligned} (a_{11} - s)\overline{w}_1 + a_{12}\overline{w}_2 + \ldots &= -w_1(0) - \overline{f}_1(s) \\ a_{21}\overline{w}_1 + (a_{22} - s)\overline{w}_2 &= -w_2(0) - \overline{f}_2(s) \\ a_{31}\overline{w}_1 + a_{23}\overline{w}_2 &= -w_3(0) - \overline{f}_3(s) \\ \cdot \quad\quad\quad\quad &\quad\quad \cdot \\ \cdot \quad\quad\quad\quad &\quad\quad \cdot \\ \cdot \quad\quad\quad\quad &\quad\quad \cdot \end{aligned}$$

a set of simultaneous, linear equations for $\bar{w}_1, \bar{w}_2, \ldots \bar{w}_m, \ldots$. The determinant of these equations is a polynomial in s; and equating this polynomial to zero results in the characteristic equation for the differential equations (11). Let the roots to this characteristic equation be $\sigma_1, \sigma_2, \ldots \sigma_m, \ldots$. (For simplicity, the argument proceeds as if none of the roots is multiple, but this restriction is not at all necessary.) From the well-developed theory of the Laplace transform, the complementary solution to the equations (11) can at once be written

$$w_l = C_{1l}e^{\sigma_1 x} + C_{2l}e^{\sigma_2 x} + \ldots + C_{ml}e^{\sigma_m x} + \ldots$$

or since $x = nh$

$$w_l = C_{1l}(e^{\sigma_1 h})^n + C_{2l}(e^{\sigma_2 h})^n + \ldots + C_{ml}(e^{\sigma_m h})^n + \ldots \qquad (14)$$

where C_{ml} are constants depending on the initial conditions and the nature of the functions $f_i(x)$.

As is well known, the above developments can be viewed in a slightly different light. Write equations (11) in the matrix form

$$\vec{w}' = [A]\,\vec{w} + \vec{f} \qquad (15)$$

where $\rightarrow$ defines a column vector and the [A] matrix is defined by

$$[A] = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \ldots \\ a_{21} & a_{22} & a_{23} & \ldots \\ a_{31} & a_{32} & a_{33} & \ldots \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \end{bmatrix} \qquad (16)$$

The complementary solution (14) immediately follows where the σ_m are the eigenvalues of [A].

Numerical solution by a predictor or an implicit corrector.- Now let us solve equations (11) using a single difference-differential equation. In practice this would be a predictor or corrector with the implicit relationship somehow brought to equality. In the next section the generalization of the following to actual predictor-corrector methods will be discussed. The analysis is presented in this order because of the simplicity of the development in this section relative to that in the next.

Consider the difference-differential equation (recall that $J = k + 1 - j$)

$$u_{n+k} = \sum_{j=1}^{k+1} \beta_j u_{n+J} + h \sum_{j=1}^{k+1} \beta'_j u'_{n+J} \qquad (17)$$

which is a predictor if $\beta_1 = \beta_1' = 0$. Introduce the operator E (see eq. (9)) and rearrange. There results

$$\left\{ h \sum_{j=1}^{k+1} \beta_j' E^j \right\} u_n' = \left\{ E^k - \sum_{j=1}^{k+1} \beta_j E^j \right\} u_n$$

or

$$u_n' = \left\{ \frac{E^k - \sum_{j=1}^{k+1} \beta_j E^j}{h \sum_{j=1}^{k+1} \beta_j' E^j} \right\} u_n \equiv S u_n \tag{18}$$

which defines the operator S in terms of the operator E. Thus, if the difference-differential equation is applied to the differential equations (11), there results at the nth step the set of linear difference equations

$$\begin{aligned}
(a_{11} - S)w_{1n} + a_{12}w_{2n} + \cdots &= -f_1(nh) \\
a_{21}w_{1n} + (a_{22} - S)w_{2n} &= -f_2(nh) \\
a_{31}w_{1n} + a_{32}w_{2n} &= -f_3(nh) \\
\vdots & \quad \vdots
\end{aligned}$$

Clearly, the roots to the characteristic equation in S are once again the eigenvalues of the matrix [A]. In other words

$$(S - \sigma_1)(S - \sigma_2) \cdots (S - \sigma_m) \cdots = 0$$

This leads to the rather remarkable result that the numerical method isolates the individual roots of the exact solution and operates on each of them individually as if the others were not even present!

Recall the definition of S and construct the "subcharacteristic" equation for E in terms of σ_m. Thus

$$E^k - \sum_{j=1}^{k+1} (\beta_j + \sigma_m h \beta_j') E^j = 0$$

which has the "subroot" structure

$$(E - \lambda_{1m})(E - \lambda_{2m}) \cdots = 0$$

One of these roots will approximate the Taylor series expansion of the term $e^{\sigma_m h}$ with a truncation error appropriate to the degree of the polynomial embedded in the difference-differential equation. This root is commonly referred to as the principal root, designated λ_{1m}, and can always be expressed in the form

$$\lambda_{1m} = 1 + \sigma_m h + \frac{1}{2}\sigma_m^2 h^2 + \frac{1}{6}\sigma_m^3 h^3 + \ldots = e^{\sigma_m h} \pm \tag{19}$$

where the $\pm$ after the $e^{\sigma_m h}$ indicates the existence of a truncation error. The remaining roots $\lambda_{2m}, \lambda_{3m}, \ldots$ are spurious. They are introduced by the numerical method and depend on the choice of the difference-differential equation, both as to number and magnitude.

In summary, the exact solution to the differential equations is

$$w_{ln} = C_{1l}(e^{\sigma_1 h})^n + \ldots + C_{ml}(e^{\sigma_m h})^n + \ldots \tag{20}$$

and the exact solution to the difference equations is

$$\begin{aligned} w_{ln} = {} & C_{11,l}(e^{\sigma_1 h \pm})^n + \ldots + C_{1m,l}(e^{\sigma_m h \pm})^n + \ldots \\ & + C_{21,l}(\lambda_{21})^n + \ldots + C_{2m,l}(\lambda_{2m})^n + \ldots \\ & + C_{31,l}(\lambda_{31})^n + \ldots + C_{3m,l}(\lambda_{3m})^n + \ldots \\ & \vdots \end{aligned} \tag{21}$$

The results include complex roots and may be extended to multiple roots.

Numerical solution using both a predictor and corrector.- The conclusions drawn in the previous section were shown to be true when a single difference-differential equation is used to integrate a set of simultaneous differential equations. We now show the same conclusions hold for a predictor-one-corrector method.

To begin with, apply the incomplete predictor-corrector combination

$$\left.\begin{aligned} u_{n+k}^{(1)} &= \sum_{j=2}^{k+1} (\alpha_j u_{n+J} + \alpha_j' h u_{n+J}') \\ u_{n+k} &= \beta_1' h u_{n+k}^{(1)\prime} + \sum_{j=2}^{k+1} (\beta_j u_{n+J} + \beta_j' h u_{n+J}') \end{aligned}\right\} \tag{22}$$

to the single equation

$$u' = \sigma u \tag{23}$$

There results, in operator notation,

$$-\{E^k\}u_n^{(1)} + \left\{\sum_{j=2}^{k+1} \alpha_j E^J\right\}u_n + \sigma\left\{h \sum_{j=2}^{k+1} \alpha_j' E^J\right\}u_n = 0$$

$$\lambda\{h\beta_1' E^k\}u_n^{(1)} - \left\{E^k - \sum_{j=2}^{k+1} \beta_j E^J\right\} u_n + \sigma\left\{h \sum_{j=2}^{k+1} \beta_j' E^J\right\}u_n = 0$$

Define a new set of operators

$$\left.\begin{aligned} s &= -\frac{E^k}{h\sum_{j=2}^{k+1} \alpha_j' E^J} \qquad y = \frac{\sum_{j=2}^{k+1} \alpha_j E^J}{h\sum_{j=2}^{k+1} \alpha_j' E^J} \\ t &= \frac{\beta_1' E^k}{\sum_{j=2}^{k+1} \beta_j' E^J} \qquad z = \frac{E^k - \sum_{j=2}^{k+1} \beta_j E^J}{h\sum_{j=2}^{k+1} \beta_j' E^J} \end{aligned}\right\} \tag{24}$$

and the matrix equation

$$\begin{bmatrix} s & \sigma - y \\ \sigma t & \sigma - z \end{bmatrix} \begin{bmatrix} u_n^{(1)} \\ u_n \end{bmatrix} = 0 \tag{25}$$

follows. The characteristic equation is simply

$$s(\sigma - z) - \sigma t(\sigma - y) = 0 \tag{26}$$

Substituting for s, t, y, and z one can determine the root structure for E in terms of the eigenvalue of the single equation (23). This root structure determines the accuracy and stability of the method defined by equations (22) as it applies to a single differential equation.

For the two equations

$$\left.\begin{aligned} u_1' &= a_{11}u_1 + a_{12}u_2 \\ u_2' &= a_{21}u_1 + a_{22}u_2 \end{aligned}\right\} \tag{27}$$

the matrix expands to

$$\begin{bmatrix} s & 0 & a_{11} - x & a_{12} \\ 0 & s & a_{21} & a_{22} - x \\ a_{11}t & a_{12}t & a_{11} - y & a_{12} \\ a_{21}t & a_{22}t & a_{21} & a_{22} - y \end{bmatrix} \begin{bmatrix} u_{1n}^{(1)} \\ u_{2n}^{(1)} \\ u_{1n} \\ u_{2n} \end{bmatrix} = 0 \tag{28}$$

If the eigenvalues of [A], where

$$[A] = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

are σ_1 and σ_2, one can use the identities

$$\sigma_1\sigma_2 = a_{11}a_{22} - a_{21}a_{12}$$

$$\sigma_1 + \sigma_2 = a_{11} + a_{12}$$

$$\sigma_1^2 + \sigma_2^2 = a_{11}^2 + 2a_{11}a_{21} + a_{22}^2$$

and show that the determinant of the matrix in equation (28) reduces to the product

$$\prod_{j=1}^{2} \begin{vmatrix} s & \sigma_j - y \\ \sigma_j t & \sigma_j - z \end{vmatrix} \tag{29}$$

having the characteristic equation

$$[s(\sigma_1 - z) - t\sigma_1(\sigma_1 - y)][s(\sigma_2 - z) - t\sigma_2(\sigma_2 - y)] = 0 \tag{30}$$

This shows that when the method defined by equations (22) is applied to two coupled differential equations, the stability and accuracy of the result depend entirely on the eigenvalues of [A] and are not related (in any other way) to the magnitude of the individual elements.

The generalization of the above to larger groups of simultaneous equations depends upon a proof of the conjecture

$$\begin{vmatrix} s & 0 & 0 & \cdots & a_{11}-x & a_{12} & a_{13} & \cdots \\ 0 & s & 0 & \cdots & a_{21} & a_{22}-x & a_{23} & \cdots \\ 0 & 0 & s & \cdots & a_{31} & a_{32} & a_{33}-x & \cdots \\ \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot & \\ a_{11}t & a_{12}t & a_{13}t & \cdots & a_{11}-y & a_{12} & a_{13} & \cdots \\ a_{21}t & a_{22}t & a_{23}t & \cdots & a_{21} & a_{22}-y & a_{23} & \cdots \\ a_{31}t & a_{32}t & a_{33}t & \cdots & a_{31} & a_{32} & a_{33}-y & \cdots \\ \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot & \end{vmatrix} = \prod_j [s(\sigma_j - z) - t\sigma_j(\sigma_j - y)] \qquad (31)$$

where the σ_j are the eigenvalues of the A matrix (see eq. (16)). Although the author knows[5] of no formal proof of (31), its equality can easily be demonstrated, within the bounds of numerical error, by means of a digital computer. Simply evaluate and compare the left- and right-hand sides of (31) (subroutines that calculate the eigenvalues of real matrices are standard equipment) using random numbers for the terms t,s,y,z and the elements a_{ij}. This "numerical proof" has been carried out for a variety of determinants with order 2 through 10.

The preceding analysis leads to a final hypothesis. First consider the following: Let

$$[B] = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1m} \\ b_{21} & b_{22} & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \\ b_{m1} & \cdots & & b_{mm} \end{bmatrix}, \qquad [C] = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1m} \\ \lambda_{21} & \lambda_{22} & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \\ \lambda_{m1} & \cdots & & \lambda_{mm} \end{bmatrix}$$

$$\vec{w} = \begin{vmatrix} w_1(x) \\ w_2(x) \\ \cdot \\ \cdot \\ \cdot \\ w_m(x) \end{vmatrix}, \qquad \vec{f} = \begin{vmatrix} f_1(x) \\ f_2(x) \\ \cdot \\ \cdot \\ \cdot \\ f_m(x) \end{vmatrix}$$

[5]Knew at this writing (see end of this section).

where b_{ij} and λ_{ij} are constants. Then if

$$\vec{w}' = [A]\vec{w} + [B]\vec{f} \tag{32}$$

where

$$[A] = [B][C][B]^{-1} \tag{33}$$

$\vec{u}$ is independent of [B] where

$$\vec{u} = [B]^{-1}\vec{w} \tag{34}$$

The proof follows immediately by multiplying equation (32) by $[B]^{-1}$. Now reduce equation (32) to a set of difference equations by applying any combination of linear, difference-differential equations representing complete or incomplete, predictor, multiple-corrector methods. Solve these equations for $\vec{w}$ and apply equation (34). Regardless of the choice of [B], we hypothesize that the numerical value of $\vec{u}$ will be identical at every step except for roundoff error. Again a "numerical proof" of this conjecture was obtained by solving five linear, simultaneous equations with real and complex eigenvalues for several choices of [C] and $\vec{f}$ and a variety of [B]. With double-precision arithmetic the differences in $\vec{u}$ caused by the various choices of [B] could safely be attributed to roundoff.

During the preparation of this report a simple analytical proof of the above was presented to the author. This proof, which changes the hypothesis into a theorem, was prepared by Dr. William A. Mersman[6] and proceeds as follows.

Combine equation (15) with the predictor-corrector sequence defined by equations (22). Using [I] to designate the unit matrix, we have

$$\vec{w}^{(1)}_{n+k} = \sum_{j=2}^{k+1} \left\{ (\alpha_j[I] + h\alpha'_j[A])\vec{w}_{n+J} + h\alpha_j[I]\vec{f}_{n+J} \right\}$$

$$\vec{w}_{n+k} = \beta'_1 h[A]\vec{w}^{(1)}_{n+k} + \sum_{j=2}^{k+1} \left\{ (\beta_j[I] + h\beta'_j[A])\vec{w}_{n+J} \right\} + h \sum_{j=1}^{k+1} \beta'_j \vec{f}_{n+J}$$

Eliminating $\vec{w}^{(1)}_{n+J}$ and introducing an operational notation, we derive the equation

$$\left\{ E^k[I] - \sum_{j=2}^{k+1} (\beta_j[I] + h(\beta'_1\alpha_j + \beta'_j)[A] + h^2\beta'_1\alpha'_j[A]^2)E^J \right\} \vec{w}_n$$

$$= h \left\{ \sum_{j=1}^{k+1} \beta'_j[I]E^J + \beta'_1 h[A] \sum_{j=2}^{k+1} \alpha'_j E^J \right\} \vec{f}_n$$

[6]Chief, Problem Definition and Analysis Branch.

Define the terms

$$\rho(E) = E^k - \sum_{j=2}^{k+1} \beta_j E^j$$

$$\sigma(E) = -h \sum_{j=2}^{k+1} (\beta_1' \alpha_j + \beta_j') E^j$$

$$\tau(E) = -h^2 \beta_1' \sum_{j=2}^{k+1} \alpha_j' E^j$$

and the characteristic equation (in terms of E) for the system of difference equations results from setting the determinant of the matrix

$$[P] \equiv [\rho(E)[I] + \sigma(E)[A] + \tau(E)[A]^2]$$

equal to zero. As is well known (see any text on linear algebra) both the determinant and the eigenvalues of $[P]$ are identical to those for $[P^*]$ where

$$[P^*] = [B][P][B]^{-1}$$

for all nonsingular $[B]$. Furthermore, whether or not $[A]$ has multiple eigenvalues, $[B]$ can always be chosen so that

$$[B][A][B]^{-1} = [T]$$

where $[T]$ is upper triangular and the elements of its diagonal are, of course, the eigenvalues of $[A]$. But since

$$[B][A]^2[B] = [B][A][B]^{-1}[B][A][B]^{-1} = [T]^2$$

this choice of $[B]$ also makes $[P^*]$ upper triangular. It follows that

$$|P| = |P^*| = \prod_i (\rho(E) + \lambda_i \sigma(E) + \lambda_i^2 \tau(E))$$

where the λ_i are the distinct eigenvalues of $[A]$. This proves the stated hypothesis.

Discussion

The results just presented show that a predictor-corrector process applied to a set of simultaneous, linear, differential equations with constant coefficients automatically "detects" the eigenvalues of the differential equations and the success or failure of the numerical method is measured by

1. Its accuracy in resolving the eigenvalue for which it is most inaccurate

2. Its stability with respect to the eigenvalue for which it is most unstable

Therefore, to study and compare numerical methods as they apply to the solution of systems of differential equations (11), it is sufficient to study and compare them as they apply to a single differential equation. In the case of nonmultiple roots, this single equation is simply

$$u' = \lambda u + f(x) \tag{35}$$

where λ may be complex and represents the "worst" eigenvalue of the system.

That it is generally impossible to estimate the magnitude of the eigenvalues by the magnitude of the individual elements of a matrix is shown by the following examples in which all three matrices have (except for the limitation of 8 significant figures) the same eigenvalues, -1, -10, and -100.

$$\left.\begin{bmatrix} -2451.5752 & 9523.1044 & -2225.7133 \\ -765.51692 & 2970.5032 & -695.60020 \\ -690.98884 & 2671.4410 & -629.92802 \end{bmatrix} \right\} \tag{36}$$

$$\begin{bmatrix} -41.774674 & 65.261307 & 15.813105 \\ 22.518472 & -39.291311 & -11.264427 \\ 40.012615 & -75.965121 & -29.934014 \end{bmatrix}$$

$$\begin{bmatrix} -95.492090 & -118.40717 & 6.8224432 \\ -8.6303246 & -12.144652 & 0.26226723 \\ -85.039375 & -114.33994 & -3.3632580 \end{bmatrix}$$

This can be an important consideration in some programs that attempt to control step size automatically using norms based on the elements in individual rows or columns (see, e.g., ref. 8).

Equation (35) is, basically, the representative equation used throughout this paper. (For practical reasons, a more convenient expression will actually be analyzed, see eq. (37).) In most papers the representative equation is presented as (1a), of which equation (35) is a very special form. Nearly always, however, commitments about the nonlinear equation are based on local linearization for the simple reason that the nonlinear form is intractable. In this approach, the reader may regard equation (35) as characterizing equation (1a) when λ symbolizes the average value of $\partial F/\partial u$ over some interval (ref. 4, p. 207), or the Lipshitz constant over the same interval (ref. 5, p. 216).

Accuracy.- For the purpose of studying the accuracy of a difference-differential approximation we choose equation (35) as the representative form and permit λ to be complex throughout the analysis. Equation (35) is still unnecessarily general, however, since, for testing the accuracy of a numerical method, we can replace $f(x)$ in some interval with its equivalent Fourier series and draw out from the latter the term $A_n e^{i\mu_n x}$ which represents the highest frequency in $f(x)$ we wish to resolve by the numerical computation in

the chosen interval. All lower frequencies are automatically more accurately determined. Hence, for studying accuracy, equation (35) may be replaced by the simpler form

$$u' = \lambda u + Ae^{\mu x} \tag{37}$$

which has the general solution

$$u = ce^{\lambda x} + \frac{Ae^{\mu x}}{\mu - \lambda} \tag{38}$$

where c is an arbitrary constant and both λ and μ can be complex. A measure of accuracy of any numerical method is given by its worst approximation to either term in equation (38) in the presence of the other.

Stability.- Equation (35) is also used as the representative form for analyzing the stability of difference-differential approximations. For such studies the term $f(x)$ can be omitted since it does not affect the stability. Whether or not a method is stable depends upon the magnitude of the roots to the characteristic equation for the difference equations which the method generates when combined with the representative form. There is a large and important subset to equations (11), associated with positive definite forms, for which all the eigenvalues are real and positive, that is, for which the λ in equation (35) is real and negative. For this subset the stability criterion can be developed from the simple normalized equation

$$u' = -u \tag{39}$$

in which the independent variable is real. This form is often used (see refs. 2, 3, 9), and for a numerical method to be stable when applied to equations (11), it must certainly be stable for equation (39). But it is not sufficient, and in a later section some methods are shown to be stable for equation (39) but not for equations (11).

When the eigenvalues of equations (11) are complex, the λ in equation (35) becomes complex. The generalization of the numerical stability criterion, under these conditions is presented in reference 10. For an analysis that applies this generalized criterion to several numerical methods, see reference 11.

Two ways of approaching the study of the stability problem for complex eigenvalues are suggested. One is to consider the simultaneous equations

$$\left.\begin{aligned} w_1' &= w_2 \\ w_2' &= -w_1 + 2\nu w_2 \end{aligned}\right\} \tag{40}$$

where ν is always real. The advantage of this way is a given set of difference-differential equations can actually be used numerically to check the results. The necessary and sufficient condition that a numerical method be stable for equations (11) is that it be stable for equations (40) in which ν is real. When $|\nu| \geq 1$, equations (40) have real eigenvalues. When

$|\nu| < 1$, equations (40) have two conjugate complex eigenvalues, and the case $\nu = 0$ represents the condition for pure imaginary eigenvalues.

A second way to study the general stability problem is to consider the normalized equation

$$u' = e^{i\omega} u \tag{41}$$

which is a special form of equation (35) where

$$\lambda = e^{i\omega} \tag{42}$$

and $f(x) = 0$. The study of equation (41) avoids much of the algebra required to analyze equations (40) but necessitates the use of complex arithmetic. However, in modern computer languages the latter is not a serious disadvantage. The necessary and sufficient condition that a numerical method be stable for equations (11) is that it be stable for equation (42) for all values of ω.

THE GENERAL ANALYSIS OF INCOMPLETE, MULTISTEP, PREDICTOR, ONE-CORRECTOR METHODS

General Discussion

The following study applies to all numerical methods that integrate ordinary differential equations making use of a predictor followed by just one corrector, with both the function and its derivative in the same final family. (A study of complete methods is presented in a later section.) These incomplete methods include, for example, Hamming's method, the Adams-Bashforth-Moulton methods, Milne's method, etc. In fact, the result of using any predictor in table I(a) followed by any corrector in table I(b) can readily be determined both as to accuracy and stability. The analysis permits one to calculate exactly (except for roundoff error) what a digital computer will produce after any number of steps when any one of the methods is applied to equation (37).

The operational form.- Consider a predictor-corrector sequence forming the first family $u^{(1)}$

$$u^{(1)}_{n+k} = \sum_{j=2}^{k+1} (\alpha_j u_{n+J} + \alpha'_j h u'_{n+J}) , \qquad J = k + 1 - j \tag{43}$$

and the final family u

$$u_{n+k} = \bar{\beta}_1 u^{(1)}_{n+k} + \bar{\beta}'_1 h u^{(1)'}_{n+k} + \sum_{j=2}^{k+1} (\bar{\beta}_j u_{n+J} + \bar{\beta}'_j h u'_{n+J}) \tag{44}$$

Notice that by multiplying the top equation by $\bar{\beta}_1$ and subtracting the result from the second equation, we can form a third equation

$$u_{n+k} = \beta_1' h u_{n+k}^{(1)\prime} + \sum_{j=2}^{k+1} (\beta_j u_{n+J} + \beta_j' h u_{n+J}') \tag{45}$$

This indicates the lack of uniqueness in the values for the coefficients in the difference-differential equations for combined predictor-corrector methods.

The following is a simple example of this lack of uniqueness. An Euler predictor followed by a modified Euler corrector is generally written in the form

$$\left.\begin{aligned} u_{n+1}^{(1)} &= u_n + h u_n' \\ u_{n+1} &= u_n + \frac{h}{2}\left(u_{n+1}^{(1)\prime} + u_n'\right) \end{aligned}\right\} \tag{46}$$

Clearly, the analytical result of such a method is not altered by the modification

$$\left.\begin{aligned} u_{n+1}^{(1)} &= u_n + h u_n' \\ u_{n+1} &= c_o\left(u_{n+1}^{(1)} - u_n - h u_n'\right) + u_n + \frac{h}{2}\left(u_{n+1}^{(1)\prime} + u_n'\right) \end{aligned}\right\} \tag{47}$$

where c_o is an arbitrary constant. As an example, setting $c_o = 1/2$ gives the combination

$$\left.\begin{aligned} u_{n+1}^{(1)} &= u_n + h u_n' \\ u_{n+1} &= \frac{1}{2}\left(u_{n+1}^{(1)} + u_n + h u_{n+1}^{(1)\prime}\right) \end{aligned}\right\} \tag{48}$$

If the corrector is used only once, equations (46) and (48) are identical in accuracy and stability when applied to equations (11).

Returning to our development, and choosing, without loss of generality, equations (43) and (45), we introduce the operator E and the representative equation (37). This combination produces the linear difference equations which can be written in matrix form

$$\begin{bmatrix} E^k & -\sum_{j=2}^{k+1} (\alpha_j + \lambda h \alpha_j') E^J \\ -\lambda h \beta_1' E^k & E^k - \sum_{j=2}^{k+1} (\beta_j + \lambda h \beta_j') E^J \end{bmatrix} \begin{bmatrix} u_n^{(1)} \\ u_n \end{bmatrix} = A h e^{\mu h n} \begin{bmatrix} \sum_{j=2}^{k+1} \alpha_j' e^{J\mu h} \\ \sum_{j=1}^{k+1} \beta_j' e^{J\mu h} \end{bmatrix} \tag{49}$$

Solving for u_n, one can divide out E^k from the left column with the result

$$\left\{E^k - \sum_{j=2}^{k+1} (L_{1j} + \lambda h L_{2j} + \lambda^2 h^2 L_{3j}) E^J\right\} u_n = hAe^{\mu hn} \sum_{j=1}^{k+1} (R_{1j} + \lambda h R_{2j}) e^{J\mu h} \quad (50)$$

where

$$\left.\begin{aligned} L_{1j} &= \beta_j & L_{11} &= L_{21} = L_{31} = 0 \\ L_{2j} &= \alpha_j \beta_1' + \beta_j' & R_{1j} &= \beta_j' \\ L_{3j} &= \alpha_j' \beta_1' & R_{2j} &= \alpha_j' \beta_1' = L_{3j} \end{aligned}\right\} \quad (51a)$$

Equation (50) is the operational form of multistep, two-iteration methods and the L and R are the coefficients in the operational form. Coefficients in the operational forms of a variety of methods are given in table II.

Equation (51a) can be inverted. That is, the coefficients in the difference-differential equations can be expressed explicitly in terms of the coefficients in the operational form, provided $R_{11} \neq 0$. Thus

$$\left.\begin{aligned} \alpha_j &= (L_{2j} - R_{1j})/R_{11} \\ \alpha_j' &= R_{2j}/R_{11} \\ \beta_j &= L_{1j} \\ \beta_j' &= R_{1j} \end{aligned}\right\} \quad (51b)$$

Fortunately, cases for which $R_{11} = 0$ appear to have little practical use. This inversion is the key to the construction of optimum numerical methods, since both stability and accuracy depend fundamentally only on the operational form (discussed later), not on the difference-differential equations. We now seek only to analyze given methods, not to develop new ones.

Before proceeding, it is convenient to introduce the following two definitions.

$$\left.\begin{aligned} DE(E) &= \text{the coefficient of } u_n \text{ in any operational form} \\ NU &= (\mu - \lambda) \text{ times the coefficient of } Ae^{\mu hn} \text{ in any operational form} \end{aligned}\right\} \quad (52)$$

In particular, for multistep, two-iteration methods

$$DE(E) = E^k - \sum_{j=2}^{k+1} (L_{1j} + \lambda h L_{2j} + \lambda^2 h^2 L_{3j}) E^J \quad (53a)$$

$$NU = h(\mu - \lambda) \sum_{j=1}^{k+1} (R_{1j} + \lambda h R_{2j}) e^{J\mu h} \tag{53b}$$

The particular solution.- Referring to the section on the operational solution of difference equations, we can immediately write the particular solution in the form

$$u_{n_p} = \frac{hAe^{\mu hn} \sum_{j=1}^{k+1} (R_{1j} + \lambda h R_{2j}) e^{J\mu h}}{e^{k\mu h} - \sum_{j=2}^{k+1} (L_{1j} + \lambda h L_{2j} + \lambda^2 h^2 L_{3j}) e^{J\mu h}}$$

or, using the definitions in equations (53),

$$u_{n_p} = \left(\frac{Ae^{\mu hn}}{\mu - \lambda}\right) \frac{NU}{DE(e^{\mu h})} \tag{54}$$

The complementary solution.- The explicit evaluation of the complementary solution requires a knowledge of the roots to the characteristic equation. In the case under consideration, the characteristic equation is

$$E^k - \sum_{j=2}^{k+1} (L_{1j} + \lambda h L_{2j} + \lambda^2 h^2 L_{3j}) E^J = 0$$

or, using equation (53a),

$$DE(E) = 0. \tag{55a}$$

Let the roots to equation (55) be such that

$$(E - \lambda_1)(E - \lambda_2)(E - \lambda_3) \ldots = 0 \tag{55b}$$

The complementary solution is then

$$u_{n_c} = c_1(\lambda_1)^n + c_2(\lambda_2)^n + c_3(\lambda_3)^n + \ldots \tag{56}$$

where $c_1, c_2, c_3, \ldots$ are constants fixed by the initial conditions, and, conventionally, λ_1 is the principal root while $\lambda_2, \lambda_3, \ldots$ are all spurious roots introduced by the particular numerical method.

Accuracy

The error in the particular solution.- Comparing equations (38) and (54), one can now derive the error in the particular solution. Defining the error term

$$\overline{er}_\mu = \frac{\text{Exact particular solution of difference equation}}{\text{Exact particular solution of differential equation}} - 1 \tag{57}$$

it follows that

$$\overline{er}_\mu = \frac{NU - DE(e^{\mu h})}{DE(e^{\mu h})} \tag{58}$$

Introducing the values of NU and $DE(e^{\mu h})$ into the numerator of (58), and collecting coefficients of λh, gives

$$\overline{er}_\mu = \frac{\left[-e^{\mu hk} + \sum_{j=1}^{k+1} (L_{1j} + \mu h R_{1j}) e^{J\mu h}\right] + \lambda h \left[\sum_{j=1}^{k+1} (L_{2j} - R_{1j} + \mu h R_{2j}) e^{J\mu h}\right]}{DE(e^{\mu h})} \tag{59}$$

The coefficient of the $\lambda^2 h^2$ term is zero since $R_{2j} = L_{3j}$ (see eq. (51a)).

Equation (59) is the exact error which any incomplete multistep, two-iteration numerical method makes in calculating the particular solution to equation (37). However, since all methods being studied here are polynomial approximations, it is more significant as well as more convenient, to express the errors in powers of h.

Setting

$$e^{\mu h} = \sum_{l=0}^{\infty} \frac{(\mu h)^l}{l!}$$

one can show

$$DE(e^{\mu h}) = 1 - \sum_{j=2}^{k+1} L_{1j} + h\left(\mu k - \mu \sum_{j=2}^{k+1} J L_{1j} - \lambda \sum_{j=2}^{k+1} L_{2j}\right) + O(h^2) \tag{60}$$

As will be shown presently (see eq. (72)), to have any polynomial fit, at all the equalities

$$\left.\begin{aligned} \sum_{j=2}^{k+1} L_{1j} &= 1 \\ \sum_{j=2}^{k+1} [L_{2j} + JL_{1j}] &= k \end{aligned}\right\} \tag{61}$$

must hold. Under these conditions

$$\lim_{h\to o} \frac{DE(e^{\mu h})}{h} = (\mu - \lambda) \sum_{j=2}^{k+1} (j - 1)L_{1j} = (\mu - \lambda) \sum_{j=2}^{k+1} L_{2j} \tag{62}$$

Next, noting

$$\sum_{l=o}^{\infty} \frac{(\mu h)^{l+1} J^l}{l!} = \sum_{l=o}^{\infty} \frac{l(\mu h)^l J^{l-1}}{l!}$$

one can show

$$\overline{er}_\mu = \frac{er_{\mu 1} + er_{\mu 2}}{h(\mu - \lambda)} = \frac{er_\mu}{h(\mu - \lambda)} \tag{63}$$

where

$$er_{\mu 1} = \frac{(\mu h)^l}{l!} \frac{\left\{ \sum_{j=1}^{k+1} [lJ^{l-1}R_{1j} + J^l L_{1j}] - k^l \right\}}{\sum_{j=2}^{k+1} (j - 1)L_{1j}} \tag{64}$$

$$er_{\mu 2} = \frac{\lambda h(\mu h)^l}{l!} \frac{\left\{ \sum_{j=1}^{k+1} [lJ^{l-1}R_{2j} + J^l (L_{2j} - R_{1j})] \right\}}{\sum_{j=2}^{k+1} (j - 1)L_{1j}} \tag{65}$$

Now it is important to notice that $\overline{er}_\mu$ is a "global" error; that is, it is the precise error in the particular solution at a given x no matter how many steps are taken. The local error, or error in the embedded polynomial, such as er_p in equation (4), is repeated n times after n steps. Because $x = nh$, the polynomial error can grow, for a given x, to

approximately $(x)(er_p)/h$. Hence, rather than $\overline{er}_\mu$, the term er_μ represents the accuracy of the local polynomial in a given method.[7]

If a method is to represent a polynomial approximation of order L to the particular solution between $n + k - 1$ and $n + k$, then the terms inside {} in equation (64) must sum to zero for all $l = 0, 1, 2, \ldots, L$ and the terms inside {} in equation (65) must sum to zero for all $l = 0, 1, 2, \ldots, L - 1$. This gives at once the equations that must be satisfied by the coefficients in the operational form for a method to provide a polynomial approximation of a given degree to the particular solution. Specifically,

$$\sum_{j=1}^{k+1} [l(k + 1 - j)^{l-1} R_{1j} + (k + 1 - j)^{l} L_{1j}] = k^{l} , \qquad l = 0, 1, 2, \ldots, L \tag{66}$$

$$\sum_{j=1}^{k+1} [l(k + 1 - j)^{l-1} R_{2j} + (k + 1 - j)^{l} (L_{2j} - R_{1j})] = 0 , \qquad l = 0, 1, 2, \ldots, L - 1 \tag{67}$$

We shall see later (in the discussion of the complementary solution) that the fulfillment of these conditions guarantees a polynomial fit of degree L to both the complementary and particular solutions.

Notice that equations (66) and (67) are independent of λ and μ so the coefficients of R and L can be tabulated once and for all. Table III does just that for any one- through five-step, predictor, one-corrector method. The table can be used both to provide the conditions that a given polynomial be embedded in a method, and, with equations (64) and (65), to find the error in the result. Similar tables were used to calculate the error er_μ for the methods presented in table II.

The error in the complementary solution.- Just how well the complementary solution is approximated by a numerical method depends upon how close the principal root, λ_1, lies to $e^{\lambda h}$, since the analytic solution is proportional to $e^{\lambda x} = (e^{\lambda h})^n$, and the numerical solution is proportional to $(\lambda_1)^n$. Let er_λ be defined by

$$er_\lambda = \lambda_1 - e^{\lambda h} \tag{68}$$

Substituting $e^{\lambda h}$ in equations (55b) and (55a)

$$er_\lambda(e^{\lambda h} - \lambda_2)(e^{\lambda h} - \lambda_3) \ldots = e^{\lambda h k} - \sum_{j=2}^{k+1} e^{J\lambda h} \sum_{m=1}^{3} L_{mj}(\lambda h)^{m-1}$$

[7]The term $(\mu - \lambda)$ in equation (63) can be misleading. Recall that it comes from the expansion of $DE(e^{\mu h})$ in powers of h. As $\mu \to \lambda$, the term $O(h^2)$ in equation (60) takes over, and the error does increase. When $\mu = \lambda$, the solution of the representative equation degenerates and the analysis proceeds along different lines.

or

$$er_\lambda = \frac{-DE(e^{\lambda h})}{\prod_{i=2}^{k}(e^{\lambda h} - \lambda_i)} \tag{69}$$

Again we are interested in finding only the lowest order error term. The value of $\prod_{i=2}^{k}(e^{\lambda h} - \lambda_i)$ for $h = 0$ can be determined as follows:

$$\prod_{i=2}^{k}[1 - (\lambda_i)_{h=0}] = \lim_{E\to 1} \frac{DE(E)}{E - 1}$$

since the principal root must equal one at $h = 0$. This, in turn, reduces to

$$\left[\frac{dDE(E)}{dE}\right]_{E=1} = \left[\frac{d}{dE}\left(E^k - \sum_{j=2}^{k+1} L_{1j}E^j\right)\right]_{E=1}$$

or, since $\sum_{j=2}^{k+1} L_{1j} = 1$,

$$\lim_{h\to 0}\left[\prod_{i=2}^{k}(e^{\lambda h} - \lambda_i)\right] = \sum_{j=2}^{k+1}(j - 1)L_{1j} \tag{70}$$

Note the similarity with equation (62) which appears in the denominator of er_μ. Putting equations (52) and (70) in equation (69), and expanding in powers of λh, one can show

$$er_\lambda = \frac{-\frac{(\lambda h)^l}{l!}\left\{\sum_{j=2}^{k+1}[J^l L_{1j} + lJ^{l-1}L_{2j} + l(l - 1)J^{l-2}L_{3j}] - k^l\right\}}{\sum_{j=2}^{k+1}(j - 1)L_{1j}} \tag{71}$$

The first nonvanishing term in (71) determines the order of the error in the complementary solution. (Notice this is a "local" error, comparable to the term er_μ. In fact, one can show that $er_\lambda = er_\mu$ when $\mu = \lambda$.)

If a method is to provide a polynomial approximation of order L to the complementary solution between $n + k - 1$ and $n + k$, the terms inside {} in equation (71) must sum to zero for all $l = 0, 1, 2, \ldots, L$. This derives the equations

$$\sum_{j=1}^{k+1} [(k+1-j)^{l} L_{1j} + l(k+1-j)^{l-1} L_{2j} + l(l-1)(k+1-j)^{l-2} L_{3j}] = k^{l}$$

$$l = 0, 1, 2, \ldots, L \quad (72)$$

Equations (72) are independent of λ and μ so, again, the coefficients of L can be tabulated once and for all (for one- through five-step methods, see table IV). By means of such tables and equation (71), the errors in the complementary solution of predictor, one-corrector methods can be readily determined. Examples are presented in table II.

Discussion of accuracy.- One can show (by using $R_{2j} = L_{3j}$) that equations (66), (67), and (72) are not independent. Hence, as was mentioned earlier, the satisfaction of equations (66) and (67) is sufficient to guarantee a local polynomial fit of degree L to both the particular and complementary solutions. Thus two sets of conditions must be fulfilled to assure a given accuracy for a predictor, one-corrector process. If another corrector is added, another set of conditions must be met, etc., as will be shown later.

On the other hand, if we wish to present the conditions for a predictor only (or for a corrector only which has somehow been brought into balance), we can set the α_j and α'_j terms in equations (43) and (51a) equal to zero. (In this degenerate case the β values become the coefficients of a predictor with the condition $\beta'_1 = 0$.) Then equation (67) is an identity, and $L_{1j} = \beta_j$, $R_{1j} = \beta'_j$. Equations (66) and (72) are identical, and they both amount to successive columns in the table preceding equation (4) adding up to zero (or, in equation (4), $er_p = 0$ for $l = 0, 1, 2, \ldots, L$). In this case the first unmatched term for $er_{\mu 1}$ or er_{λ} is the same as the "error constant" term given by Henrici on page 223 in reference 5.

In summary, equations (66) and (67) are the most general conditions for accuracy imposed upon the coefficients of two difference-differential equations forming a predictor-corrector sequence under the conditions:

(a) The differential equations are of the form given by equations (11).

(b) Polynomial approximation is used.

(c) The difference-differential equations are of a form represented by equations (43) and (45).

Stability

A general discussion of stability is given in the last section of the report where the Dahlquist criterion is extended to cover all linear, numerical, quadrature formulas with combined effects of predictors, correctors, modifiers, etc. In this section the specific stability of predictor, one-corrector methods is considered.

Some definitions.- To begin with, let us examine some terminology commonly used in discussions of numerical stability. A set of ordinary differential equations is inherently unstable if the real part of one or more of the eigenvalues of equations (11) is positive. There are two classes to consider.

Class 1. The initial or boundary conditions are such that the analytical solution grows exponentially. Then the numerical solution must also grow exponentially and is, therefore, by definition, unstable -- but it is not necessarily inaccurate. If these inherently unstable differential equations are transformed into difference equations, the latter are relatively inaccurate (sometimes referred to as "relatively unstable," see the discussion below on induced instability) if any of the spurious roots are greater in absolute value than the largest principal root.

Class 2. The initial or boundary conditions are such that the destabilizing eigenvectors are suppressed and the exact analytic solution has no terms which grow exponentially. Under these conditions the numerical solution will still eventually increase exponentially, usually because of small truncation errors that excite the inherent unstable terms, but, if for no other reason, because of errors brought about by roundoff. This can be a particularly vicious form of instability and its control requires methods outside the scope of this paper.

When a set of ordinary differential equations is reduced to a set of ordinary difference equations, the latter have an induced instability if the real parts of all the eigenvalues of the differential equations are all less than or equal to zero, but one or more of the eigenvalues of the difference equations has an absolute value greater than one. This instability is obviously associated with the particular form of the difference-differential equation chosen for the computations and can, therefore, be controlled.

Remembering that the eigenvalues of the difference equations are functions of h, we find two ways of providing this control. One is to develop methods that are stable when $h = 0$. Then for small enough[8] h, the method is always stable. The Dahlquist stability theorem applies to this study. The other aspect is to develop methods that are stable for as large a value of h as possible. In order to discuss this problem, we will refer to a stability boundary, $|\lambda h|_c$, which defines a critical value of h, any increase of which causes the absolute value of one or more eigenvalues in the difference equation to exceed unity. Thus a method has induced instability when

$$\left.\begin{aligned} \lambda &= e^{i\omega} \\ \pi/2 \leq \omega &\leq \pi \\ |\lambda h| &> |\lambda h|_c \end{aligned}\right\} \qquad (73)$$

[8]To avoid argument, we either omit the possibility of neutral stability at $h = 0$, or further require that all eigenvalues that lie on the unit circle when $h = 0$ move into it as h starts to increase. A method having the latter property is defined by equations (120).

A discussion of the significance of this stability boundary is presented in a later section entitled "The relationship between accuracy and stability."

A final remark concerns the case when the real parts of all the eigenvalues from the differential equation are less than zero and the absolute values of all the eigenvalues from the difference equations are less than one, but some of the difference eigenvalues are greater in magnitude than the differential eigenvalues. This condition is sometimes referred to as being "relatively unstable." If we are to maintain a consistent definition of stability, this terminology is misleading. Such cases are stable since all solutions, as well as their errors, approach zero with increasing x. They may possibly be relatively inaccurate, however, the possibility occurring when the $f(x)$ terms in equations (11) are negligible and the offending eigenvalues are the least heavily damped. In any case one should be cautious about trusting the numerical solution of the asymptotic decay of a function when the level falls beneath the product of the truncation error of the method and its maximum resolved amplitude.

The unit circle.- Since the induced stability of a method depends upon the magnitude of the roots to the characteristic equation $DE(E) = 0$, a quick, visual, representation of the stability of a method is displayed if we plot all the roots to the characteristic equation on a complex plane. For the construction of these plots let $0 \leq h \leq 1$ and

$$\lambda = e^{i\omega}$$

where $0 \leq \omega \leq \pi$. Describe a unit circle with center at the origin of this plane. In the range $0 \leq \omega \leq \pi/2$ the differential equation is inherently unstable and the principal root, at least, must fall outside the circle for $h > 0$. In the range $\pi/2 \leq \omega \leq \pi$ the differential equation is inherently stable and in this range any point falling outside the circle presents a value of h and λ (in complex form) for which the numerical method has induced instability.

The two extremes: Adams-Moulton methods, Milne methods.- The accuracy of most of the popularly used predictor-corrector formulas has been compromised to avoid induced instability. That this compromise can be resolved in many ways is evident from the number of predictor-corrector methods in common use. However, all of these formulas lie between two extremes which we will refer to as the Milne methods and the Adams-Moulton methods. These extremes have certain identifying features which appear immediately in the stability plots described above, but which can also be traced to coefficients in the operational form and even to coefficients in the difference-differential equations themselves.

The Adams-Moulton methods are defined as those for which all the spurious roots fall on the origin when $h = 0$. The principal root at $h = 0$ must always, of course, fall on the intersection of the unit circle and the positive real axis. When $h = 0$, the characteristic equation (see (53)) reduces to

$$E^k - \sum_{j=2}^{k+1} L_{1j}E^j = 0$$

The necessary conditions that make the spurious roots zero and the principal root one are

$$L_{12} = 1$$

$$L_{1j} = 0 , \quad j = 3, 4, \ldots, k + 1$$

In such a case the characteristic equation reduces to just

$$(E - 1)E^{k-1} = 0$$

which clearly satisfies the conditions. Inspecting equations (51b), we see this means

$$\beta_2 = 1$$

$$\beta_j = 0 , \quad j = 3, 4, \ldots, k + 1$$

Thus, in Adams-Moulton methods, only the value of u nearest the corrected point is given a nonzero weight in the corrector.

In this report methods which have at least one spurious root on the unit circle when $h = 0$ are referred to as Milne methods, and those which have all the spurious roots on the unit circle are referred to as total Milne methods.[9] In the latter case the characteristic equation becomes

$$E^k - \sum_{j=2}^{k+1} L_{1j}E^j = (E - 1)\prod_{j=2}^{k}(E - e^{i\theta_j}) = 0$$

and the θ_j are determined such that the L_{1j} are real and the accuracy is optimum. In these cases $L_{i,k+1} \neq 0$ and, from equations (51), $\beta_{k+1} \neq 0$. Thus, in the total Milne methods the value of u farthest from the corrected point is given a nonzero weight in the corrector.

All multistep methods, with induced stability at $h = 0$, lie between these two extremes. On the one hand they have no spurious roots on the unit circle so they are more likely to be stable for nonzero h than the Milne methods. On the other hand, if some of the roots at $h = 0$ are permitted to lie off the origin - but still within the unit circle - they gain some freedom which can be used to choose the L_{1j}, $j = 2, 3, \ldots, k + 1$ so that they will be more accurate than Adams-Moulton methods with equivalent step number.

[9]What are called total Milne methods in this report are referred to as optimal methods in reference 5. The author prefers the former description since, in practice, stability troubles generally prevent such methods from being optimum.

The relationship between accuracy and stability.- To appreciate the role that induced instability plays in assessing the merits of various methods, we should distinguish carefully between the requirements for stability and accuracy. The error in any method, as it is given by the lowest order, nonvanishing, truncation terms (e.g., the values for er_μ and er_λ in table II), loses its significance when these terms exceed about one tenth. In fact, to rely on such error estimates, the step size should be chosen so that

$$(|\mu h|, |\lambda h|)_{max} < 0.1$$

in methods applied to the integration of

$$u' = \lambda u + Ae^{\mu x}$$

Now one usual requirement for predictor-corrector methods, programmed for general use, is that the induced stability boundary $|\lambda h|_c$, see (73), be as large as possible, generally greater than 0.6. The question immediately arises: Of what value is a method that is stable in a region where it is inaccurate? The answer is supplied if we consider the application of the method to the integration of simultaneous equations. Let us consider a predictor-corrector process for which $|\lambda h|_c = 0.61$. To prohibit induced instability and give an accuracy measured by the values of er_μ and er_λ pertaining to the method, it is certainly sufficient that

$$(|\mu_i h|, |\sigma_i h|) < 0.1 , \quad \text{for all } i \tag{74a}$$

where the μ_i and σ_i are determined by the differential equations (see eqs. (11) and (16)). Although (74a) is a sufficient condition, it is not a necessary one. The necessary conditions for both stability and accuracy (in the sense used above) are that

$$\left.\begin{array}{ll} (1) & (|\mu_i h|, |\sigma_i h|)_{max} < 0.1 \text{ , for all } i \text{ representing those } \mu \text{ and } \sigma \\ & \text{one seeks to calculate to the specified accuracy} \\ (2) & |\sigma_i h|_{max} < |\lambda h|_c \text{ , for all } i \end{array}\right\} \tag{74b}$$

This distinction between accuracy and stability is sometimes important.[10] Perhaps the easiest way to describe the situation for those unfamiliar with it is to give an example. Consider the equations

$$\left.\begin{aligned} w_1' &= -1.38w_1 - 0.81w_2 \\ w_2' &= -2.16w_1 - 1.92w_2 \end{aligned}\right\} \tag{75a}$$

with the initial values

$$\left.\begin{aligned} w_1(0) &= -2.9905 \\ w_2(0) &= 4.0010 \end{aligned}\right\} \tag{75b}$$

[10]For example, the so-called "stiff" equations arising in the study of nonequilibrium fluid flow.

For many purposes the step size $h = 0.2$ is perfectly acceptable for solving this problem in the range, say, $0 \leq x \leq 10$, using a predictor-corrector method for which $|\lambda h|_c = 0.61$. For the equations as presented, the reason is certainly not obvious. But if the eigenvectors of the equations are formed by the relations

$$\left.\begin{aligned} w_1 &= 5u_1 - 3u_2 \\ w_2 &= 10u_1 + 4u_2 \end{aligned}\right\} \tag{76}$$

one can show equations (75) are equivalent to

$$\left.\begin{aligned} u_1' &= -3u_1 \\ u_2' &= -0.3u_2 \end{aligned}\right\} \tag{77a}$$

with the initial conditions

$$\left.\begin{aligned} u_1(0) &= 0.0001 \\ u_2(0) &= 1.0000 \end{aligned}\right\} \tag{77b}$$

The analytic solutions of these are, of course,

$$u_1 = 0.0001e^{-3x}$$

$$u_2 = 1.0000e^{-0.3x}$$

Now suppose we are not interested in values of u_1 and u_2 when they are ≤ 0.0001. As we have seen, the stability of predictor-corrector methods, when applied to simultaneous equations, depends upon the "worst" eigenvalue of the differential system, in this case -3. Since $-(-3)(0.2) = 0.6 < 0.61$, stability is assured. This corresponds to the second condition in (74b). As far as the required accuracy is concerned, stability is all that is necessary for u_1, since it is smaller than the allowable error to begin with and, being stable, cannot grow. The accuracy of u_2 will be acceptable for any method that makes $|er_\lambda| < 0.0001$ for $|\lambda h| = 0.06$. This corresponds to the first condition in (74b).

Of course, if the initial conditions for equations (75a) are changed to

$$w_1(0) = 2.0000$$

$$w_2(0) = 6.0000$$

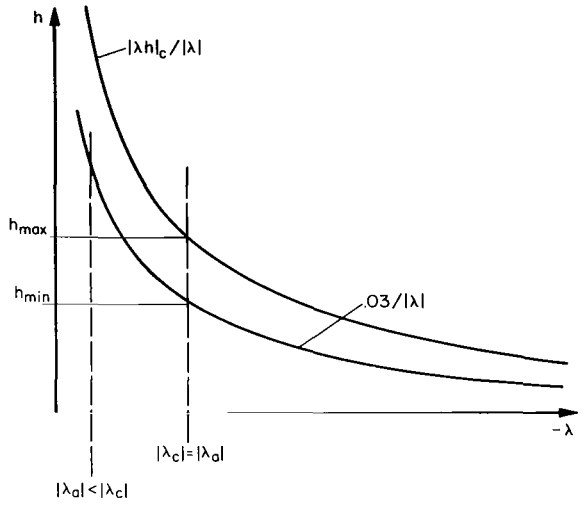
the situation is quite different. Now the analytic solutions are

$$u_1 = 1.0000e^{-3x}$$

$$u_2 = 1.0000e^{-0.3x}$$

and in the interval $0 \leq x \leq 3.1$, $u_1 > 0.0001$. To get the same resolution as before, the step size would have to start at 0.02. By the time $x \approx 3.1$, however, a step size of 0.2 would again be satisfactory.

Clearly, the importance of these considerations depends upon the size and complexity of the set of differential equations being studied. We can calculate, however, the most one can gain by using the necessary conditions (74b) rather than the sufficient conditions (74a). Let $|\lambda_a|$ represent the largest eigenvalue or frequency (i.e., $|\lambda_a| = (|\lambda|, |\mu|)_{max}$) we wish to resolve in a given problem. Let the maximum[11] step size be given by $|h\lambda_a| = 0.03$. Consider the two curves in sketch (c). They represent the maximum step size for a given λ determined on the basis of accuracy $(0.03/|\lambda|)$ and stability $(|\lambda h|_c/|\lambda|)$, where $|\lambda h|_c$ is the stability boundary of the numerical method. If λ_c is the negative eigenvalue largest in magnitude and we wish to resolve it, then $|\lambda_a| \geq |\lambda_c|$ and the accuracy curve governs the step size, giving h_{min}. However, if $|\lambda_a| < |\lambda_c|$ it may be possible to increase the step size until it is governed by the stability curve, giving h_{max}. It is impossible to increase h further by the methods discussed in this report. Under these conditions the ratio of the maximum to the minimum step size is



Sketch (c)

$$\frac{h_{max}}{h_{min}} = \frac{|\lambda h|_c}{0.03} \tag{78}$$

which has a significant effect in the solution of lengthy problems.

Some examples.- Next, we construct the unit circle in the complex plane and plot in the same plane the roots to the characteristic equations for a variety of predictor-corrector methods. In each example the roots are calculated for a step size equal to zero and the locus of these points is indicated by flagged symbols. The step size is then incremented by 0.1 and each root is recorded accordingly. The number of points plotted varies because of scale limitations, but in no case does $|\lambda h|$ exceed one. The value of ω in the representative equation $u' = e^{i\omega}u$ is varied through the range $\pi \geq \omega \geq \pi/2$ to study induced instability, and through $0 \geq \omega \geq -\pi/2$ to study the region of inherent instability.

The location of the principal root under these conditions is shown in figure 1. This root must, of course, be common to all methods, and a measure of the accuracy of a method is displayed by how closely one of its set of points falls on those in figure 1.

[11]Simple two-step methods with $er_\mu \approx er_\lambda \approx 0.01(h\lambda_a)^4$ are available. If $|h\lambda_a| = 0.03$ such methods have an error $\approx 0.81(10)^{-8}$. Smaller step sizes in these (or more accurate) methods in a machine using eight-place floating arithmetic would be useless.

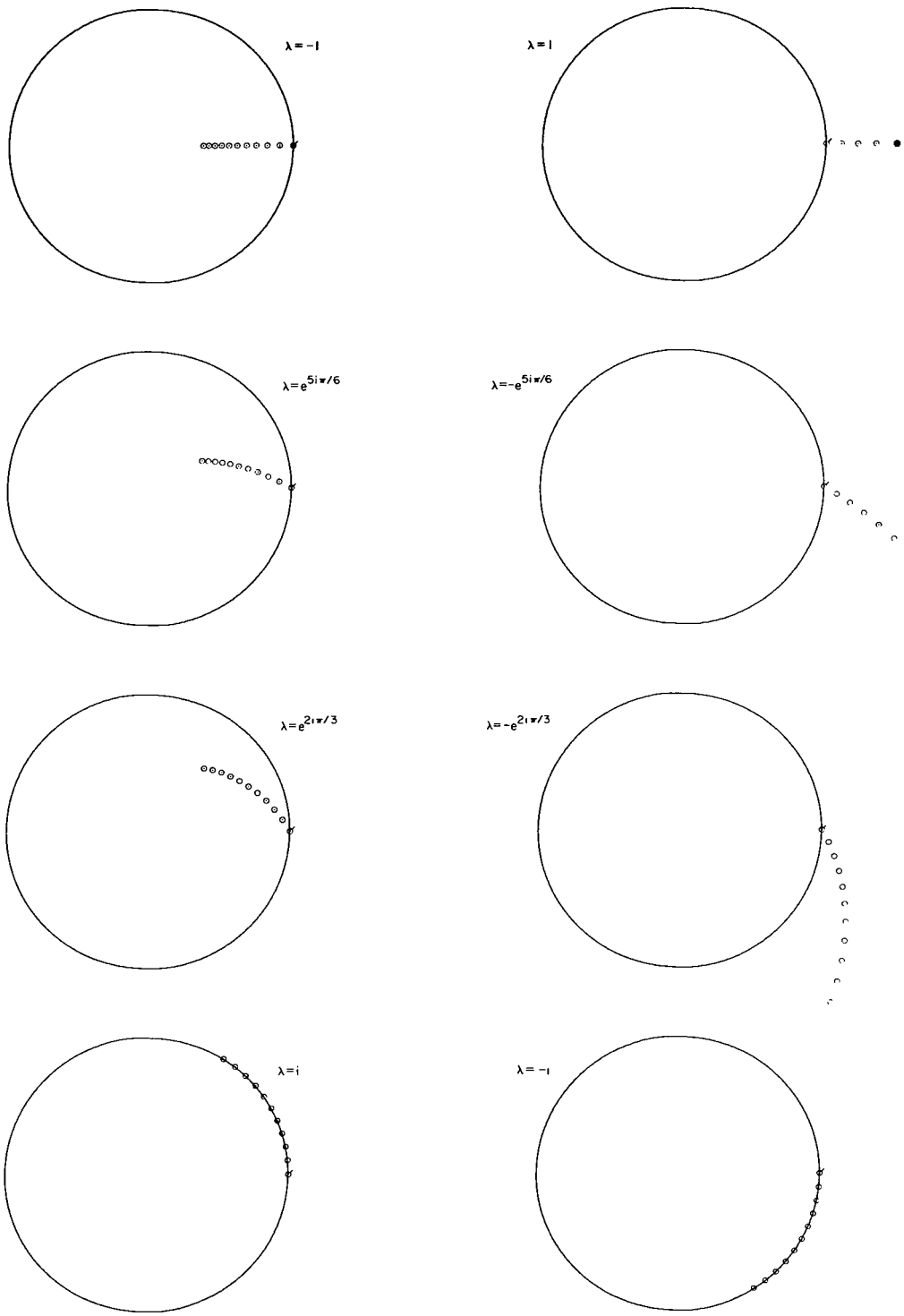


Figure 1.- Location of principal root.

Figure 2 represents the method resulting from the combination of a four-step, Adams-Bashforth predictor (row 5, table I(a)) followed by a four-step, Adams-Moulton corrector (row 4, table I(b)). All of the spurious roots lie on the origin at $h = 0$, forming a triple root there and causing the variation of the roots with h to be quite different for small h ($\partial\lambda_i/\partial h \to ch^{-2/3}$ as $h \to 0$, $i \neq 1$) from what it is for $h \approx 0.1$ and higher. The method presents no stability problem until $|\lambda h| \gtrsim 0.7$. Beyond this value a spurious root exceeds unity for $\lambda = \pm i$. Notice that the upper left-hand circle ($\lambda = -1$) shows the method is stable for λh even less than -1.0 for <u>real</u> eigenvalues. This is a case for which a stability analysis that makes use of only real λ would give quite erroneous results regarding the value of a method for general simultaneous equations.

Figure 3 displays the root structure of Hamming's method without modifiers (row 6, table I(a), and row 6, table 1(b)). Four roots are involved and two of the three spurious roots do not fall on the origin at $h = 0$. When $\lambda = -1$, one of the spurious roots starts at 0.422 and proceeds in a positive direction along the real axis for increasing h. It crosses the principal root at $h \approx 0.265$ causing a degenerate instability there. For $\lambda h \lesssim -0.5$ it falls outside the unit circle and the method becomes definitely unstable. This is a case for which the general stability boundary is determined by considering only real eigenvalues.

The method Hamming finally proposed, and the one usually programmed and referred to by his name, uses two "modifiers." An analysis and discussion of modifiers is given in the next section, where it is shown that Hamming's method with modifiers is equivalent to predicting with row 7, table I(a) and correcting with row 7, table I(b). The roots to the characteristic equation for this method are shown in figure 4. The root structure is completely different from that shown for the unmodified method. There are now five roots. One of the four spurious roots lies on the origin when $h = 0$. The eigenvalue limiting the stability is now complex, occurring when $\lambda = e^{2i\pi/3}$ and the stability boundary is $|\lambda h| \lesssim 0.6$, slightly greater than the unmodified one.

Typical of what can be done to increase the stability of these four- and five-root methods is shown in figure 5. This figure illustrates the roots to the characteristic equation for a method proposed by Crane and Klopfenstein (ref. 11) which amounts to using the four-step predictor in row 8, table I(a), and the four-step Adams-Moulton corrector (row 4, table I(b)). For this method, all of the spurious roots fall in the unit circle for $|\lambda h| < 0.9$.

The classical Milne method (row 6, table I(a), and row 5, table I(b)) has the characteristic roots shown in figure 6. Four roots are generated; two of the spurious roots lie on the origin and one at -1 when $h = 0$. (So, in our terminology, it is a Milne method, but not a total Milne method.) It is well known that the corrector alone is unstable for all negative λh. Chase (ref. 3) showed that for <u>real</u> λh the combined, predictor-corrector process is stable for $-0.3 > \lambda h > -0.8$, a conclusion which is represented here in the upper circle on the left. Clearly, however, a glance at the entire left column in the figure shows that for almost all complex eigenvalues with

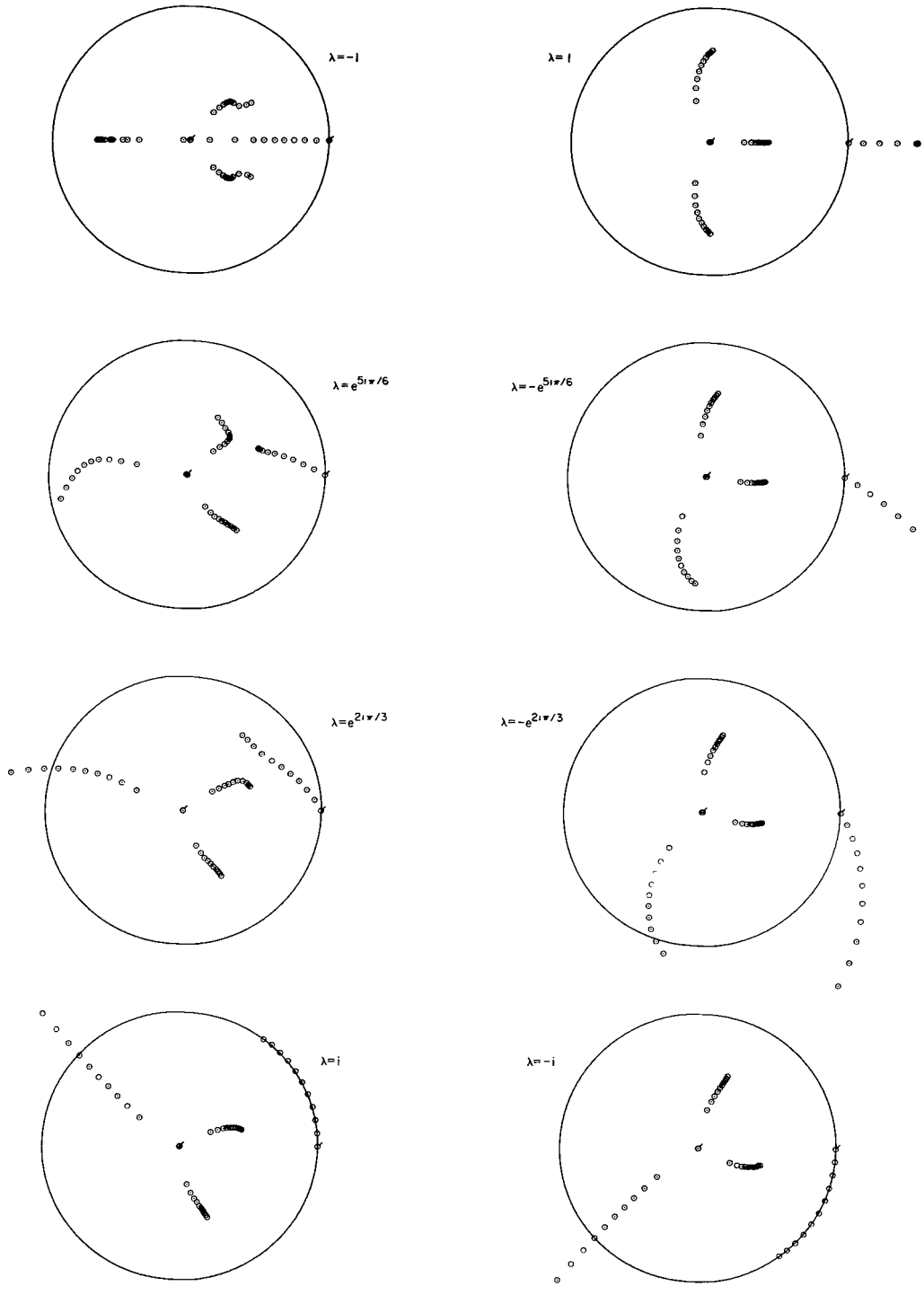


Figure 2.- Adams-Bashforth four-step predictor combined with Adams-Moulton four-step corrector.

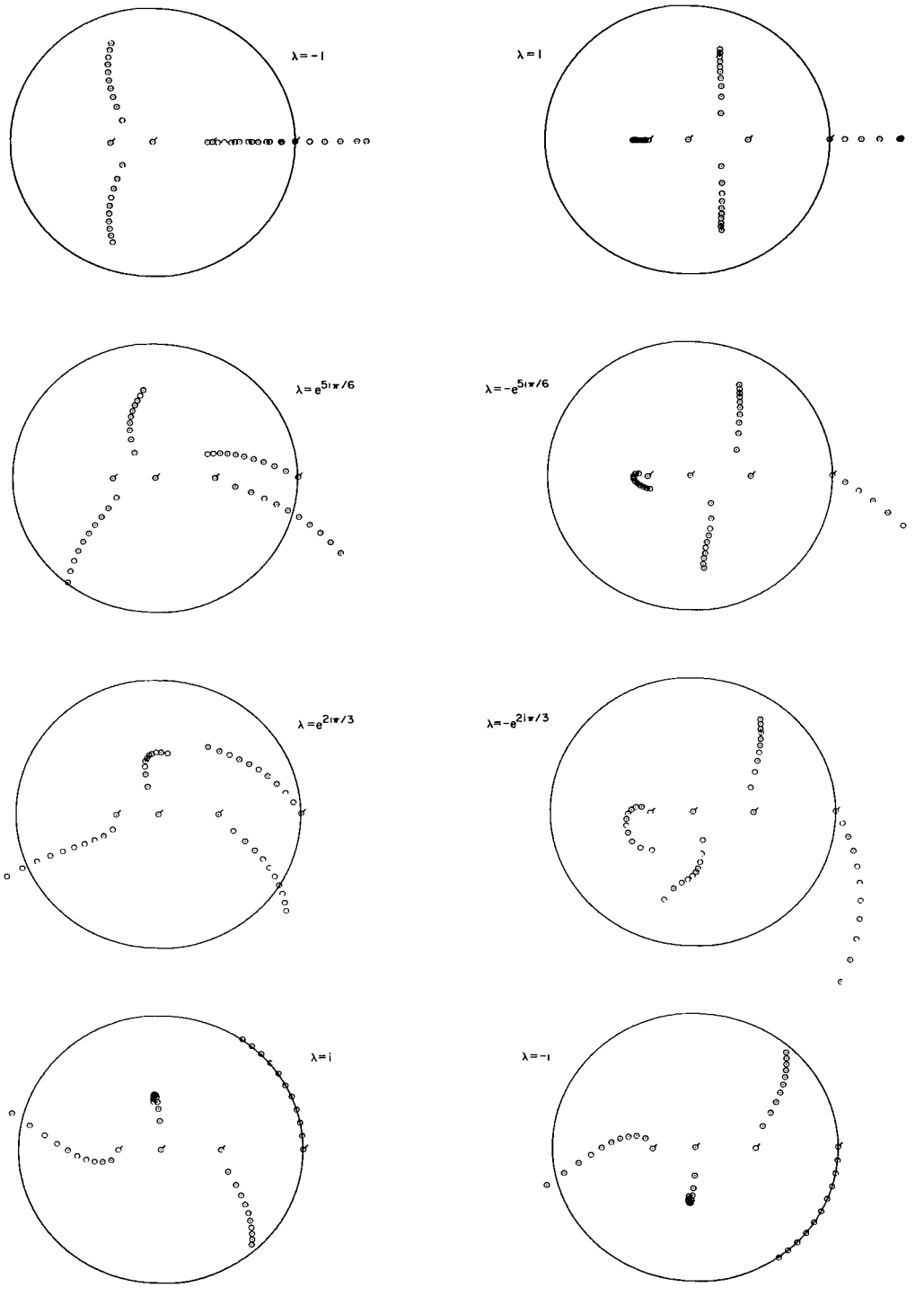


Figure 3.- Hamming's method without modifiers.

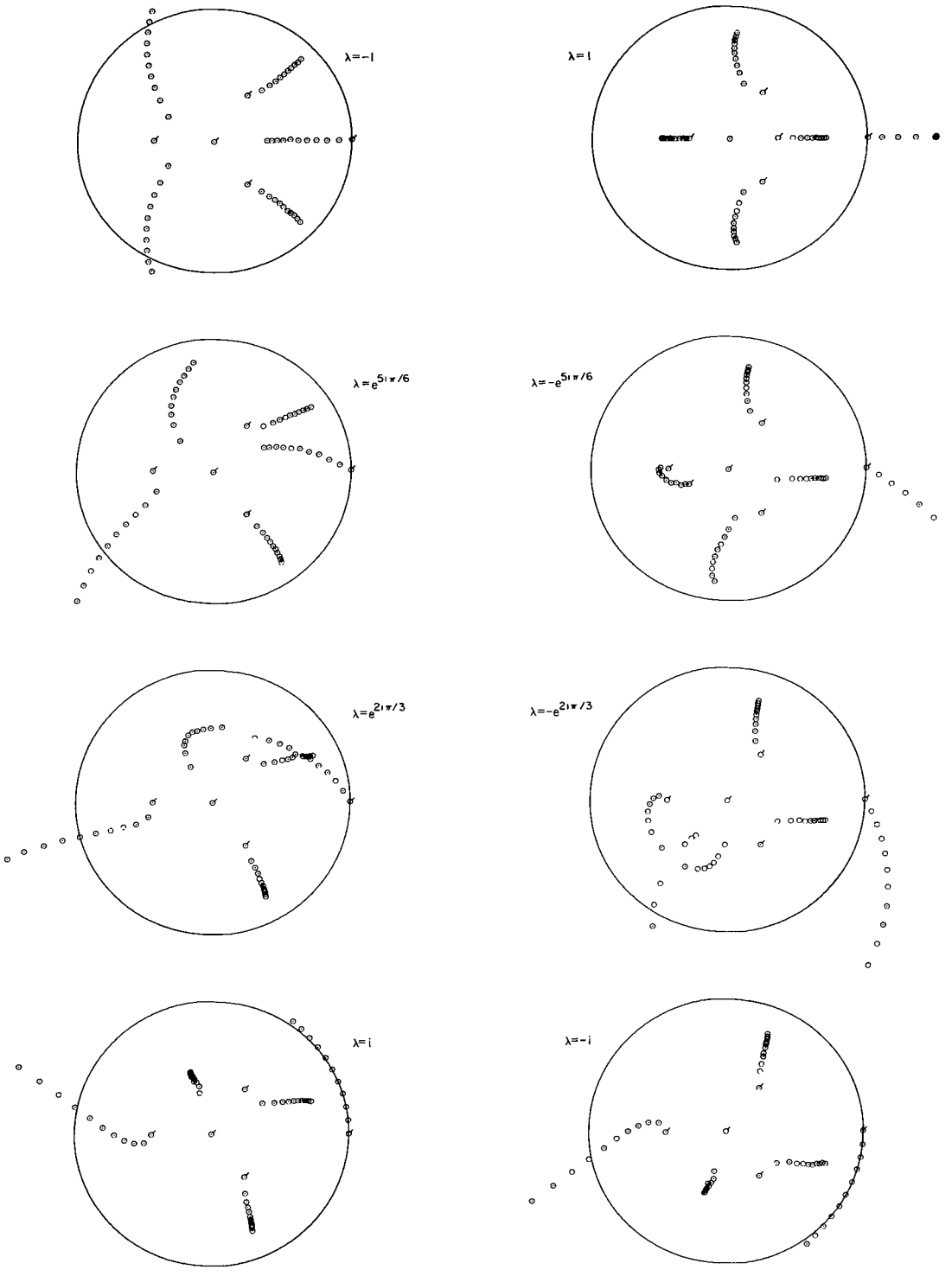


Figure 4.- Hamming's method with modifiers.

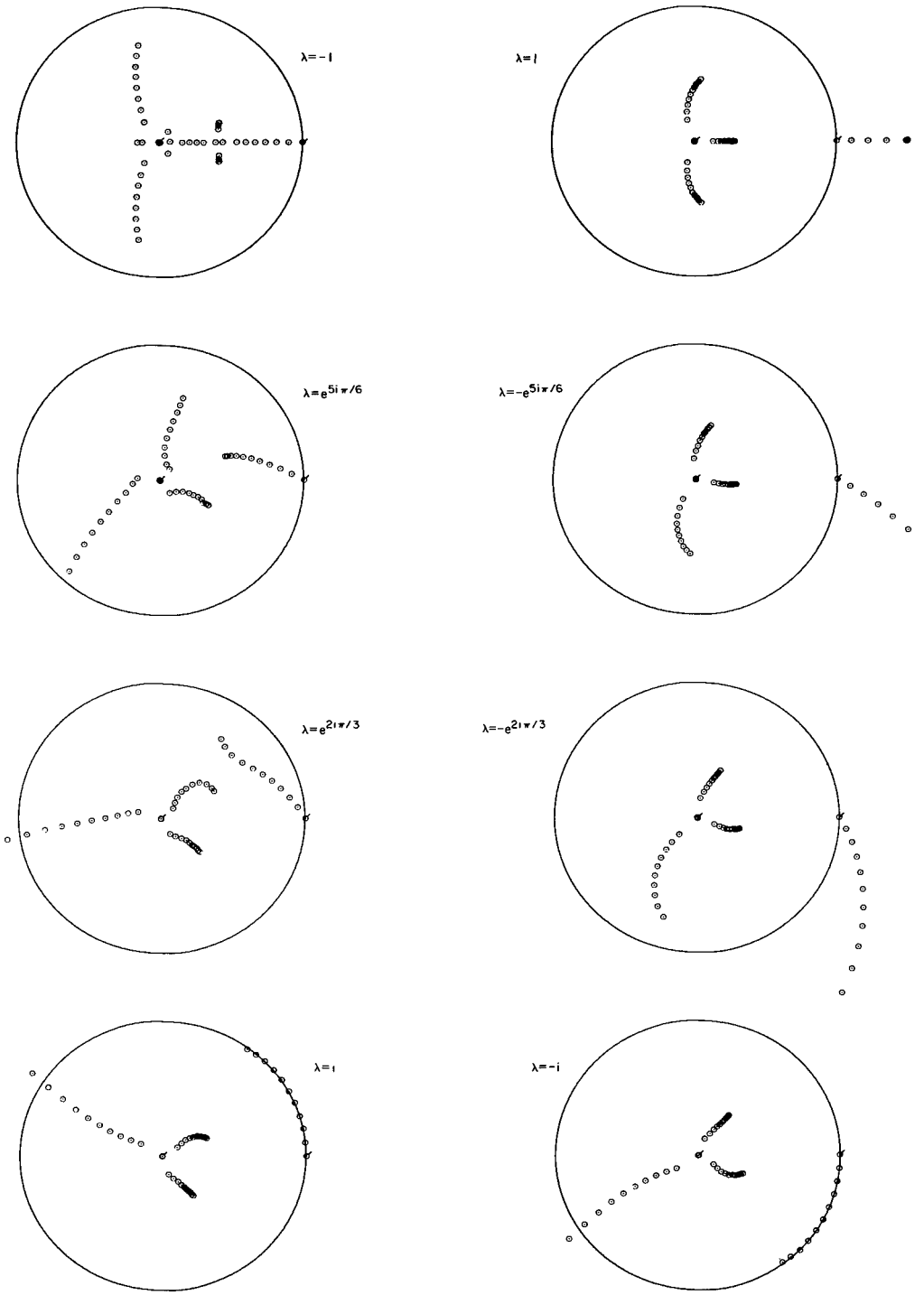


Figure 5.- Method of Crane and Klopfenstein.

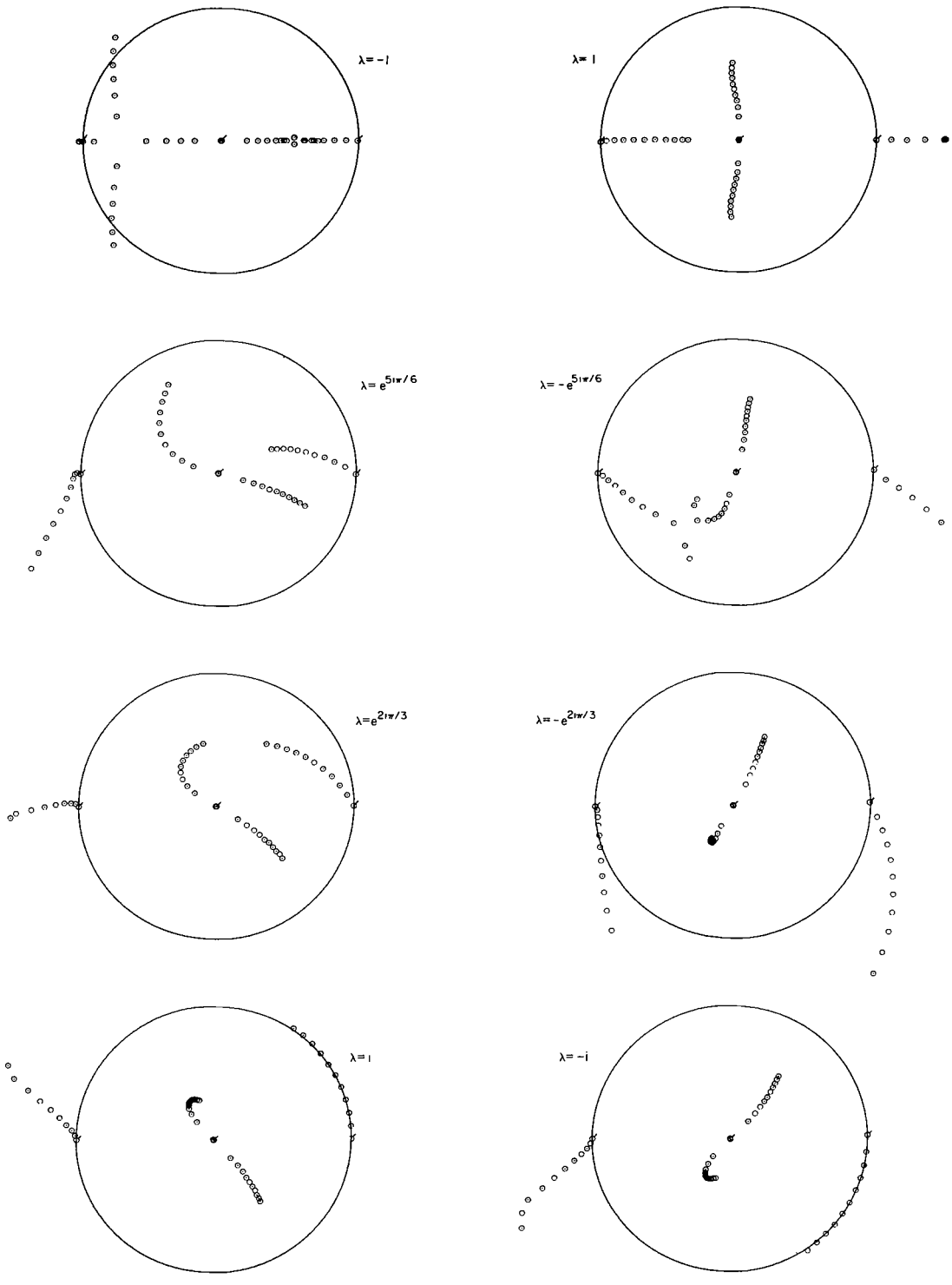


Figure 6.- Milne's method without modifiers.

negative real parts, the Milne predictor-corrector method is unstable for all $h > 0$. This result is also presented in reference 11.

The necessity for considering the stability problem in the entire complex plane has certainly been demonstrated. We will present one more example, however, since it provides a background for some of our subsequent discussion. Consider the total Milne method composed of the two-step predictor in row 10, table I(a), followed by the Milne, two-step corrector (row 5, table I(b)). This method has been proposed by Stetter (ref. 9). Since now both the predictor and the corrector have only two steps, only two roots appear in the characteristic equation. They are shown in figure 7, the left and right columns of the previous figures being collapsed into two circles for convenience. For all real λh the method is stable for $0 \geq \lambda h \geq -1$. It is the most accurate conventional (i.e., incomplete) predictor, one-corrector method that can be devised for arbitrarily small h (er_μ and er_λ are given in table II(k)), and may be used if one is sure that all the eigenvalues of the differential equation are real. As the left circle in figure 7 shows, however, like the classical Milne method, it is unstable for all imaginary eigenvalues.

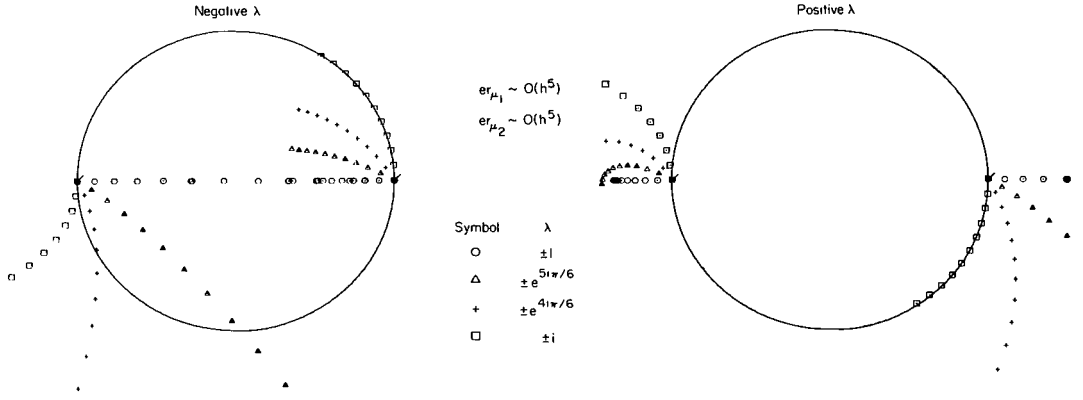


Figure 7.- Stetter's two-step method.

METHODS WITH MODIFIERS OR NONFUNDAMENTAL FAMILIES

Incomplete, Predictor, One-Corrector Methods With Modifiers

Only fundamental families were used to construct the methods studied in the preceding section. It is quite possible, of course, to hold in memory, and subsequently use, combinations of u and u' which were calculated in a previous cycle of computation but are not members of the final families. The equations relating these combinations are often referred to as modifiers. In the terminology of this report, using a modifier corresponds to constructing a family that is not fundamental. The principal purpose of this section is to show (by operational methods) that modifiers are artificial in the following sense:

Any method[12] with modifiers or nonfundamental families can be identified with a method without modifiers composed only of fundamental families.

One of the simplest types of modifiers is that which weighs only past families of the dependent variable and not its derivative. It requires no further iterations, but it does, of course, require more memory. The set of difference-differential equations with modifiers that we will now analyze can be written

$$u^{(1)}_{n+k} = \sum_{j=2}^{k+1} (\alpha_j u_{n+J} + h\alpha'_j u'_{n+J}) \tag{79a}$$

$$u^{(2)}_{n+k} = \tau_2 u^{(1)}_{n+k} + \tau_3 u^{(3)}_{n+k-1} + \tau_4 u^{(1)}_{n+k-1} \tag{79b}$$

$$u^{(3)}_{n+k} = h\beta'_1 u^{(2)'}_{n+k} + \sum_{j=2}^{k+1} (\beta_j u_{n+J} + h\beta'_j u'_{n+J}) \tag{79c}$$

$$u_{n+k} = \sigma_1 u^{(3)}_{n+k} + \sigma_2 u^{(1)}_{n+k} + \sigma_3 u^{(3)}_{n+k-1} + \sigma_4 u^{(1)}_{n+k-1} \tag{79d}$$

Equations (79a) and (79c) are the conventional predictor-corrector equations studied in the previous section. Equations (79b) and (79d) are the modifiers, weighing previously calculated, nonfundamental families by the constants τ_j and σ_j. Applying these equations to the representative equation and introducing the operator E, one derives the matrix equality

$$\begin{bmatrix} E^k & 0 & 0 & -C_\alpha \\ -\tau_2 E^k - \tau_4 E^{k-1} & E^k & -\tau_3 E^{k-1} & 0 \\ 0 & -\lambda h\beta'_1 E^k & E^k & -C_\beta \\ -\sigma_2 E^k - \sigma_4 E^{k-1} & 0 & -\sigma_1 E^k - \sigma_2 E^{k-1} & E^k \end{bmatrix} \begin{bmatrix} u_n^{(1)} \\ u_n^{(2)} \\ u_n^{(3)} \\ u_n \end{bmatrix} = Ahe^{\mu hn} \begin{bmatrix} F_\alpha \\ 0 \\ F_\beta \\ 0 \end{bmatrix} \tag{80}$$

where

$$\left.\begin{aligned} C_\alpha &= \sum_{j=2}^{k+1} (\alpha_j + \lambda h\alpha'_j)E^J & F_\alpha &= \sum_{j=2}^{k+1} \alpha'_j e^{J\mu h} \\ C_\beta &= \sum_{j=2}^{k+1} (\beta_j + \lambda h\beta'_j)E^J & F_\beta &= \sum_{j=1}^{k+1} \beta'_j e^{J\mu h} \end{aligned}\right\} \tag{81}$$

[12] We only consider incomplete methods but the same conclusions apply to complete ones.

If we solve equations (80) for u_n, we can divide out E^{k-1} from the first and third columns, and E^k from the second column. There results

$$\left\{E^{k+2} - \sum_{j=0}^{k+1} E^j \sum_{m=1}^{3} L^*_{mj}(\lambda h)^{m-1}\right\} u_n = hAe^{\mu hn} \sum_{j=-1}^{k+1} e^{J\mu h} \sum_{m=1}^{2} R^*_{mj}(\lambda h)^{m-1} \tag{82}$$

where L^* and R^* are fairly simple combinations of the constants α, τ, β, and σ in equations (79). It is clear that, if we set

$$\left.\begin{aligned} k_e &= k + 2 \\ L_{m,j+2} &= L^*_{mj} \\ R_{m,j+2} &= R^*_{mj} \\ J_e &= k_e + 1 - j \end{aligned}\right\} \tag{83}$$

equation (82) can be written

$$\left\{E^{k_e} - \sum_{j=2}^{k_e+1} E^{J_e} \sum_{m=1}^{3} L_{mj}(\lambda h)^{m-1}\right\} u_n = hAe^{\mu hn} \sum_{j=1}^{k_e+1} e^{J_e\mu h} \sum_{m=1}^{2} R_{mj}(\lambda h)^{m-1} \tag{84}$$

Except for the subscript e, this is the same as equation (50), the operational form of the difference-differential equations (43) and (45), composed only of fundamental families. Thus the k-step method with modifiers given by equations (79) can be identified exactly with a higher step method without modifiers.

More general forms of modifiers, weighing more past families of the function as well as its derivative, can be analyzed. They would simply increase the number of terms with powers of E in the square matrix in equation (80) and lead to characteristic equations in (82) of higher order. By substitutions similar to those in (83), the final operational form can be again identified with equation (50). This correspondence of methods (as they apply to linear differential equations) is always established when the operational forms are identical.

Hamming's Method With Modifiers

A good example of equivalent methods, one using two modifiers, and the other using no modifiers but having one more step is given by analyzing Hamming's method as it is usually programmed (see ref. 2). Hamming's modified method can be written

$$
\left.
\begin{aligned}
u^{(1)}_{n+4} &= u_n + \frac{4}{3}\,h\left(2u'_{n+3} - u'_{n+2} + 2u'_{n+1}\right) \\
u^{(2)}_{n+4} &= u^{(1)}_{n+4} - \frac{112}{121}\left(u^{(1)}_{n+3} - u^{(3)}_{n+3}\right) \\
u^{(3)}_{n+4} &= \frac{1}{8}\left[9u_{n+3} - u_{n+1} + 3h\left(u^{(2)\prime}_{n+4} + 2u'_{n+3} - u'_{n+2}\right)\right] \\
u_{n+4} &= \frac{1}{121}\left[9u^{(1)}_{n+4} + 112u^{(3)}_{n+4}\right]
\end{aligned}
\right\} \qquad (85)
$$

By a straightforward calculation, using the formulae in the previous section, one can show that these have the coefficients in an operational form given by table II(c). Using equations (51b), we immediately find two difference-differential equations that have the same operational form. These are

$$
\left.
\begin{aligned}
u^{(1)}_{n+5} &= u_{n+4} + u_{n+1} - u_n + \frac{4}{3}\,h(2u'_{n+4} - 3u'_{n+3} + 3u'_{n+2} - 2u'_{n+1}) \\
u_{n+5} &= \frac{1}{121}\left[126u_{n+4} - 14u_{n+2} + 9u_{n+1} + h\left(42u^{(1)\prime}_{n+5} + 108u'_{n+4} - 54u'_{n+3} + 24u'_{n+2}\right)\right]
\end{aligned}
\right\} \qquad (86)
$$

Equations (86) represent a conventional, five-step, incomplete, predictor-corrector method composed of two fundamental families which, except for round-off errors, gives results identical to those obtained using Hamming's modified method when applied to equations (11).

Discussion

Consideration of the previous sections raises the question as to the nature of the relationship between families and steps. In what was just presented, a five-step method was duplicated by a four-step method with additional families. How far can this be carried? The answer is that any k-step method can be reduced to a one-step method if the number of families is increased appropriately. (The converse is not true; the minimum number of families is the number of iterations in incomplete methods and one plus the number of iterations in complete ones.) The next question that arises is whether or not this introduction of families serves any really useful purpose. After all, any given method basically evaluates a polynomial of a certain degree using an amount of data stored in memory necessary to attain that degree. From this point of view, there is little difference between a method expressed with modifiers and the same method reduced to fundamental families. Possibly, a few storage locations can be saved and a few arithmetic or logic manipulations eliminated by using one or the other. Most likely, roundoff accumulations will differ, but these are not considered here.

There is quite another point of view, however, from which the family concept can play a valuable role. One can show that, if the proper families are constructed, any polynomial method can be reduced to a several-family, one-step method, the step size of which can be changed at will after each advance. To derive such constructions systematically, however, requires a theory that falls outside this report.

COMPLETE MULTISTEP PREDICTOR-CORRECTOR METHODS

Introduction

We define complete predictor-corrector methods as those in which the final values of the function and its derivative are not members of the same family. A k-step, two-family, complete method is represented by the two equations

$$\left.\begin{aligned} u_{n+k}^{(1)} &= \sum_{j=2}^{k+1} \left(\alpha_j u_{n+J} + \bar{\alpha}_j u_{n+J}^{(1)} + \bar{\alpha}_j' h u_{n+J}^{(1)'} \right) \\ u_{n+k} &= \beta_1' h u_{n+k}^{(1)'} + \sum_{j=2}^{k+1} \left(\beta_j u_{n+J} + \bar{\beta}_j u_{n+J}^{(1)} + \bar{\beta}_j' h u_{n+J}^{(1)'} \right) \end{aligned}\right\} \tag{87}$$

This method requires only one iteration per step. Thus, if the calculation of the derivative dominates the computing time, when compared to incomplete methods, complete ones can

(1) Use twice as many steps for the same amount of computing time, or

(2) Cover the same interval with the same step size in half the computing time.

We seek to find whether or not, on the basis of accuracy and stability, these gains are real or fictitious.

Notice that equations (87) are not composed of (what has been defined in this report to be) fundamental families, since both u_{n+J} and $u_{n+J}^{(1)}$ are retained in memory. Complete methods that use only fundamental families would be

$$\left.\begin{aligned} u_{n+k}^{(1)} &= \sum_{j=2}^{k+1} \left(\alpha_j u_{n+J} + \alpha_j' h u_{n+J}^{(1)'} \right) \\ u_{n+k} &= \beta_1' h u_{n+k}^{(1)'} + \sum_{j=2}^{k+1} \left(\beta_j u_{n+J} + \beta_j' h u_{n+J}^{(1)'} \right) \end{aligned}\right\} \tag{88}$$

for one iteration, and

$$\left.\begin{aligned} u_{n+k}^{(1)} &= \sum_{j=2}^{k+1} \left(\alpha_j u_{n+J} + \alpha'_j h u_{n+J}^{(2)'}\right) \\ u_{n+k}^{(2)} &= \beta'_1 h u_{n+k}^{(1)'} + \sum_{j=2}^{k+1} \left(\beta_j u_{n+J} + \beta'_j h u_{n+J}^{(2)'}\right) \\ u_{n+k} &= \gamma_1 u_{n+k}^{(2)} + \gamma'_1 h u_{n+k}^{(2)'} + \sum_{j=2}^{k+1} \left(\gamma_j u_{n+J} + \gamma'_j h u_{n+J}^{(2)'}\right) \end{aligned}\right\} \tag{89}$$

for two iterations. We show in the next section that almost all operational forms given by equations (87) can be constructed from some form of equations (88).

Analysis and Discussion

Introduce the representative equation (37) into equations (87) and there follows the matrix equation

$$\begin{bmatrix} E^k - \sum_{j=2}^{k+1} (\bar{\alpha}_j + \lambda h \bar{\alpha}'_j) E^J & - \sum_{j=2}^{k+1} \alpha_j E^J \\ -\lambda h \beta'_1 E^k - \sum_{j=2}^{k+1} (\bar{\beta}_j + \lambda h \bar{\beta}'_j) E^J & E^k - \sum_{j=2}^{k+1} \beta_j E^J \end{bmatrix} \begin{bmatrix} u_n^{(1)} \\ u_n \end{bmatrix} = A h e^{\mu h n} \begin{bmatrix} \sum_{j=2}^{k+1} \bar{\alpha}'_j e^{J\mu h} \\ \beta'_1 e^{k\mu h} + \sum_{j=2}^{k+1} \bar{\beta}'_j e^{J\mu h} \end{bmatrix} \tag{90}$$

Compare equations (90) and (49) and the difference between the incomplete and the complete methods begins to appear. The left column in (49) contains the term E^k which can be factored out, leaving a characteristic polynomial of order k. In equation (90) no such factoring can be made, the left column is completely (hence the terminology) filled with terms E^k, E^{k-1}, . . ., E^0, and the characteristic polynomial is now of order $2k$.

Solving equation (90) for u_n gives the operational form

$$\left\{E^{2k} - \sum_{j=2}^{2k+1} (L_{1j} + \lambda h L_{2j}) E^{2k+1-j}\right\} u_n = h A e^{\mu h n} \sum_{j=1}^{2k+1} R_{1j} e^{\mu h (2k+1-j)} \tag{91}$$

where L and R are combinations of the α and β. In the simplest case when k = 1 these combinations are

$$\left.\begin{aligned} L_{12} &= \bar{\alpha}_2 + \beta_2 & R_{11} &= \beta_1' \\ L_{13} &= \bar{\alpha}_2' + \alpha_2\beta_1' & R_{12} &= \bar{\beta}_2' - \bar{\alpha}_2\beta_1' \\ & & R_{13} &= \bar{\beta}_2\bar{\alpha}_2' - \bar{\alpha}_2\bar{\beta}_2' \\ L_{22} &= \beta_2\bar{\alpha}_2 - \alpha_2\bar{\beta}_2 & & \\ L_{23} &= \beta_2\bar{\alpha}_2' - \alpha_2\bar{\beta}_2' & & \end{aligned}\right\} \tag{92}$$

Although there are exactly seven terms on both sides of the equations, inverting them, so as to express α,β in terms of L,R, would be difficult because of their nonlinear form.

If we use only fundamental families, equations (87) reduce to equations (88), and the combinations

$$\left.\begin{aligned} L_{1j} &= \beta_j \\ L_{2j} &= \bar{\alpha}_j' + \beta_1'\alpha_j + \sum_{i=2}^{j-1} (\bar{\beta}_i'\alpha_{j+1-i} - \bar{\alpha}_i'\beta_{j+1-i}) \\ R_{11} &= \beta_1' , \qquad R_{1j} = \bar{\beta}_j' , \qquad j = 2, \ldots \end{aligned}\right\} \tag{93}$$

are formed. But these equations can be inverted, in general, since

$$\beta_1' = R_{11} , \qquad \bar{\beta}_j' = R_{1j} , \qquad \beta_j = L_{1j} , \qquad j = 2, \ldots, k+1 \tag{94a}$$

and

$$\bar{\alpha}_j' + R_{11}\alpha_j + \sum_{i=2}^{j-1} (R_{1j}\alpha_{j+1-i} - L_{1j}\bar{\alpha}_i') = L_{2j} , \qquad j = 1, \ldots, 2k+1 \tag{94b}$$

the latter being a set of linear simultaneous equations for α. For example, if k = 2, equations (94) reduce to

$$\left.\begin{aligned} \beta_1' &= R_{11} , & \bar{\beta}_2' &= R_{12} , & \beta_2 &= L_{12} \\ & & \bar{\beta}_3' &= R_{13} , & \beta_3 &= L_{13} \end{aligned}\right\} \tag{95a}$$

and

$$\begin{bmatrix} R_{11} & 0 & 1 & 0 \\ R_{12} & R_{11} & -L_{12} & 1 \\ R_{13} & R_{12} & -L_{13} & -L_{12} \\ 0 & R_{13} & 0 & -L_{13} \end{bmatrix} \begin{bmatrix} \alpha_2 \\ \alpha_3 \\ \bar{\alpha}_2' \\ \bar{\alpha}_3' \end{bmatrix} = \begin{bmatrix} L_{22} \\ L_{23} \\ L_{24} \\ L_{25} \end{bmatrix} \tag{95b}$$

respectively. In this inverted form L and R are arbitrary and, in particular, L_{24} and L_{25} can be equated to zero. This leads to a set of equations from which α and β (coefficients in the difference-differential equations) can at once be derived for any combination of L and R on the left side of equations (93), provided only that the determinant of the matrix in equation (95b) is not zero. In other words, except for the restriction just given, any operational form contained in equations (87) for $k = 1$ can be identified with an operational form from equations (88) for $k = 2$. This situation remains true for higher values of k, so that equations (88) are sufficiently general to represent almost[13] all complete, two-family, predictor-corrector methods. Further, the relations between the coefficients in its operational form and the coefficients in the difference-differential equation can always be inverted by means of equations (94), provided only that the determinant of (94b) is not zero.

The accuracy and stability of the two-family, complete methods can be studied using the same analysis as that presented for equation (50). For the complete case, simply set L_{3j} and R_{2j} equal to zero. It is apparent from equations (50) and (91) that a k-step complete method will have accuracy and stability features associated with a 2k-step, incomplete method. Superficially, this appears to violate the Dahlquist criterion. When the latter is applied to operational forms, however, we see that in reality no such violation occurs. The difference in stability between complete and incomplete forms is discussed later.

Examples

An example of a complete, two-step, predictor-corrector method contained in equations (88) is

$$\left.\begin{aligned} u_{n+2}^{(1)} &= \frac{1}{4}\left[-14u_{n+1} + 18u_n + h\left(15u_{n+1}^{(1)'} + 7u_n^{(1)'}\right)\right] \\ u_{n+2} &= u_{n+1} + \frac{h}{12}\left(5u_{n+2}^{(1)'} + 8u_{n+1}^{(1)'} - u_n^{(1)'}\right) \end{aligned}\right\} \tag{96}$$

[13] Cases in which equation (94b) are overdetermined have not been investigated.

The coefficients in its operational form are

m \ j	L_{mj} 2	3	4	5	R_{mj} 1	2	3	4	5
1	24	0	0	0	10	6	-2	0	0
2	55	-59	37	-9					

Divide by 24

$$er_{\mu} = \frac{1}{192}\mu^3(\mu - \lambda)H^4 , \qquad er_{\lambda} \sim O(H^5)$$

and the error terms, referenced to the computing step h, are

$$er_{\mu 1} = \frac{1}{24}(\mu h)^4 , \qquad er_{\mu 2} = -\frac{1}{24}\mu^3\lambda h^4 , \qquad er_{\lambda} \sim O(h^5) \tag{97}$$

Equations (96) employ a two-step predictor and an Adams-Moulton, two-step corrector; and they have Adams-Moulton type stability. An incomplete, two-step, predictor-corrector method using the Adams-Moulton corrector can be written

$$\left.\begin{aligned} u^{(1)}_{n+2} &= -2.8u_{n+1} + 3.8u_n + h(3.4u'_{n+1} + 1.4u'_n) \\ u_{n+2} &= u_{n+1} + \frac{h}{12}\left(5u^{(1)\prime}_{n+2} + 8u'_{n+1} - u'_n\right) \end{aligned}\right\} \tag{98}$$

and the coefficients in its operational form are

m \ j	L_{mj} 2	3	R_{mj} 1	2	3
1	12	0	5	8	-1
2	-6	18	0	17	7
3	17	7			

Divide by 12

$$er_{\mu} = \frac{1}{24}\mu^3(\mu - \lambda)H^4 , \qquad er_{\lambda} \sim O(H^5)$$

The lowest order error terms for the method given in (98) are identical to those given in equations (97) when referenced to the computing step h. (To order h^4 this is the minimum possible error if an Adams-Moulton corrector is used.) We see that equations (96) have the same (lowest order) accuracy as equations (98), but require one less iteration per step.

For a fair comparison, the error terms in all methods should be referenced to the representative step size H, defined as the distance the solution is advanced by two iterations. For equations (96) we have

$$h = \frac{1}{2} H$$

and for equations (98)

$$h = H$$

A remark is in order here with regard to the adjustment of er_μ and er_λ to conform with the representative step. Since we are now concerned with the efficiency of a method as it applies to the complete calculation of the differential equations, the global, rather than the local, error should be used as a measure. This will account for the fact that if one method uses half the step size of another, it commits its local error twice as many times. Hence, if in general, h = Ha

$$\left.\begin{aligned} er_\mu(h^n) &\to \frac{1}{a}\, er_\mu(a^n H^n) \\ \text{and} \qquad er_\lambda(h^n) &\to \frac{1}{a}\, er_\lambda(a^n H^n) \end{aligned}\right\} \qquad (99)$$

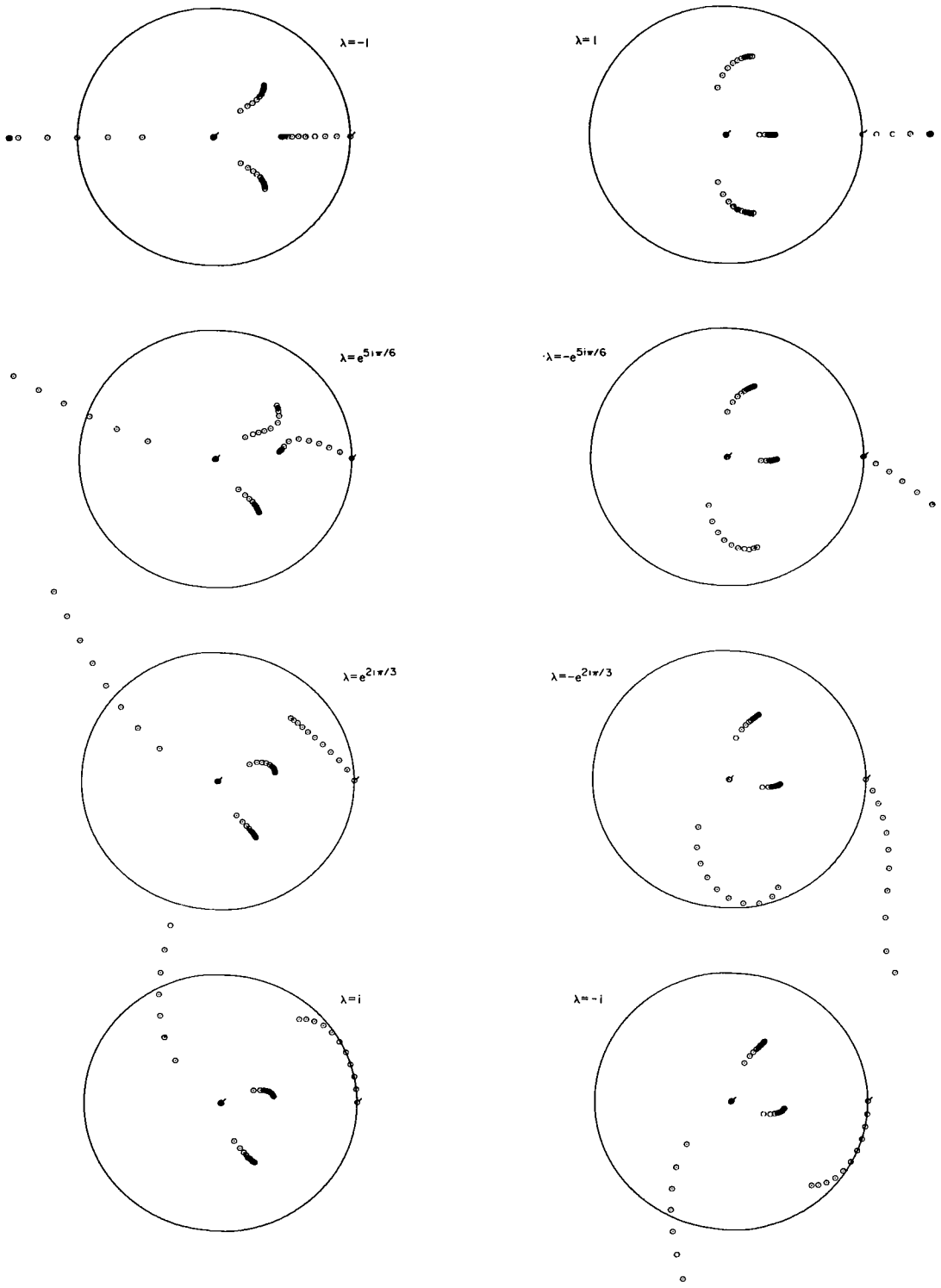
Errors referenced to H are given for the two methods in tables following equations (96) and (98). Clearly, on the basis of accuracy, the complete method is to be preferred, having 1/8 the error of the incomplete method with the same corrector. Since both have Adams-Moulton type stability, all the spurious roots vanish at h = 0. There remains the question, however, regarding the magnitudes of the spurious roots for $h \neq 0$. The characteristic equations DE(E) = 0 for the two methods are

$$24E^4 - (24 + 55\lambda h)E^3 + \lambda h(59E^2 - 37E - 9) = 0 \qquad (100)$$

for the complete method and

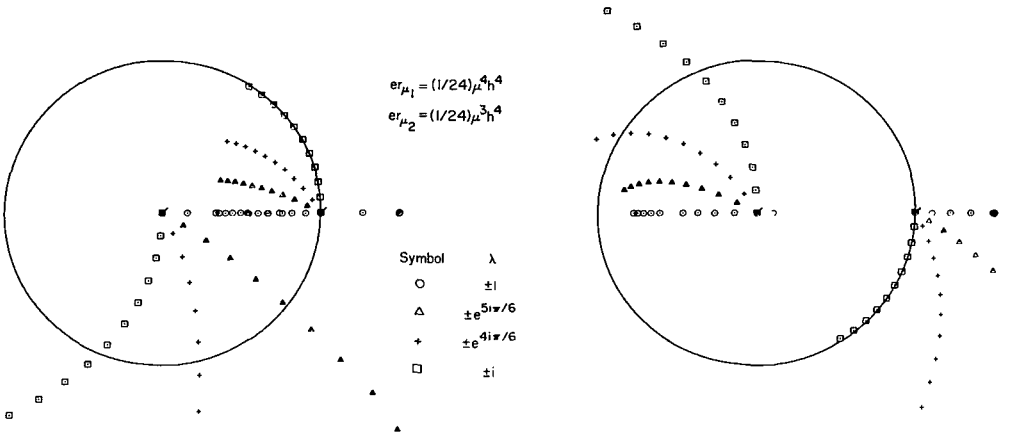
$$12E^2 - (12 - 6\lambda h + 17\lambda^2 h^2)E - \lambda h(18 + 7\lambda h) = 0 \qquad (101)$$

for the incomplete one. We at once see the complete method has two more spurious roots than its counterpart if $h \neq 0$. The magnitudes of the roots for real and complex λh are shown in figures 8(a) and 8(b). In terms of h, the calculation step size, the complete method has induced instability when $|\lambda h| > 0.3$, whereas the incomplete method has no induced instability until $|\lambda h| = 0.5$. However, if we again refer our measurements to the representative step size, the boundaries are $|\lambda H| = 0.6$ and $|\lambda H| = 0.5$, respectively.



(a) Complete, two-step, one-iteration method given by equations (96).

Figure 8.- Stability plots for two different two-step methods.



(b) Incomplete, two-step, two-iteration method given by equations (98).

Figure 8.- Concluded.

In summary, if the evaluation of the derivative dominates the computing time, then, for the same computing time, a complete method with the two-step, Adams-Moulton corrector is more stable than, and has about 1/8 the error of, an incomplete method with the same error. The statement regarding error is based on only the first term in a series expansion. Experience indicates it is not reliable for $|\lambda h| \gtrsim 0.1$.

GENERAL ANALYSIS OF INCOMPLETE MULTISTEP METHODS WITH MULTIPLE CORRECTORS

Derivation of the General Solution for a Fixed Corrector

The process defined by equations (43) and (44) can be generalized such that, during the same cycle of computation, an arbitrary number of correctors -- that is, an arbitrary number of iterations -- is used. Even when each of the correctors is different, the general solution can be derived by the technique outlined below. The solution is quite complicated, however, and does not appear to be of practical interest. The special case, when two correctors, both different, are used, is analyzed later in this section. If all the correctors are the same, the general solution has a rather simple form; and it provides us with the ability to study the effect of the number of iterations on the accuracy and stability of a method.

The equations for the fundamental families of an incomplete method in which $m - 1$ correctors are used, all identical, can be written

$$\left.\begin{aligned}
u_{n+k}^{(1)} &= \sum_{j=2}^{k+1} (\alpha_j u_{n+J} + h\alpha_j' u_{n+J}') \\
u_{n+k}^{(2)} &= h\beta_1' u_{n+k}^{(1)'} + \sum_{j=2}^{k+1} (\beta_j u_{n+J} + h\beta_j' u_{n+J}') \\
&\;\;\vdots \qquad\qquad\qquad\quad \vdots \\
u_{n+k} &= h\beta_1' u_{n+k}^{(m-1)'} + \sum_{j=2}^{k+1} (\beta_j u_{n+J} + h\beta_j' u_{n+J}')
\end{aligned}\right\} \tag{102}$$

Using the notation defined in equations (81) we introduce the representative equation and the operator E and derive the matrix equation

$$\begin{bmatrix}
E^k & 0 & 0 \ldots & 0 & -C_\alpha \\
-b_1E^k & E^k & 0 & 0 & -C_\beta \\
0 & -b_1E^k & E^k & 0 & -C_\beta \\
\vdots & & & & \vdots \\
0 & 0 & 0 \ldots & -b_1E^k & E^k - C_\beta
\end{bmatrix}
\begin{bmatrix} u_n^{(1)} \\ u_n^{(2)} \\ u_n^{(3)} \\ \vdots \\ u_n \end{bmatrix}
= Ahe^{\mu hn}
\begin{bmatrix} F_\alpha \\ F_\beta \\ F_\beta \\ \vdots \\ F_\beta \end{bmatrix} \tag{103}$$

where

$$b_1 = \lambda h\beta_1' \tag{104}$$

Expand the determinant about the right-hand column and the characteristic equation simplifies to

$$\begin{aligned}
DE(E) &= E^k - C_\alpha b_1^{m-1} - C_\beta[b_1^{m-2} + b_1^{m-3} + \ldots + 1] \\
&= E^k - C_\alpha b_1^{m-1} - C_\beta \frac{1 - b_1^{m-1}}{1 - b_1} = 0
\end{aligned} \tag{105}$$

Introduce the notation

$$\left.\begin{aligned}
L_j &= (\alpha_j + \lambda h\alpha_j')b_1^{m-1} + (\beta_j + \lambda h\beta_j') \frac{1 - b_1^{m-1}}{1 - b_1} \\
R_j &= \alpha_j' b_1^{m-1} + \beta_j' \frac{1 - b_1^{m-1}}{1 - b_1}
\end{aligned}\right\} \tag{106}$$

and the operational form turns out to be

$$\left\{ E^k - \sum_{j=2}^{k+1} L_j E^J \right\} u_n = Ahe^{\mu hn} \sum_{j=2}^{k+1} R_j e^{J\mu h} \tag{107}$$

If

$$E^k - \sum_{j=2}^{k+1} L_j E^J = (E - \lambda_1)(E - \lambda_2) \cdots$$

the complete solution can be written

$$u_n = C_1(\lambda_1)^n + C_2(\lambda_2)^n + \cdots + Ahe^{\mu hn} \frac{\sum_{j=2}^{k+1} R_j e^{J\mu h}}{e^{k\mu h} - \sum_{j=2}^{k+1} L_j e^{J\mu h}} \tag{108}$$

The case $m = 1$ represents one iteration per cycle of computation, that is, a predictor without a corrector. This follows from equations (106) since β_j and β'_j disappear from L_j and R_j when $m = 1$. On the other hand, if $m \to \infty$, the equations are independent of α_j and α'_j (provided $|b_1| < 1$, which is a necessary condition for the convergence of the iterations), and the corrector equation in its implicit form emerges.

A Discussion of Some Simple Predictor-Corrector Methods

If we use equation (108) to inspect the simple predictor-corrector scheme (an Euler predictor followed by a modified Euler corrector)

$$u_{n+1}^{(1)} = u_n + hu'_n$$

$$u_{n+1}^{(2)} = u_n + \frac{h}{2}\left(u_{n+1}^{(1)\prime} + u'_n\right)$$

$$u_{n+1}^{(3)} = u_n + \frac{h}{2}\left(u_{n+1}^{(2)\prime} + u'_n\right)$$

$$\vdots \qquad\qquad \vdots$$

we find the complementary solution to be after m iterations

$$u_{n_C} = C_1 \left[\frac{1 + \frac{\lambda h}{2} - 2\left(\frac{\lambda h}{2}\right)^{m+1}}{1 - \frac{\lambda h}{2}} \right]^n \tag{109}$$

This result serves as an excellent example of the danger of analyzing a corrector, ignoring the effect of a predictor. If the modified Euler method is studied alone -- as an implicit identity -- one finds the complementary solution

$$u_{n_C} = C_1\left[\frac{1 + (\lambda h/2)}{1 - (\lambda h/2)}\right]^n \tag{110}$$

instead of equation (109). This shows at once that the modified Euler method is stable for all negative values of λh (see ref. 12). But clearly, as m becomes large, equation (109) reduces to equation (110) only when $|\lambda h| < 2$. Hence, when used as a corrector in a predictor-corrector sequence, the modified Euler method (or trapezoidal rule as it is sometimes referred to) is violently unstable for $\lambda h \ll -2$.

A further study reveals that er_μ and er_λ for the Euler-modified-Euler method behave in the following manner for increasing numbers of iterations in a cycle of computation

m	er_μ	er_λ
1	$-h^2(\mu^2/2)$	$-h^2(\lambda^2/2)$
2	$-h^3(\mu^3 - 3\lambda\mu^2)/12$	$-h^3(\lambda^3/6)$
3	$h^3(\mu^3/12)$	$h^3(\lambda^3/12)$
∞	$h^3(\mu^3/12)$	$h^3(\lambda^3/12)$

This method is often used in programming the "method of characteristics" in the study of hyperbolic partial differential[14] equations. It is of interest to notice:

1. The order of error in the predictor is one less than the order of error in the corrector, but the order of error in the method is the same as that for the corrector after one application of the corrector.

2. The coefficient of the h^3 term is improved by a second application of the corrector.

3. The error, as measured by the lowest order in the truncation, is not further reduced if the iterations are continued.

If the Euler predictor is replaced by the Nystrom predictor (row 2 in table I(a)), the error sequence with iterations is

[14]When more than one independent variable is involved, the reference step H should be redefined.

m	er_μ	er_λ	Spurious root
1	$-h^3(\mu^3/6)$	$-h^3(\lambda^3/6)$	$-1 + \lambda h$
2	$h^3(\mu^3/12)$	$h^3(\lambda^3/12)$	$-\frac{1}{2}\lambda h$
∞	$h^3(\mu^3/12)$	$h^3(\lambda^3/12)$	$-\frac{1}{2}\lambda h$

Here, we see that the predictor by itself has the same order of error as the corrector but is less accurate, and, further, is unstable. One application of the corrector (a total of 2 iterations)

1. Increases the accuracy such that no improvement on the coefficient of h^3 is made by continuing the iterations

2. Produces a stable numerical method.

Basically, this process is the one used in several computer programs to solve for the flow in front of blunt bodies travelling at high speeds.

Incomplete Multistep Predictor Two-Corrector Methods

Next, the incomplete methods previously analyzed are extended by adding one more corrector with arbitrary coefficients. Only a brief sketch of the procedure is given here, mostly for the sake of thoroughness, since the practicality of using two correctors is open to question. However, the added corrector has a decidedly stabilizing effect. In fact, it is shown that, for two-step incomplete cases, a stable, two-corrector method can be constructed having error terms one order higher than is possible for stable, one-corrector methods.

Development.- The three fundamental families are

$$\left.\begin{aligned} u^{(1)}_{n+k} &= \sum_{j=2}^{k+1} (\alpha_j u_{n+J} + h\alpha'_j u'_{n+J}) \\ u^{(2)}_{n+k} &= h\beta'_1 u^{(1)'}_{n+k} + \sum_{j=2}^{k+1} (\beta_j u_{n+J} + h\beta'_j u'_{n+J}) \\ u_{n+k} &= \gamma_1 u^{(2)}_{n+k} + h\gamma'_1 u^{(2)'}_{n+k} + \sum_{j=2}^{k+1} (\gamma_j u_{n+J} + h\gamma'_j u'_{n+J}) \end{aligned}\right\} \qquad (111)$$

where, by the argument used to derive equation (45) β_1 can be made zero without loss of generality; but, by the same argument, the term containing γ_1 cannot. Although a weighted value of the middle equation (111) added to the

final equation can be made to cancel the term involving $u_{n+k}^{(2)}$ (similar to what was done in equation (44)), a term containing $u_{n+k}^{(1)}{}'$ would then appear amounting to a method different from that given by equations (111) with $\gamma_1 = 0$.

Paralleling the one-corrector case studied in a previous section, one can construct a matrix equation and derive the operational form. There results

$$\left\{E^k - \sum_{j=2}^{k+1} E^J \sum_{m=1}^{4} L_{mj}(\lambda h)^{m-1}\right\} u_n = Ahe^{\mu hn} \sum_{j=1}^{k+1} e^{J\mu h} \sum_{m=1}^{3} R_{mj}(\lambda h)^{m-1} \tag{112}$$

from which

$$DE(E) = E^k - \sum_{j=2}^{k+1} E^J \sum_{m=1}^{4} L_{mj}(\lambda h)^{m-1} \tag{113}$$

and

$$NU = h(\mu - \lambda) \sum_{j=1}^{k+1} e^{J\mu h} \sum_{m=1}^{3} R_{mj}(\lambda h)^{m-1} \tag{114}$$

For $j = 2, 3, \ldots$ (since $L_{11}, L_{21}, \ldots = 0$)

$$\left.\begin{aligned} L_{1j} &= \beta_j\gamma_1 + \gamma_j \\ L_{2j} &= \alpha_j\beta_1'\gamma_1 + \beta_j'\gamma_1 + \beta_j\gamma_1' + \gamma_j' \\ L_{3j} &= \alpha_j'\beta_1'\gamma_1 + \alpha_j\beta_1'\gamma_1' + \beta_j'\gamma_1' \\ L_{4j} &= \alpha_j'\beta_1'\gamma_1' \end{aligned}\right\} \tag{115}$$

and for $j = 1, 2, \ldots$

$$\left.\begin{aligned} R_{1j} &= \beta_j'\gamma_1 + \gamma_j' \\ R_{2j} &= \alpha_j'\beta_1'\gamma_1 + \beta_j'\gamma_1' \\ R_{3j} &= L_{4j} \end{aligned}\right\}$$

These equations uniquely determine L and R, the coefficients in the operational form, for given α, β, and γ. Once again, the accuracy of the method is represented by equations (63) and (68), and the stability depends on the magnitude of the roots to $DE(E) = 0$.

Equations (115) can be inverted if γ_1 is set equal to zero. Thus

$$\left.\begin{aligned} \alpha_j &= (L_{3j} - R_{2j})/R_{21} & \alpha'_j &= L_{4j}/R_{21} \\ \beta_j &= (L_{2j} - R_{1j})/R_{11} & \beta'_j &= R_{2j}/R_{11} \\ \gamma_j &= L_{1j} & \gamma'_j &= R_{1j} \end{aligned}\right\} \qquad (116)$$

and difference-differential equations that represent any operational form for which R_{11} and R_{12} are not zero can at once be written. The practical consequence of these limitations is not known.

To find the error in the particular solution, the numerator of (58) is expanded in powers of λh. In this case $R_{3j} = L_{4j}$ and the coefficient of the $\lambda^3 h^3$ term vanishes identically. Choosing the L and R so the coefficients of the $(\lambda h)^0$, $(\lambda h)^1$, and $(\lambda h)^2$ terms are zero gives, respectively,

$$\sum_{j=1}^{k+1} [l(k + 1 - j)^{l-1} R_{1j} + (k + 1 - j)^l L_{1j}] - k^l = 0 ,$$
$$l = 0, 1, 2, \ldots, L \qquad (117a)$$

$$\sum_{j=1}^{k+1} [l(k + 1 - j)^{l-1} R_{2j} + (k + 1 - j)^l (L_{2j} - R_{1j})] = 0 ,$$
$$l = 0, 1, 2, \ldots, L - 1 \qquad (117b)$$

$$\sum_{j=1}^{k+1} [l(k + 1 - j)^{l-1} R_{3j} + (k + 1 - j)^l (L_{3j} - R_{2j})] = 0 ,$$
$$l = 0, 1, 2, \ldots, L - 2 \qquad (117c)$$

which are the accuracy conditions for equations (111). One can show that, as in the single corrector case, both the particular and complementary solutions are fit locally by polynomials of order L if the above conditions are satisfied.

The leading terms in the series expansion for er_μ are determined by evaluating for $l = L + 1$ the remainders in the expressions -- they are no longer equalities -- (117) multiplying these remainders by $(\mu h)^{L+1}/(L + 1)!$, $\lambda h(\mu h)^L/L!$, and $\lambda^2 h^2 (\mu h)^{L-1}/(L - 1)!$, respectively, and dividing each by $\sum_{j=2}^{k+1} (j - 1)L_{1j}$ as in the development of equations (64) and (65).

The derivation of the error in the complementary term is identical to the derivation of equation (71). Thus one can show

$$er_\lambda = \frac{-\frac{(\lambda h)^l}{l!}\left\{\sum_{j=2}^{k+1}\left[J^l L_{1j} + lJ^{l-1}L_{2j} + l(l-1)J^{l-2}L_{3j} + l(l-1)(l-2)J^{l-3}L_{4j}\right] - k^l\right\}}{\sum_{j=2}^{k+1}(j-1)L_{1j}} \tag{118}$$

Finally, to compare fairly with other methods, all error terms should be expressed in terms of the representative step H where

$$h = 3H/2 \tag{119}$$

and the adjustments are made according to equations (99).

Example.- The incomplete two-step, double-corrector method given by

$$\left.\begin{aligned} u_{n+2}^{(1)} &= \frac{1}{89}\left[-248u_{n+1} + 337u_n + h(302u'_{n+1} + 124u'_n)\right] \\ u_{n+2}^{(2)} &= \frac{1}{375}\left[-432u_{n+1} + 807u_n + h\left(89u_{n+2}^{(1)\prime} + 788u'_{n+1} + 305u'_n\right)\right] \\ u_{n+2} &= u_n + \frac{h}{3}\left(u_{n+2}^{(2)\prime} + 4u'_{n+1} + u'_n\right) \end{aligned}\right\} \tag{120}$$

has the operational form represented by

m \ j	L_{jm}		R_{jm}		
	2	3	1	2	3
1	0	1125	375	1500	375
2	1068	1182	89	788	305
3	540	642	0	302	124
4	302	124			

Divide by 1125

$$er_\mu = (0.028\mu^2 - 0.040\mu\lambda - 0.020\lambda^2)\mu^3H^5$$

$$er_\lambda = -0.032\lambda^5H^5$$

The accuracy of the method as measured by the computational step size is given by

$$\left.\begin{aligned} er_\mu &= (0.0057\mu^2 - 0.0080\lambda\mu - 0.0040\lambda^2)\mu^3h^5 \\ er_\lambda &= -0.0063\lambda^5h^5 \end{aligned}\right\} \tag{121}$$

and as measured by the representative step size as shown in the table.

There are two roots to the characteristic equation

$$1125E^2 - E(1068\lambda h + 540\lambda^2h^2 + 302\lambda^3h^3) - 1125 - 1182\lambda h - 642\lambda^2h^2 - 124\lambda^3h^3 = 0 \tag{122}$$

and they are shown in figure 9. The spurious root starts on the unit circle

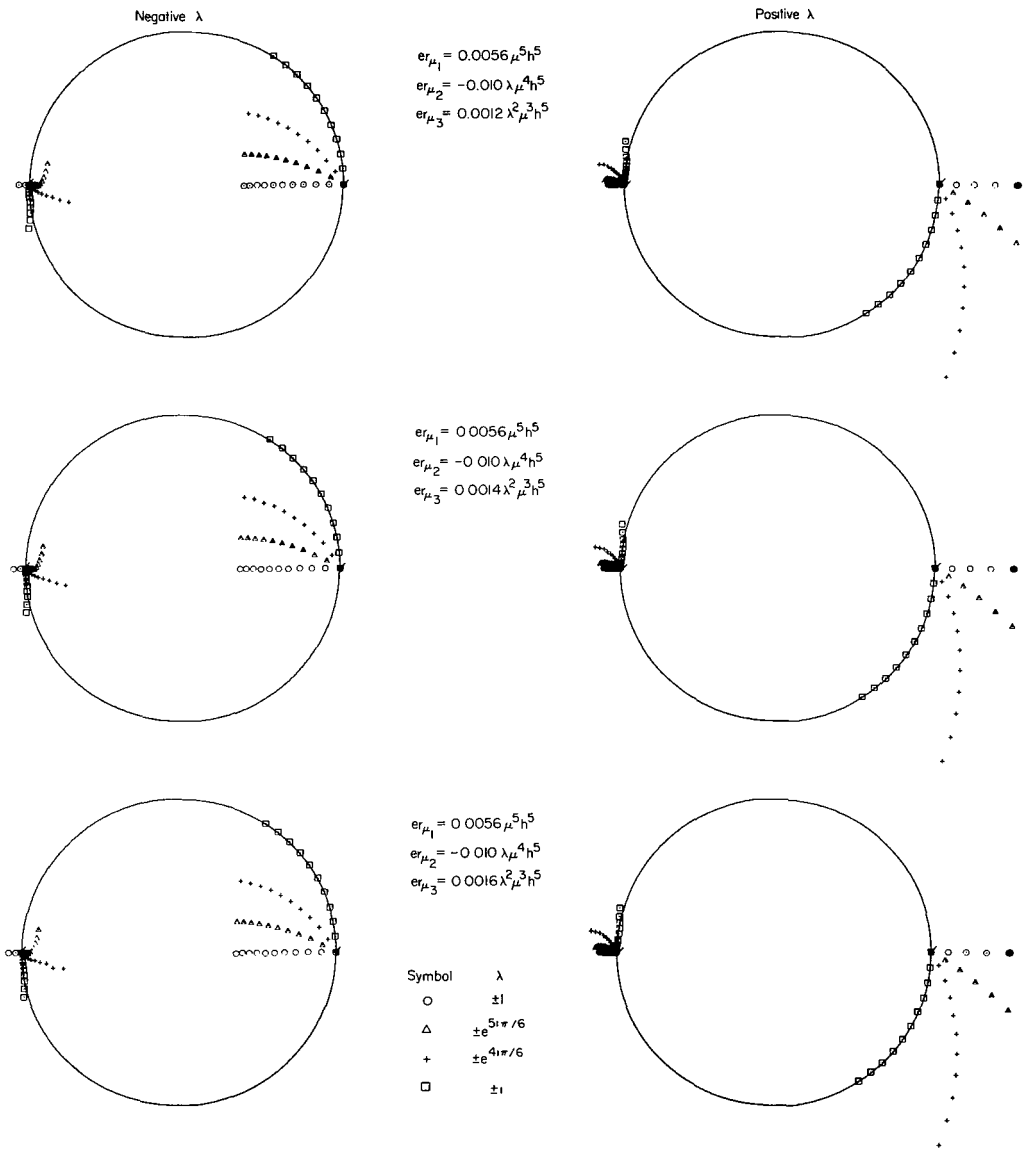


Figure 9.- Stability of predictor, two-corrector method given by equations (120).

at -1. Although it is impossible to tell from the figure, a close examination of the results shows that the spurious root moves into the unit circle

for all complex λ given by $\lambda = e^{i\omega}$ for which $\pi/2 < \omega < 3\pi/2$. Further, the spurious root remains inside the unit circle for $|\lambda h| \lesssim 0.57$, or $|\lambda H| \leq 0.38$ which establishes the stability boundary according to equation (73). This serves as an example of a total Milne method (all the spurious roots are on the unit circle when $h = 0$) that has no induced stabilities for a range of $h > 0$. The stability boundary can be increased to at least $|\lambda H|_c = 0.56$ if the maximum error is allowed to increase by a factor of about 3/2.

COMBINED RUNGE-KUTTA AND PREDICTOR-CORRECTOR METHODS

Introduction

Now let us consider the overall result of increasing the number of iterations in a cycle of computation. In the analysis of the equation $u' = F(x,u)$, it is clear that each successive iteration requires the re-evaluation of the term $F(x,u)$ at a value of u different from any used in previous iterations. As the number of iterations increases, appropriate choices of α, β, γ, etc., permit us to match the final result with a Taylor series expansion of $F(x,u)$ in the u direction through any given order -- regardless of step number. In other words, the accuracy of fit to the complementary solution depends on both the step number and the number of iterations; and its series expansion can be matched with arbitrary accuracy by increasing either the one or the other independently.

Next consider the particular solution to the differential equations. Conventional predictor-corrector formulas are constructed using a fixed and equal spacing of the independent variable, x, a spacing we have designated as h. In such cases the number of samplings of $F(x,u)$ in the x direction is determined entirely by the number of steps used in the method. No new information regarding the variation of $F(x,u)$ with x is supplied by increasing the number of correctors in a cycle of computation. Since, in general, the error associated with a method must depend on its worst fit to either the particular or the complementary solution, equispaced, predictor-corrector methods are limited in accuracy by their step number, regardless of the number of iterations.

Clearly, if both u and x are varied in the successive correctors, the complete series expansion of $F(x,u)$ in both x and u can be matched to an indefinite order, and arbitrary accuracies to both the complementary and particular solutions obtained. The development of this concept leads, at once, to equations that merge the predictor-corrector and the Runge-Kutta methods.

On the General Form of the Equations

One can write a set of difference-differential equations that represent a complete, combined method with k steps and m iterations. The result would be a set of formulas in which all families of all the values of the function and its derivative calculated in a cycle of computation and held in

memory are weighted with appropriate values of α, β, γ, etc. (see, e.g., ref. 13). By considering only the two-step case, we will see the complexity involved in such a completely general expression. For the practical purpose of actually constructing methods with fixed accuracy and stability, much simpler subsets of these general expressions appear to be satisfactory. The real crux of the problem of constructing optimum methods depends upon whether or not a set of equations can be found that explicitly relates the coefficients in the difference-differential equations to the coefficients in the operational form (e.g., eqs. (51b), (94), and (116)). This question, in turn, depends upon how the operator E is brought into the matrix equation from which the operational form is constructed.

The discussion of combined methods is more readily presented if we consider simple sketches in which the following symbols are used:

□ predicted value of the function, $u^{(1)}$

△ derivative of the predicted value, $du^{(1)}/dx$

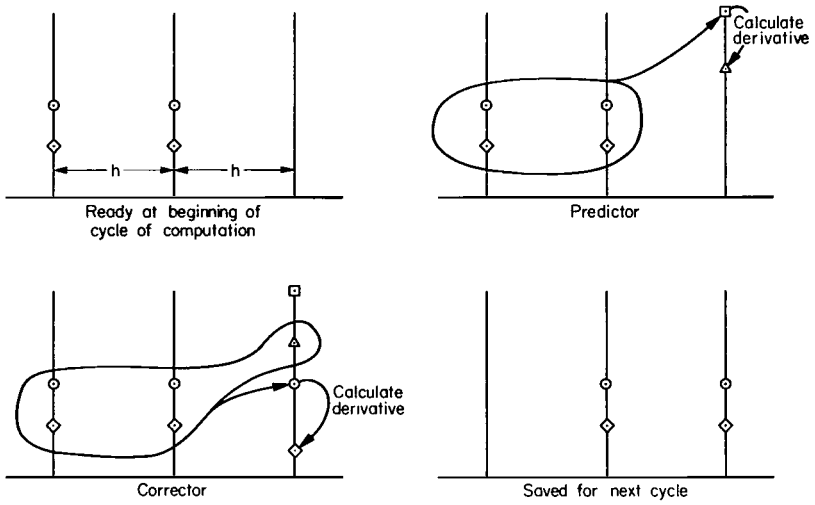
○ corrected value of the function at a point previously predicted, u

◇ derivative of the corrected value, du/dx

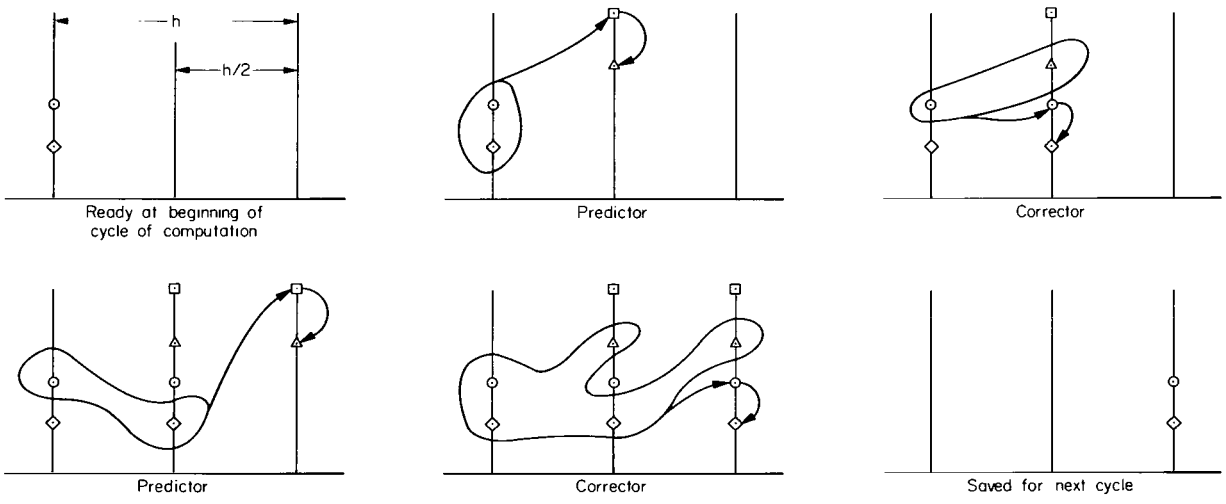
On the definition of step size in combined methods.- Using the symbols defined above, let us construct a cycle of computation for a two-step, predictor-corrector method (e.g., eq. (98)), and the standard, one-step, fourth-order, Runge-Kutta method (eqs. (136)).

When presented graphically, as in the sketches, it may at first appear that singling out a length h and calling it step size is a rather arbitrary procedure. In fact, an ambiguity in the use of the words "step size" and "step number" can easily arise when predictor-corrector methods are combined with Runge-Kutta methods; although, as we shall presently see, the term can be given a unique and quite natural definition that applies to the two different approaches, either individually or in combination.

In conventional predictor-corrector schemes, such as that shown in sketch (d), the function and its derivative are calculated at equispaced points only, and a value of the function is (or can be without loss of accuracy) "output" at each point. The spacing is quite naturally referred to as the step size and the resulting step number can be and is used as the fundamental parameter in describing both the accuracy and stability of the method. See, for example, the Dahlquist stability theorem. On the other hand, in Runge-Kutta methods the function and its derivative are calculated at points other than those at which it is most accurately represented or intended to be output. In sketch (e) it is shown at the midpoint, but this is totally unnecessary. Nevertheless, regardless of the number of intermediate points used, or their spacing, the Runge-Kutta methods are always referred to as one-step methods.



Sketch (d)



Sketch (e)

The two sketches lead us at once to a definition of step size that is common to predictor-corrector methods, Runge-Kutta methods, or any combination thereof. Thus

$$h = \text{step size} \equiv \text{the distance the integration is advanced by one cycle of computation} \tag{123}$$

This definition is in keeping with the usage throughout this report and in all references about numerical methods of which the author is aware, including the recent ones on combined methods (refs. 13, 16).

With such a definition, step size is still a useful parameter for describing the accuracy of a method, but it no longer has any basic meaning with regard to stability. In short, stable combined methods exist which embed polynomials of arbitrarily high order regardless of step number. This is further discussed in the last section of the report entitled Stability.

When comparing the accuracy and stability of combined methods, then, either with themselves or other methods, one must do so on the basis of controlled values of such things as:

1. The amount of memory required

2. The amount of arithmetic required

Examples of a two-step method.- In the following, we present a variety of two-step methods subjected to the following restraints:

1. A memory of -- at most -- four values of the function and/or its derivative.

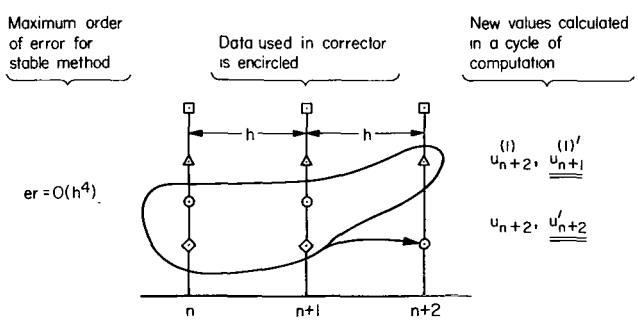
2. The calculation of -- at most -- two families in a cycle of computation.

With the addition of one more family or word of memory, the accuracy and stability of any method can probably always be improved.

Method I. *Two iterations, incomplete, uncombined*

$$u^{(1)}_{n+2} = \alpha_2 u_{n+1} + \alpha_3 u_n + h(\alpha'_2 u'_{n+1} + \alpha'_3 u'_n)$$

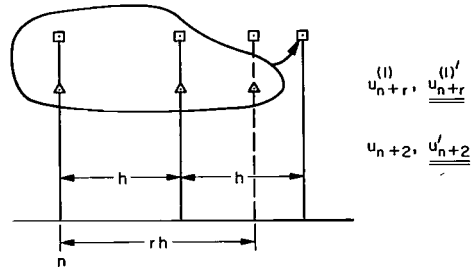
$$u_{n+2} = \beta'_1 h u^{(1)\prime}_{n+2} + \beta_2 u_{n+1} + \beta_3 u_n + h(\beta'_2 u'_{n+1} + \beta'_3 u'_n)$$



Method 2a Two iterations, incomplete, combined

$$u^{(1)}_{n+r} = \alpha_2 u_{n+1} + \alpha_3 u_n + h(\alpha'_2 u'_{n+1} + \alpha'_3 u'_n)$$

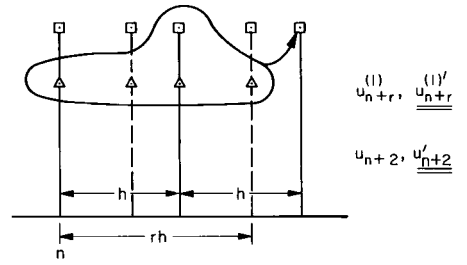
$$u_{n+2} = \beta'_1 h u^{(1)'}_{n+r} + \beta_2 u_{n+1} + \beta_3 u_n + h(\beta'_2 u'_{n+1} + \beta'_3 u'_n)$$



Method 2b Two iterations, incomplete, combined

$$u^{(1)}_{n+r} = \alpha_2 u_{n+1} + h(\alpha'_2 u'_{n+1} + \bar{\alpha}'_2 u^{(1)'}_{n+r-1} + \alpha'_3 u'_n)$$

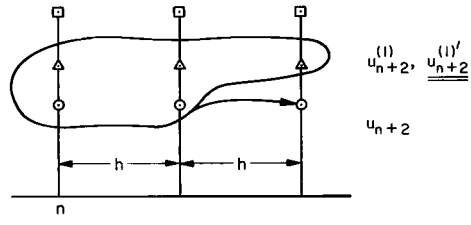
$$u_{n+2} = \bar{\beta}'_1 h u^{(1)'}_{n+r} + \beta_2 u_{n+1} + h(\beta'_2 u'_{n+1} + \bar{\beta}'_2 u^{(1)'}_{n+r-1} + \beta'_3 u'_n)$$



Method 3 One iteration, complete, uncombined

$$u^{(1)}_{n+2} = \alpha_2 u_{n+1} + \alpha_3 u_n + h(\alpha'_2 u'^{(1)'}_{n+1} + \alpha'_3 u'^{(1)'}_n)$$

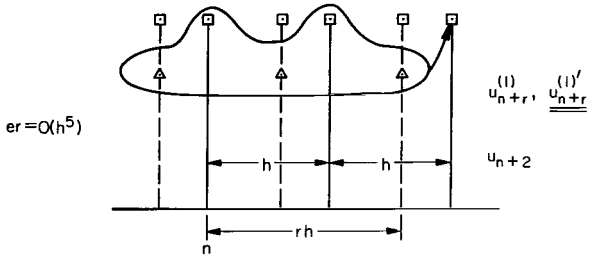
$$u_{n+2} = h\beta'_1 u^{(1)'}_{n+2} + \beta_2 u_{n+1} + \beta_3 u_n + h(\beta'_2 u^{(1)'}_{n+1} + \beta'_3 u^{(1)'}_n)$$



Method 4a. One iteration, complete, combined

$$u_{n+r}^{(1)} = \alpha_2 u_{n+1} + \alpha_3 u_n + h(\alpha_2' u_{n+r-1}^{(1)'} + \alpha_3' u_{n+r-2}^{(1)'})$$

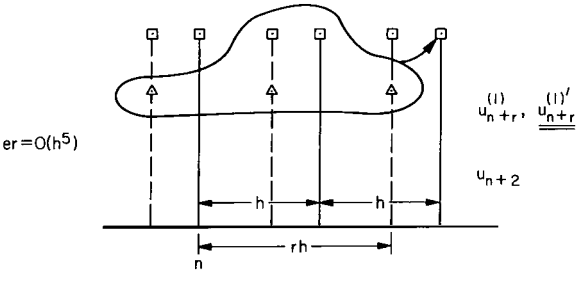
$$u_{n+2} = h\beta_1' u_{n+r}^{(1)'} + \beta_2 u_{n+1} + \beta_3 u_n + h(\beta_2' u_{n+r-1}^{(1)'} + \beta_3' u_{n+r-2}^{(1)'})$$



Method 4b. One iteration, complete, combined

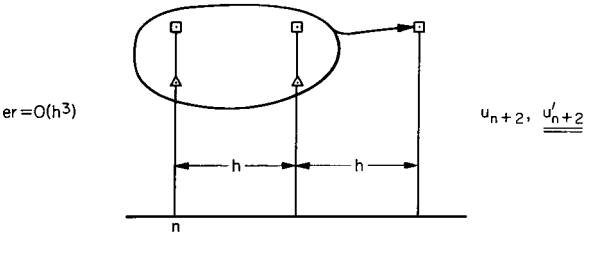
$$u_{n+r}^{(1)} = \alpha_2 u_{n+1} + \alpha_3 u_{n+r-1}^{(1)} + h(\alpha_2' u_{n+r-1}^{(1)'} + \alpha_3' u_{n+r-2}^{(1)'})$$

$$u_{n+2} = h\beta_1' u_{n+r}^{(1)'} + \beta_2 u_{n+1} + \beta_3 u_{n+r-1}^{(1)} + h(\beta_2' u_{n+r-1}^{(1)'} + \beta_3' u_{n+r-2}^{(1)'})$$



Method 5. One iteration, incomplete, uncombined

$$u_{n+2} = \alpha_2 u_{n+1} + \alpha_3 u_n + h(\alpha_2' u_{n+1}' + \alpha_3' u_n')$$



Notice that, in each case, a cycle of computation is completed when the integration has been advanced a distance h, and the number of iterations refers to the total number of evaluations of the derivative in this cycle. At first glance, it appears that the two iteration methods do not belong in the same group with the one-iteration methods because more arithmetic is certainly required to evaluate the derivative of the corrected function and this violates condition 2 of this section. However, this effect is taken care of by referring the stability and accuracy terms in all methods to the reference step size H. With this important qualification, all methods 1 through 4 can be compared on the basis of very nearly equal logical complexity,

storage and computation time, provided only that the calculations required to evaluate $F(x,u)$ in the equation $u' = F(x,u)$ are equal to or greater than those required to find $u^{(1)}_{n+r}$ and u_{n+2} in a cycle of computation.

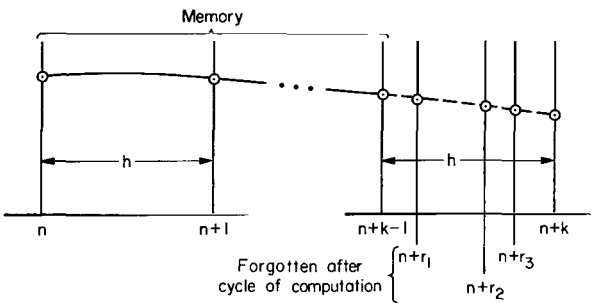
Method 1 represents the classical predictor-corrector sequence; for example, an Adams-Bashforth predictor followed by an Adams-Moulton corrector. One can show by applying the analysis presented in a previous section that no method of this type is stable for arbitrarily small but nonzero h and for arbitrary complex eigenvalues if the local error is $0(h^5)$. Method 2a is typical of the combined methods proposed in references 14, 15, and 16. At the beginning of a cycle of computation the function and its derivative are calculated at the point $n + r$, and the derivative is used to repredict the function at $n + 2$. Neither the function nor its derivative is retained in memory. This scheme can be used to develop stable methods with a local error of $0(h^5)$ (see eqs. (156)). The order of the error cannot be further increased since $0(h^5)$ is the highest order possible for any method with a five term corrector, and we are limited to correctors with a maximum of five terms by the conditions assumed. We can undoubtedly improve the magnitude of the error and increase the stability if we make methods (1) and (2) complete. However, this would necessitate the addition of another family -- again violating the assumed conditions. Method 2b falls into the same class as method 2a, but it retains in memory the value of the derivative of the function at the intermediate point, rather than the value of the function at n. A method of this type is studied under equations (159). It is the most accurate method of all those illustrated, having the leading error terms $er_\mu = 1/720(\mu H)^5$ and $er_\lambda = 1/720(\lambda H)^5$. Its induced stability boundary is $|\lambda H|_c \approx 0.3$.

Next, we consider one-iteration methods which can be made both complete and combined under the imposed conditions. The simplest one-iteration process is the incomplete, uncombined one composed of a single predictor, method 5 in our case. The accuracy is severely limited by the requirement of stability. If this process is made complete, as in method 3, the accuracy of stable methods can be increased by an order of magnitude -- see equations (96). If it is further combined with the Runge-Kutta techniques, as in method 4a, the error term for stable methods can be reduced to $0(h^5)$ just as in method 2a (see eqs. (152)). Another choice of data for a one-iteration, complete, combined method is presented as method 4b. Rather surprisingly, however, this choice is always unstable if the error is to be $0(h^5)$. Proof of the latter is given in this part commencing with equations (153).

A Special Class of Multistep, Multi-iteration Combined Methods

Let us next study a class of multi-iteration, combined methods. Consider a set of equations formed by using a memory of the function and its derivative at only equispace intervals, but, during a cycle of computation, predicting and correcting at several arbitrarily placed intermediate points. The analysis of such a process is rather simple and shows the connection between the classical predictor-corrector methods and methods for multiple numbers of steps and iterations.

Assume that values of u and u' have been computed or given at the k equispaced points hn, $h(n+1)$, . . ., $h(n+k-1)$ as in sketch (f). We seek to advance the solution one step to the point $h(n+k)$. We permit ourselves up to four iterations, or three correctors, but we allow for the possibility of evaluating the function and its derivative the first three times at the points $h(n+r_1)$, $h(n+r_2)$, and $h(n+r_3)$, where the r values need not be integers or even lie in the interval between $n+k-1$ and $n+k$. A set of expressions for four fundamental families can be written:



Sketch (f)

$$u^{(1)}_{n+r_1} = \sum_{j=2}^{k+1} (\alpha_j u_{n+J} + \alpha'_j h u'_{n+J}) \tag{124a}$$

$$u^{(2)}_{n+r_2} = \beta^*_1 u^{(1)\prime}_{n+r_1} + \sum_{j=2}^{k+1} (\beta_j u_{n+J} + \beta'_j h u'_{n+J}) \tag{124b}$$

$$u^{(3)}_{n+r_3} = \sum_{j=1}^{2} \gamma^*_j u^{(j)\prime}_{n+r_j} + \sum_{j=2}^{k+1} (\gamma_j u_{n+J} + \gamma'_j h u'_{n+J}) \tag{124c}$$

$$u_{n+k} = \sum_{j=1}^{3} \delta^*_j u^{(j)\prime}_{n+r_j} + \sum_{j=2}^{k+1} (\delta_j u_{n+J} + \delta'_j h u'_{n+J}) \tag{124d}$$

Equations (124) require four iterations per cycle of computation. All the following analysis applies directly to three-iteration, incomplete, combined methods if

$$\alpha_j = \alpha'_j = \beta^*_1 = \gamma^*_1 = \delta^*_1 = 0$$

and families 2 and 3 are replaced by 1 and 2, respectively. The two-iteration case follows by further reduction from the top.

The incomplete, uncombined, predictor-corrector methods discussed previously are obtained from equations (124) by setting $r_j = k$. Classical Runge-Kutta methods result if the memory is set equal to zero for $j > 2$. Specifically, the standard, fourth-order, Runge-Kutta method results when

(1) $k = 1$

(2) $r_1 = r_2 = \frac{1}{2}$, $r_3 = 1$

A detailed discussion of this method is presented in the next section of this part.

The matrix equation for the combined, incomplete methods defined by (124) follows exactly as it was developed in the previous parts. Extending the notation in equations (81) to include terms with γ and δ, we find

$$\begin{bmatrix} E^{r_1} & 0 & 0 & -C_\alpha \\ -\lambda h\beta_1^* E^{r_1} & E^{r_2} & 0 & -C_\beta \\ -\lambda h\gamma_1^* E^{r_1} & -\lambda h\gamma_2^* E^{r_2} & E^{r_3} & -C_\gamma \\ -\lambda h\delta_1^* E^{r_1} & -\lambda h\delta_2^* E^{r_2} & -\lambda h\delta_3^* E^{r_3} & E^k - C_\delta \end{bmatrix} \begin{bmatrix} u_n^{(1)} \\ u_n^{(2)} \\ u_n^{(3)} \\ u_n \end{bmatrix} = Ahe^{\mu hn} \begin{bmatrix} F_\alpha \\ \beta_1^* e^{\mu h r_1} + F_\beta \\ \sum_{j=1}^{2} \gamma_j^* e^{\mu h r_j} + F_\gamma \\ \sum_{j=1}^{3} \delta_j^* e^{\mu h r_j} + F_\delta \end{bmatrix} \tag{125}$$

Now if we seek only u_n, each term E^{r_j} common to a column can be factored out; and, regardless of the choice of r_j, the operational form has only integer exponents of E. Using straightforward algebraic manipulations, one can derive the expressions defined in equations (52).

$$DE(E) = E^k - \sum_{j=2}^{k+1} E^J \sum_{m=1}^{5} L_{mj}(\lambda h)^{m-1} \tag{126a}$$

$$NU = h(\mu - \lambda)\left[\sum_{m=1}^{3} (\lambda h)^{m-1} \sum_{j=1}^{4-m} R_{mj}^* e^{\mu h r_j} + \sum_{j=2}^{k+1} e^{J\mu h} \sum_{m=1}^{4} R_{mj}(\lambda h)^{m-1}\right] \tag{126b}$$

The coefficients in the operational form are

$$\left.\begin{aligned} R_{1j}^* &= \delta_j^* \\ R_{21}^* &= \delta_3^*\gamma_1^* + \delta_2^*\beta_1^* \\ R_{22}^* &= \delta_3^*\gamma_2^* \\ R_{31}^* &= \delta_3^*\gamma_2^*\beta_1^* \end{aligned}\right\} \tag{127a}$$

$$\left.\begin{aligned} R_{1j} &= \delta_j' \\ R_{2j} &= \alpha_j'\delta_1^* + \beta_j'\delta_2^* + \gamma_j'\delta_3^* \\ R_{3j} &= \alpha_j'(\beta_1^*\delta_2^* + \gamma_1^*\delta_3^*) + \beta_j'\gamma_2^*\delta_3^* \\ R_{4j} &= \alpha_j'\beta_1^*\gamma_2^*\delta_3^* \end{aligned}\right\} \tag{127b}$$

$$\left.\begin{aligned} L_{1j} &= \delta_j \\ L_{2j} &= R_{1j} + \alpha_j\delta_1^* + \beta_j\delta_2^* + \gamma_j\delta_3^* \\ L_{3j} &= R_{2j} + \alpha_j(\beta_1^*\delta_2^* + \gamma_1^*\delta_3^*) + \beta_j\gamma_2^*\delta_3^* \\ L_{4j} &= R_{3j} + \alpha_j\beta_1^*\gamma_2^*\delta_3^* \\ L_{5j} &= R_{4j} \end{aligned}\right\} \tag{127c}$$

They can be inverted (provided R_{31}^*, R_{22}^* and R_{13}^* are not zero) to form the expressions

$$\left.\begin{aligned} \beta_1^* &= R_{31}^*/R_{22}^* \\ \gamma_1^* &= (R_{21}^* - \beta_1^*R_{12}^*)/R_{13}^* \\ \gamma_2^* &= R_{22}^*/R_{13}^* \\ \delta_j^* &= R_{1j}^* \end{aligned}\right\} \tag{128}$$

$$\left.\begin{aligned} \alpha_j &= (L_{4j} - R_{3j})/R_{31}^* \\ \beta_j &= (L_{3j} - R_{2j} - \alpha_jR_{21}^*)/R_{22}^* \\ \gamma_j &= (L_{2j} - R_{1j} - \alpha_jR_{11}^* - \beta_jR_{12}^*)/R_{13}^* \end{aligned}\right\} \tag{129}$$

$$\left.\begin{aligned} \delta_j &= L_{1j} \\ \alpha_j' &= L_{5j}/R_{31}^* \\ \beta_j' &= (R_{3j} - \alpha_j'R_{21}^*)/R_{22}^* \\ \gamma_j' &= (R_{2j} - \alpha_j'R_{11}^* - \beta_j'R_{12}^*)/R_{13}^* \\ \delta_j' &= R_{1j} \end{aligned}\right\} \tag{130}$$

Again we use equation (58) to calculate the error in the particular solution. Expanding the numerator in powers of λh we find the conditions for making the coefficients to $(\lambda h)^j$ vanish for $j = 0, \ldots, 3$. Since $L_{5j} = R_{4j}$, the coefficient to $(\lambda h)^4$ is identically zero.

$$\sum_{j=2}^{k+1}\left[l(k+1-j)^{l-1}R_{1j}+(k+1-j)^{l}L_{1j}\right]+\sum_{j=1}^{3}lr_j^{l-1}R^*_{1j}-k^l=0\ ,$$

$$l = 0, 1, \ldots, L \qquad (131)$$

$$\sum_{j=2}^{k+1}\left[l(k+1-j)^{l-1}R_{2j}+(k+1-j)^{l}(L_{2j}-R_{1j})\right]-\sum_{j=1}^{3}\left[r_j^{l}R^*_{1j}-lr_j^{l-1}R^*_{2j}\right]=0\ ,$$

$$l = 0, 1, \ldots, L-1 \qquad (132)$$

$$\sum_{j=2}^{k+1}\left[l(k+1-j)^{l-1}R_{3j}+(k+1-j)^{l}(L_{3j}-R_{2j})\right]-\sum_{j=1}^{2}\left[r_j^{l}R^*_{2j}-lr_j^{l-1}R^*_{3j}\right]=0\ ,$$

$$l = 0, 1, \ldots, L-2 \qquad (133)$$

$$\sum_{j=2}^{k+1}\left[l(k+1-j)^{l-1}R_{4j}+(k+1-j)^{l}(L_{4j}-R_{3j})\right]-r_1^{l}R^*_{31}=0\ ,$$

$$l = 0, 1, \ldots, L-3 \qquad (134)$$

These are the accuracy conditions for equations (124). Once more one can show that both the particular and complementary solutions are fit locally by polynomials of order L if the above conditions are satisfied. Except for the addition of one more term (which is multiplied by $\lambda^3h^3(\mu h)^{L-2}/(L-3)!$) the value of er_μ is determined just as in the discussion under equation (117c). The error in the complementary solution is

$$er_\lambda=\left\{\frac{(\lambda h)^l}{l!}\sum_{j=2}^{k+1}\left[(k+1-j)^{l}L_{1j}+l(k+1-j)^{l-1}L_{2j}+l(l-1)(k+1-j)^{l-2}L_{3j}\right.\right.$$

$$\left.\left.+l(l-1)(l-2)(k+1-j)^{l-3}L_{4j}+l(l-1)(l-2)(l-3)(k+1-j)^{l-4}L_{5j}\right]-k^l\right\}\Bigg/\sum_{j=2}^{k+1}(j-1)L_{1j} \qquad (135)$$

To compare with other methods all error terms should be expressed in terms of the representative step H where

$$h = 2H$$

since four iterations are made to advance one step h.

Accuracy of the Standard Fourth-Order Runge-Kutta Method

The standard Runge-Kutta formula (see, e.g., ref. 17) can be written in the predictor-corrector notation used in equation (124). It is

$$\left.\begin{aligned} u_{n+0.5}^{(1)} &= u_n + \frac{1}{2} h u_n' \\ u_{n+0.5}^{(2)} &= u_n + \frac{1}{2} h u_{n+0.5}^{(1)'} \\ u_{n+1}^{(3)} &= u_n + 2\left(\frac{1}{2} h\right) u_{n+0.5}^{(2)'} \\ u_{n+1} &= u_n + \frac{1}{3}\left(\frac{1}{2} h\right)\left[u_{n+1}^{(3)'} + 2\left(u_{n+0.5}^{(2)'} + u_{n+0.5}^{(1)'}\right) + u_n'\right] \end{aligned}\right\} \tag{136}$$

If $\frac{1}{2} h$ is replaced by h, $n + 0.5$ by $n + 1$, and $n + 1$ by $n + 2$, these equations are immediately recognized in predictor-corrector terminology to be an Euler predictor, an Euler corrector, a Nystrom predictor, and a Milne corrector, respectively. Families (1), (2), and (3) are not so accurate as the final family and are discarded at the end of each cycle of computation.

When equations (124) and (136) are compared, it is clear that

$$r_1 = r_2 = \frac{1}{2}$$

$$r_3 = 1$$

and the coefficients in the difference-differential equations are

$$\left.\begin{aligned} &\alpha_2 = \beta_2 = \gamma_2 = \delta_2 = 1 \\ &\alpha_2' = \frac{1}{2}, \quad \beta_2' = \gamma_2' = 0, \quad \delta_2' = \frac{1}{6} \\ &\beta_1^* = \frac{1}{6} \\ &\gamma_1^* = 0, \quad \gamma_2^* = 1 \\ &\delta_1^* = \delta_2^* = \frac{1}{3}, \quad \delta_3^* = \frac{1}{6} \end{aligned}\right\} \tag{137}$$

The coefficients in the operational form are

m	L_{m2}	R^*_{m1}	R^*_{m2}	R^*_{m3}	R_{m2}
1	1	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{6}$
2	1	$\frac{1}{6}$	$\frac{1}{6}$		$\frac{1}{6}$
3	$\frac{1}{2}$	$\frac{1}{12}$			$\frac{1}{12}$
4	$\frac{1}{6}$				$\frac{1}{24}$
5	$\frac{1}{24}$				

(138)

The error terms are easily calculated. Making use of equation (99), one finds for the error in the particular solution

$$er_\mu = \frac{-\mu^2 H^5}{180} [\mu^3 - 5\lambda\mu^2 + 10\lambda^2\mu - 30\lambda^3] \tag{139}$$

and for the error in the complementary solution

$$er_\lambda = \frac{24\lambda^5 H^5}{180} \tag{140}$$

Notice that the fourth-order Runge-Kutta method is more accurate as a simple integrator ($\lambda = 0$, $er_\mu = -(\mu H)^5/180$) than it is as a differential analyzer ($\mu = 0$, $er_\lambda = 24(\lambda H)^5/180$). On the basis of lowest order error estimates (since all methods are referenced to H, they are directly comparable) it is not so good as Hamming's unmodified method (table II(b)), and a full order of magnitude worse than Hamming's modified method (table II(c)). But, of course, when expressed in fundamental families, these are four- and five-step methods, respectively. It is comparable in accuracy ($46/180 \approx 0.255$) to a three-step, Adams-Bashforth-Moulton, predictor-corrector combination (table II(b)).

Stability of Runge-Kutta Methods

The fourth-order Runge-Kutta approximation to the complementary solution of the representative equation can be constructed at once from (138) and is

$$\left\{E - 1 - \lambda h - \frac{1}{2}\lambda^2 h^2 - \frac{1}{6}\lambda^3 h^3 - \frac{1}{24}\lambda^4 h^4\right\} u_n = 0$$

There is only one root, the principal one. At first, it might seem that stability is not an issue in such a method since there are no spurious roots and the principal root approximates $e^{\lambda h}$ which certainly falls on or inside the unit circle for negative λ and small enough h. However, if we consider the accuracy and stability criterion given by conditions (74b), the question

of stability again arises, even for one-root methods. In fact, in cases for which this criterion holds (i.e., when a negative eigenvalue in a set of differential equations has a large absolute value relative to those that are driving the system), the principal root itself may determine the stability boundary.

The variation of the principal root in the vicinity of the unit circle is shown in figure 10 for these one-root, one-step methods. In each case the points are separated by an increment of 0.1 in h, and $|\lambda| = |e^{i\omega}| = 1$. One reference value of h is given in each set of points so quantitative estimates can be made. As indicated, ω varies from $\pi/2$ to π in steps of $\pi/10$.

Figure 10(a) shows the principal root behavior for

$$\lambda_1 = 1 + \lambda h + \frac{1}{2}\lambda^2 h^2$$

which results from the method formed by combining an Euler predictor with a modified Euler corrector. Such a method is self-starting and extremely easy to program. It is often used in exploratory numerical research. For real negative λ the method is seen to be stable for $0 < -\lambda h = -\lambda H < 2$. For imaginary λ, however, the principal root actually falls outside the unit circle for all $|\lambda h| > 0$. For imaginary λ and $h < 0.5$ the method is accurate enough so that more than 200 steps would be required for the instability to become serious for most cases. Nevertheless, strictly speaking, its stability boundary is zero and, as a completely general method, it should be used with caution. (It is unsatisfactory, for example, for studying problems containing high-frequency, low-amplitude noise.)

Figure 10(b) shows the principal root behavior for

$$\lambda_1 = 1 + \lambda h + \frac{1}{2}\lambda^2 h^2 + \frac{1}{6}\lambda^3 h^3$$

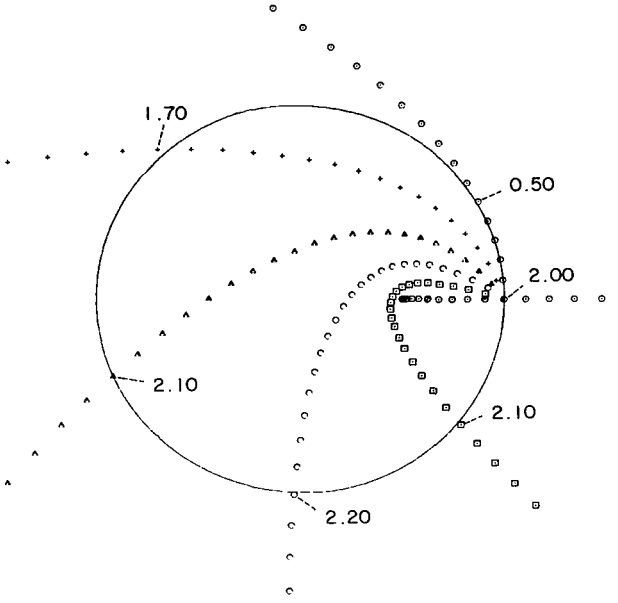
which results from third-order Runge-Kutta methods (e.g., Heun's method, see ref. 4, page 236). For real negative λ this method is stable for $0 < -\lambda h = -1.5\lambda H < 2.5$. Now when λ is imaginary the principal root falls inside the unit circle until $h \approx 1.7$. This method has the stability boundary $|\lambda h|_c = |1.5\lambda H|_c = 1.7$.

Finally, figure 10(c) shows the behavior for

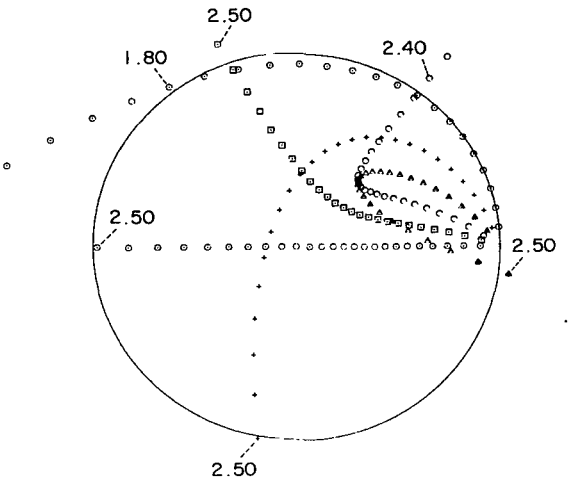
$$\lambda_1 = 1 + \lambda h + \frac{1}{2}\lambda^2 h^2 + \frac{1}{6}\lambda^3 h^3 + \frac{1}{24}\lambda^4 h^4$$

which results from the standard, fourth-order Runge-Kutta method represented by equations (136). This method shows excellent stability for all complex λ. For real negative λ the method is stable for $-\lambda h = -2\lambda H < 2.8$. The worst case occurs when $\omega \approx 0.8\pi$ and limits the stability boundary to $|\lambda h|_c = |2\lambda H|_c = 2.6$.

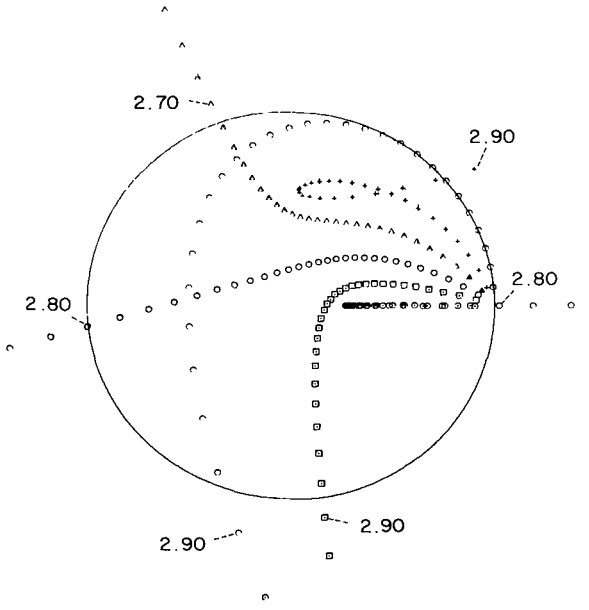
In conclusion, the fourth-order Runge-Kutta method is accurate, easy to program, has a higher stability boundary (even when compared on the basis



(a) Euler predictor followed by a modified Euler corrector.



(b) Heun's method.



(c) Fourth-order Runge-Kutta method.

$\lambda = e^{i\omega}$

Symbol	ω
⊙	$i\pi$
□	$0.9i\pi$
○	$0.8i\pi$
△	$0.7i\pi$
+	$0.6i\pi$
⊙	$0.5i\pi$

Figure 10.- Principal root structure of three well-known one-root, one-step methods.

of H) than any of the predictor, one-corrector combinations given in table I, and presents no problem in starting or step modification. For many practical purposes it is quite satisfactory.

The Four Iteration, One-Step Incomplete Method in General

In the simple one-step method the equation for the error in the complementary solution, equation (135), takes a remarkably simple form. Noting from the derivation that $(0)^0 = 1$, we see that if $er_\lambda \to 0(h^5)$, we must have

$$L_{m2} = \frac{1}{(m-1)!} \qquad m = 1, \ldots, 5 \tag{141}$$

regardless of the choice of any of the other parameters. These values of L_{m2} are the first five coefficients of x^j in the expansion of e^x. Hence, er_λ is always given to the lowest order by

$$er_\lambda = \frac{(\lambda h)^5}{5!} = \frac{24(\lambda H)^5}{180} \tag{142}$$

regardless of the choice of the sampling points r_1, r_2, and r_3. This proves that there exists in equations (124) no other one-step, incomplete four-iteration method that is more accurate than the standard, fourth-order Runge-Kutta method (given by eqs. (136)) in calculating the complementary solution. There can be improvements in the accuracy of the particular solution, but, since equations (139) and (140) show the error in the complementary solution is nearly the largest, methods that provide these improvements are of limited interest.

Multistep, One-Iteration, Complete Combined Methods

In this part let us consider the multistep form of the methods described above as methods 4a and 4b. These methods are both complete and combined, but are easy to analyze and instructive.

In terms of fundamental families, the difference-differential equations for the multistep form of method 4a are

$$\left.\begin{aligned} u^{(1)}_{n+r} &= \sum_{j=2}^{k+1} \left(\alpha_j u_{n+J} + \alpha'_j h u^{(1)'}_{n+r+1-j} \right) \\ u_{n+k} &= h\beta'_1 u^{(1)'}_{n+r} + \sum_{j=2}^{k+1} \left(\beta_j u_{n+J} + \beta'_j h u^{(1)'}_{n+r+1-j} \right) \end{aligned}\right\} \tag{143}$$

Inserting equation (37) and using operational notation, one can show

$$\begin{bmatrix} E^{r-k}\left(E^{k} - \lambda h \sum_{j=2}^{k+1} \alpha'_j E^{J}\right) & -\sum_{j=2}^{k+1} \alpha_j E^{J} \\ -E^{r-k}\lambda h \sum_{j=1}^{k+1} \beta'_j E^{J} & E^{k} - \sum_{j=2}^{k+1} \beta_j E^{J} \end{bmatrix} \begin{bmatrix} u_n^{(1)} \\ \\ u_n \end{bmatrix} = Ahe^{\mu hn} e^{\mu h(r-k)} \begin{bmatrix} \sum_{j=2}^{k+1} \alpha'_j e^{\mu hJ} \\ \\ \sum_{j=1}^{k+1} \beta'_j e^{\mu hJ} \end{bmatrix} \tag{144}$$

Compare this with equation (90). If $\bar{\alpha}_j$ and $\bar{\beta}_j$ are set equal to zero, and the bars are deleted from the primed terms, the equations are identical except for the term E^{r-k} multiplying the left column and the term $e^{\mu h(r-k)}$ multiplying the entire right-hand side. The characteristic equations for the two methods differ only by the factor E^{r-k}, which means that for given values of α and β, the roots are identical. Further, if we solve only for the final family, the factor E^{r-k} has no effect on the particular integral, since it appears in both the numerator and denominator of the solution. Finally, the term $e^{\mu h(r-k)}$ simply multiplies the particular integral, so the complete solutions to equation (144) can at once be written from the analysis of equation (90); there results

$$\left\{E^{2k} - \sum_{j=2}^{2k+1} (L_{1j} + \lambda h L_{2j}) E^{2k+1-j}\right\} u_n = hAe^{\mu hn} \sum_{j=1}^{k+1} R_{1j} e^{\mu h(k+r+1-j)} \tag{145}$$

where

$$\left.\begin{aligned} L_{1j} &= \beta_j , && j = 2, k+1 \\ R_{1j} &= \beta'_j , && j = 1, k+1 \\ L_{2j} &= \alpha'_j + \beta'_1 \alpha_j + \sum_{i=2}^{j-1} (\beta'_i \alpha_{j+1-i} - \alpha'_i \beta_{j+1-i}) , && j = 2, 2k+1 \end{aligned}\right\} \tag{146}$$

The inverted equations, giving α and β in terms of L and R, are the simultaneous, linear equations in (94).

The only influence of r is contained in the exponent of e in the right-hand side of equation (145). However, this has a profound effect on the results. The equations which must be satisfied for a local polynomial fit of order L are now

$$\sum_{j=1}^{k+1} \left[l(k + r + 1 - j)^{l-1} R_{1j} + (2k + 1 - j)^{l} L_{1j} \right] - (2k)^{l} = 0 ,$$

$$l = 0, 1, \ldots, L \qquad (147a)$$

$$\sum_{j=1} \left[(2k + 1 - j)^{l} L_{2j} - R_{1j}(k + r + 1 - j)^{l} \right] = 0 ,$$

$$l = 0, 1, \ldots, L - 1 \qquad (147b)$$

instead of (66) and (67), respectively. The errors are given by evaluating the left side of equations (147) when they are no longer zero and substituting the results for the terms inside {} in equations (64) and (65).

To get some idea of the effect of r on the accuracy and stability, one can construct tables for the coefficients of L and R for k = 2 and compare them with table III. Thus for equations (147) with k = 2:

$er_{\mu 1}$

l	R_{11}	R_{12}	R_{13}	L_{12}	L_{13}	$(4)^l$
0	0	0	0	1	1	1
1	1	1	1	3	2	4
2	2(2+r)	2(1+r)	2r	9	4	16
3	$3(2+r)^2$	$3(1+r)^2$	$3r^2$	27	8	64
4	$4(2+r)^3$	$4(1+r)^3$	$4r^3$	81	16	256
5	$5(2+r)^4$	$5(1+r)^4$	$5r^4$	243	32	1024

$er_{\mu 2}$

l	L_{22}	L_{23}	L_{24}	L_{25}	R_{11}	R_{12}	R_{13}
0	1	1	1	1	1	1	1
1	3	2	1	0	2+r	1+r	r
2	9	4	1	0	$(2+r)^2$	$(1+r)^2$	r^2
3	27	8	1	0	$(2+r)^3$	$(1+r)^3$	r^3
4	81	16	1	0	$(2+r)^4$	$(1+r)^4$	r^4

Note that this table reduces to portions of table III when $r = 2$. A table for er_λ is not constructed, since it is still true that $er_\lambda = er_\mu$ when $\lambda = \mu$.

When $k = 2$, the characteristic equation for (144) is

$$E^4 - E^3(L_{12} + \lambda h L_{22}) - E^2(L_{13} + \lambda h L_{23}) - \lambda h[L_{24}E + L_{25}] = 0 \qquad (148)$$

and the stability at $\lambda h = 0$ is given by the roots to the quadratic

$$E^2 - L_{12}E - L_{13} = 0 \qquad (149)$$

If we satisfy equations (147) for $L = 4$ and various r values between 1.6 and 2.0, we can construct the curves for L_{13}, $er_{\mu 1}$, and $er_{\mu 2}$ shown in figure 11. The error expressed by $er_{\mu 2}$ completely dominates the accuracy of the method and varies from about $-0.17\lambda\mu^4h^5$ at $r = 2$ to about $-1.3\lambda\mu^4h^5$ at $r = 1.6$. The variation of er_μ shown by the solid curve is derived using only the first term in the truncation error. The actual values of $er_\lambda \times 10^5$ (for $\lambda h = 0.1$), plotted in the circles, indicate that the first term in the truncation error is accurate for values of $\lambda h = 0.1$ and lower. At around $r = 1.635$, L_{13} goes to zero and the three spurious roots all lie on the origin when $\lambda h = 0$, giving Adams-Moulton type stability. The critical stability boundary $|\lambda h|_c$, see (73), is around 0.3 when L_{13} is zero. This boundary is determined from plots such as those shown in figure 12.

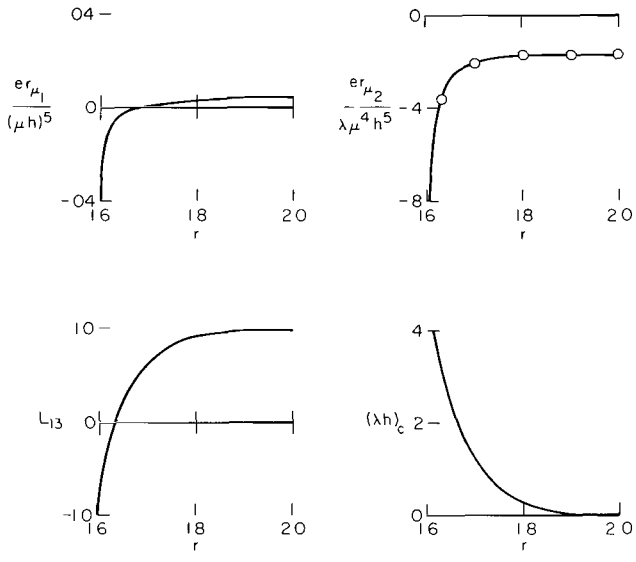


Figure 11.- Variation of terms controlling stability and accuracy of methods given by equations (143).

All things taken into consideration, the value $r = 1.635$ appears to be a good compromise for this method. If we reference the accuracy and stability to H (= 2h) -- according to equation (99) -- we find

$$|er_\mu| \approx \frac{1}{3}\left(\frac{\lambda\mu^4H^5}{16}\right) = \frac{1}{48}\lambda\mu^4H^5 \approx 0.0208\lambda\mu^4H^5 \qquad (150)$$

$$0 < (-\lambda H)_c \approx 0.6 \qquad (151)$$

The coefficients in the operational form are, again for $r = 1.635$

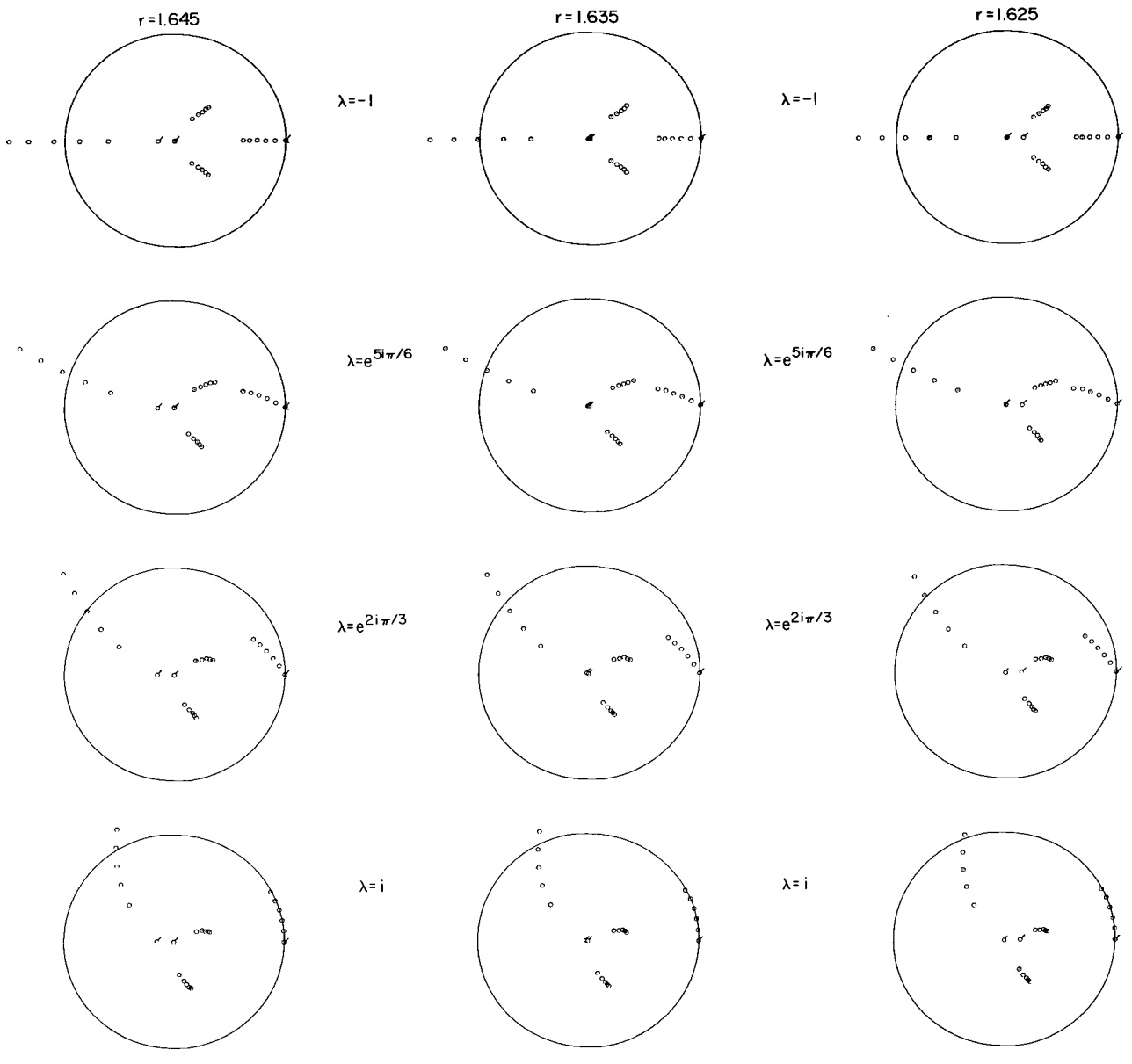


Figure 12.- Typical stability plots for two-step form of equations (143).

j \ m	L_{mj}			
	2	3	4	5
1	0.98446029	0.01553971	0	0
2	2.29749376	-2.44603026	1.53842844	-0.37435229

j \ m	R_{mj}				
	1	2	3	4	5
1	0.84801929	0.18240329	-0.01488289	0	0

and using equations (94), one finds the corresponding difference-differential equations

$$\left.\begin{aligned} u^{(1)}_{n+1.635} &= \quad + \alpha_2 u_{n+1} + \alpha_3 u_n + \alpha_2' h u^{(1)'}_{n+0.635} + \alpha_3' h u^{(1)'}_{n-0.365} \\ u_{n+2} &= \beta_1' h u^{(1)'}_{n+1.635} + \beta_2 u_{n+1} + \beta_3 u_n + \beta_2' h u^{(1)'}_{n+0.635} + \beta_3' h u^{(1)'}_{n-0.365} \end{aligned}\right\} \tag{152a}$$

where

$$\left.\begin{aligned} \alpha_2 &= -21.44241590 & \beta_2 &= 0.98446029 \\ \alpha_3 &= 22.44241590 & \beta_3 &= 0.01553971 \\ \alpha_2' &= 20.48107725 & \beta_1' &= 0.84801929 \\ \alpha_3' &= 2.59633865 & \beta_2' &= 0.18240329 \\ & & \beta_3' &= -0.01488289 \end{aligned}\right\} \tag{152b}$$

A proof that method 4b is unstable for $O(h^5)$ proceeds along the following lines. The matrix equation for the operational form of the method, as applied to the representative equation (37), can be written

$$\begin{bmatrix} E^{r-2}\{E^2-(\alpha_3+\lambda h\alpha_2')E-\lambda h\alpha_3'\} & -\alpha_2 E \\ -E^{r-2}\{\lambda h\beta_1' E^2+(\beta_3+\lambda h\beta_2')E+\lambda h\beta_3'\} & E(E-\beta_2) \end{bmatrix} \begin{bmatrix} u_n^{(1)} \\ u_n \end{bmatrix} = Ahe^{\mu h n} e^{\mu h(r-2)} \begin{bmatrix} \sum_{j=2}^{3} \alpha_j' e^{\mu h(3-j)} \\ \sum_{j=1}^{3} \beta_j' e^{\mu h(3-j)} \end{bmatrix} \tag{153}$$

One can easily show that the characteristic equation, $DE(E) = 0$, reduces to

$$E^3 - E^2(L_{12} + \lambda h L_{22}) - E(L_{13} + \lambda h L_{23}) - \lambda h L_{24} = 0 \tag{154}$$

where L are simple combinations of the α and β. Now the error in the complementary solution is determined by substituting $e^{\lambda h}$ for E in equation (154) and finding to what order the expanded result does not match the expansion of $e^{\lambda h}$ itself. This simply amounts to entering table IV for $k = 3$ and finding the first row that does not sum to zero. If we are to make the first five rows all sum to zero, so the order of the error is h^5, all five values of L are completely determined; they must satisfy the equations

$$\left.\begin{aligned} L_{12} &= -8 \\ L_{13} &= 9 \\ L_{22} &= 17/3 \\ L_{23} &= 14/3 \\ L_{24} &= -1/3 \end{aligned}\right\} \qquad (155)$$

regardless of the values chosen for r and R. Using equations (155) and setting $h = 0$, we find equation (154) reduces to

$$E^2 + 8E - 9 = (E - 1)(E + 9) = 0$$

which has a violent instability given by the root $E = -9$.

Two-Step, Two-Iteration, Incomplete, Combined Methods

The final section of this part is devoted to a discussion of methods 2a and 2b. Method 2a is really a special form of equations (124) when the latter are simplified to two iterations, or one value of r. Applying the analysis of those equations, one can show that all methods of type 2a having an error of $O(h^5)$ have operational coefficients given by (the * has been omitted from R^*_{11} in the following)

m \ j	L_{mj}		R_{mj}		
	2	3	1	2	3
1	$\frac{8(r-2)}{1-2r}$	$\frac{17-10r}{1-2r}$	$\frac{2}{r(1-r)(1-2r)}$	$\frac{2(4r^2-9r+4)}{(1-r)(1-2r)}$	$\frac{-2(2r^2-4r+1)}{r(1-2r)}$
2	$\frac{-4(r-2)}{1-2r}$	$\frac{2(5-4r)}{1-2r}$	0	$\frac{-2r}{1-2r}$	$\frac{2(1-r)}{1-2r}$
3	$\frac{-2r}{1-2r}$	$\frac{2(1-r)}{1-2r}$			

and difference-differential equations given by

$$\left.\begin{aligned} u^{(1)}_{n+r} &= \alpha_2 u_{n+1} + \alpha_3 u_n + h(\alpha'_2 u'_{n+1} + \alpha'_3 u'_n) \\ u_{n+2} &= \beta_2 u_{n+1} + \beta_3 u_n + h\left(\beta'_1 u^{(1)'}_{n+r} + \beta'_2 u'_{n+1} + \beta'_3 u'_n\right) \end{aligned}\right\} \tag{156a}$$

where

$$\left.\begin{aligned} \alpha_2 &= r^2(3 - 2r) & \beta_2 &= 8(r - 2)/(1 - 2r) \\ \alpha_3 &= (1 + 2r)(1 - r)^2 & \beta_3 &= (17 - 10r)/(1 - 2r) \\ \alpha'_2 &= -r^2(1 - r) & \beta'_1 &= 2/[r(1 - r)(1 - 2r)] \\ \alpha'_3 &= r(1 - r)^2 & \beta'_2 &= 2[4r^2 - 9r + 4]/[(1 - r)(1 - 2r)] \\ & & \beta'_3 &= -2[2r^2 - 4r + 1]/[r(1 - 2r)] \end{aligned}\right\} \tag{156b}$$

The leading error terms are

$$\left.\begin{aligned} \mathrm{er}_\mu &= \frac{[(5r^2 - 11r + 4)\mu + 5r(1 - r)\lambda]\mu^4 H^5}{180[2(2r - 3)]} \\ \mathrm{er}_\lambda &= \frac{(2 - 3r)\lambda^5 H^5}{180(2r - 3)} \end{aligned}\right\} \tag{157}$$

and the characteristic equation at $h = 0$ is given by

$$(E - 1)\left(E + \frac{17 - 10r}{1 - 2r}\right) = 0 \tag{158}$$

so the spurious root starts at

$$\lambda_2 = -\frac{17 - 10r}{1 - 2r}$$

which vanishes at $r = 1.7$, giving Adams-Moulton type stability there. A table of the error terms for various values of r is included and stability

r	$\frac{\mathrm{er}_{\mu 1}}{\mu^5 H^5}$	$\frac{\mathrm{er}_{\mu 2}}{\lambda\mu^4 H^5}$	$\frac{\mathrm{er}_\lambda}{\lambda^5 H^5}$
2.00	0.0056	-0.0278	-0.0222
1.90	.0040	-.0297	-.0257
1.85	.0030	-.0312	-.0282
1.80	.0019	-.0333	-.0314
1.75	.0003	-.0364	-.0361
1.70	-.0017	-.0413	-.0430

plots in the same range are shown in figure 13. On inspection, the figure shows that the choice $r = 1.75$ gives the most stable numerical method, having an induced stability boundary $|\lambda H|_c = 0.6$.

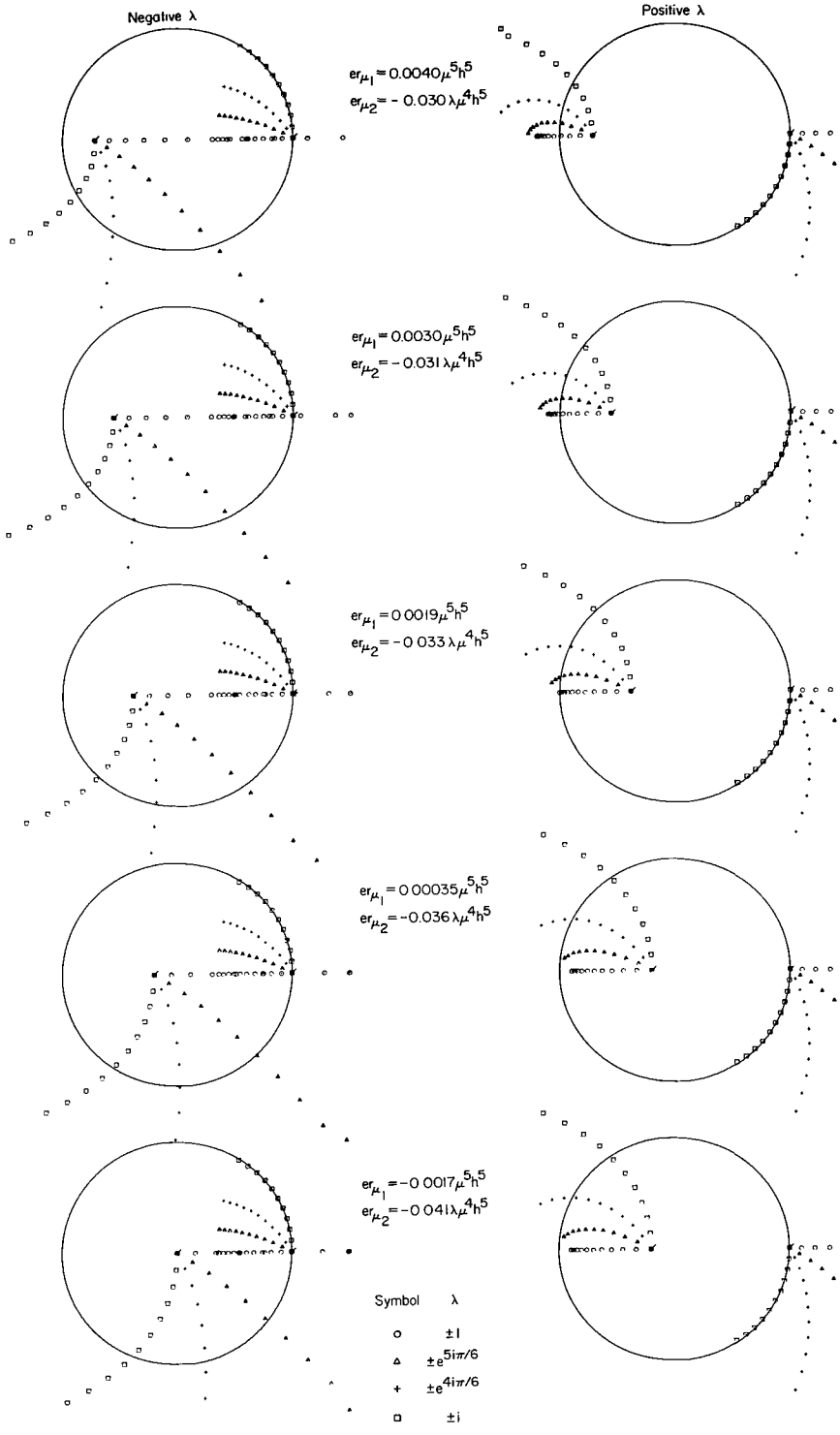


Figure 13.- Stability plots for the method defined by equations (156).

Method 2b can be written

$$\left.\begin{aligned} u_{n+r}^{(1)} &= \alpha_2 u_{n+1} + h\left(\alpha_2' u_{n+1}' + \alpha_3' u_n' + \bar{\alpha}_2' u_{n+r-1}^{(1)'}\right) \\ u_{n+2} &= \beta_2 u_{n+1} + h\left(\beta_2' u_{n+1}' + \beta_3' u_n' + \bar{\beta}_1' u_{n+r}^{(1)'} + \bar{\beta}_2' u_{n+r-1}^{(1)'}\right) \end{aligned}\right\} \tag{159}$$

and reduced to the operational form

$$\{E^3 - E^2(L_{12} + \lambda h L_{22} + \lambda^2 h^2 L_{32}) - E(\lambda h L_{23} + \lambda^2 h^2 L_{33}) - \lambda^2 h^2 L_{34}\} u_n$$

$$= Ahe^{\mu h n}[(R_{11} + \lambda h R_{21})e^{2\mu h} + (R_{12} + \lambda h R_{22})e^{\mu h} + \lambda h R_{23} + \bar{R}_{11} e^{\mu h(r+1)} + \bar{R}_{12} e^{\mu h r}] \tag{160}$$

where

$$\left.\begin{aligned} R_{11} &= \beta_2' & L_{12} &= \beta_2 \\ R_{12} &= \beta_3' & L_{22} &= \beta_2' + \bar{\alpha}_2' + \alpha_2 \bar{\beta}_1' \\ \bar{R}_{11} &= \bar{\beta}_1' & L_{32} &= \alpha_2' \bar{\beta}_1' \\ \bar{R}_{12} &= \bar{\beta}_2' & L_{23} &= \beta_3' - \beta_2 \bar{\alpha}_2' + \bar{\beta}_2' \alpha_2 \\ R_{21} &= L_{32} & L_{33} &= -\beta_2' \bar{\alpha}_2' + \bar{\beta}_1' \alpha_3' + \bar{\beta}_2' \alpha_2' \\ R_{22} &= L_{33} & L_{34} &= \bar{\beta}_2' \alpha_3' - \beta_3' \bar{\alpha}_2' \\ R_{23} &= L_{34} \end{aligned}\right\} \tag{161}$$

The accuracy conditions formed by collecting the terms independent of λh in equation (160) are

$$er_{\mu 1} = \frac{(\mu h)^l}{l!}\left\{\sum_{j=1}^{2}\left[R_{1j} l (3-j)^{l-1} + \bar{R}_{1j} l (r+2-j)^{l-1}\right] + 2^l L_{12} - 3^l\right\}, \quad l = 0, 1, \ldots \tag{162a}$$

The term corresponding to $\sum_{j=2}^{k+1}(j-1)L_{1j}$ in equations (64) and (65) is unity.

The second set of conditions formed when the coefficients to λh are collected in equation (166) are

$$er_{\mu 2} = \frac{\lambda h(\mu h)^{l}}{l!}\left\{ \sum_{j=1}^{3} R_{2j}l(3 - j)^{l-1} - \sum_{j=1}^{2} [R_{1j}(3 - j)^{l} + \bar{R}_{1j}(r + 2 - j)^{l}] \right.$$

$$\left. + \sum_{j=2}^{3} L_{2j}(4 - j)^{l} \right\} \qquad l = 0, 1, \ldots \tag{162b}$$

The coefficient to $(\lambda h)^2$ in equation (60) is identically zero.

If er_μ is made to be of order h^5, then equations (162) provide nine equations for the ten unknown constants (four α, five β, and r) in equations (159). Hence, one can calculate the errors $er_{\mu 1}$ and $er_{\mu 2}$ as functions of the parameter r. The result is shown in figure 14, where we see that $er_{\mu 2}$ completely dominates the total error except in the region $r \sim 1.8$.

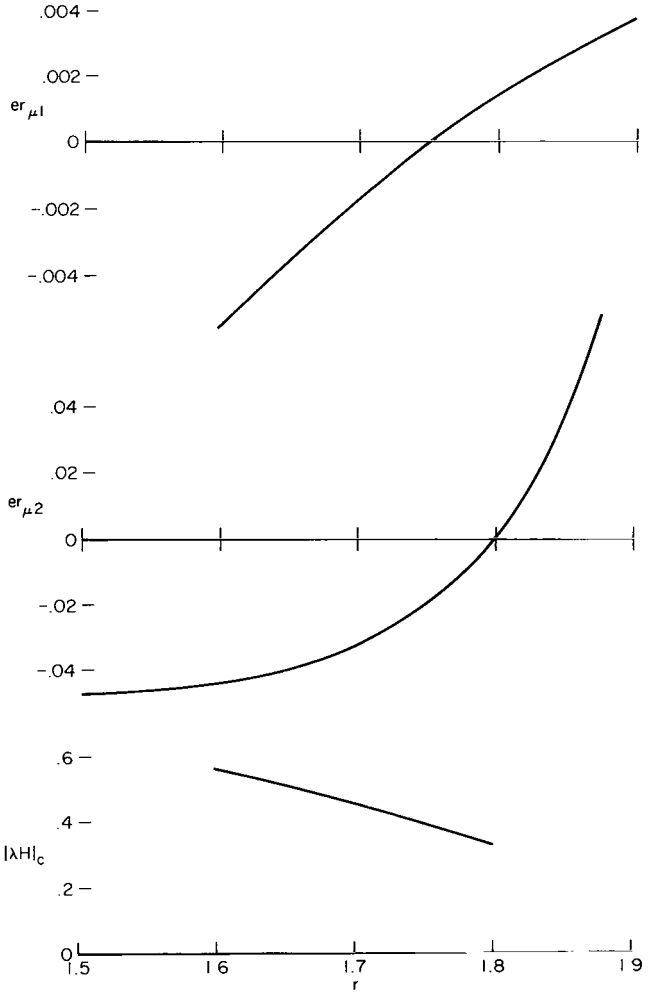


Figure 14.- Variation of terms controlling stability and accuracy of methods given by equations (159).

Notice, from equation (160), that the method has Adams-Moulton type stability for all r, since at h = 0 the characteristic equation reduces to $E(E-1)=0$. If r is set equal to 1.8, we find the following values for L and R

m \ j	L_{mj}			R_{mj}			$\bar{R}_{mj}$	
	2	3	4	1	2	3	1	2
1	720	0	0	705	15	0	375	-375
2	-1128	1848	0	1340	932	-64	0	0
3	1340	932	-64					

Divide by 720

$$er_\mu = \frac{1}{720}(\mu H)^5 , \qquad er_\lambda = \frac{1}{720}(\lambda H)^5$$

and the corresponding set of difference-differential equations

$$\left.\begin{aligned} u^{(1)}_{n+1.8} &= u_{n+1} + \frac{h}{75}\left(268u'_{n+1} + 22u'_n - 230u^{(1)'}_{n+0.8}\right) \\ u_{n+2} &= u_{n+1} + \frac{h}{48}\left(47u'_{n+1} + u'_n + 25u^{(1)}_{n+1.8} - 25u^{(1)}_{n+0.8}\right) \end{aligned}\right\} \tag{163}$$

The error terms for these equations are by far the smallest errors ($1/720 \approx 0.0014$) for any of the methods considered above. However, there is the usual sacrifice in stability. From a study of results such as those given in figure 15, one can calculate the curve for the induced stability boundary shown in figure 14. We see that equations (163) are limited by the stability boundary $|\lambda H|_c = 0.3$.

THE OPERATIONAL FORM

Definition and Discussion

A variety of systems of difference-differential equations have been analyzed as they applied to the solution of ordinary differential equations. In every case the method being studied was associated with an equation of the form

$$P\left(L_{mj}(\lambda h)^{m-1}E^{k+1-j}\right) = Ahe^{\mu hn}Q\left(R_{mj}(\lambda h)^{m-1}e^{\mu h(r_{i+1}-j)}\right) \tag{164}$$

where P and Q symbolize polynomials with terms such as those within the arguments. In every case this equation was the sole basis for determining the accuracy and stability of the method. In general, the maximum value of m is determined by the number of iterations. If a method uses M iterations,

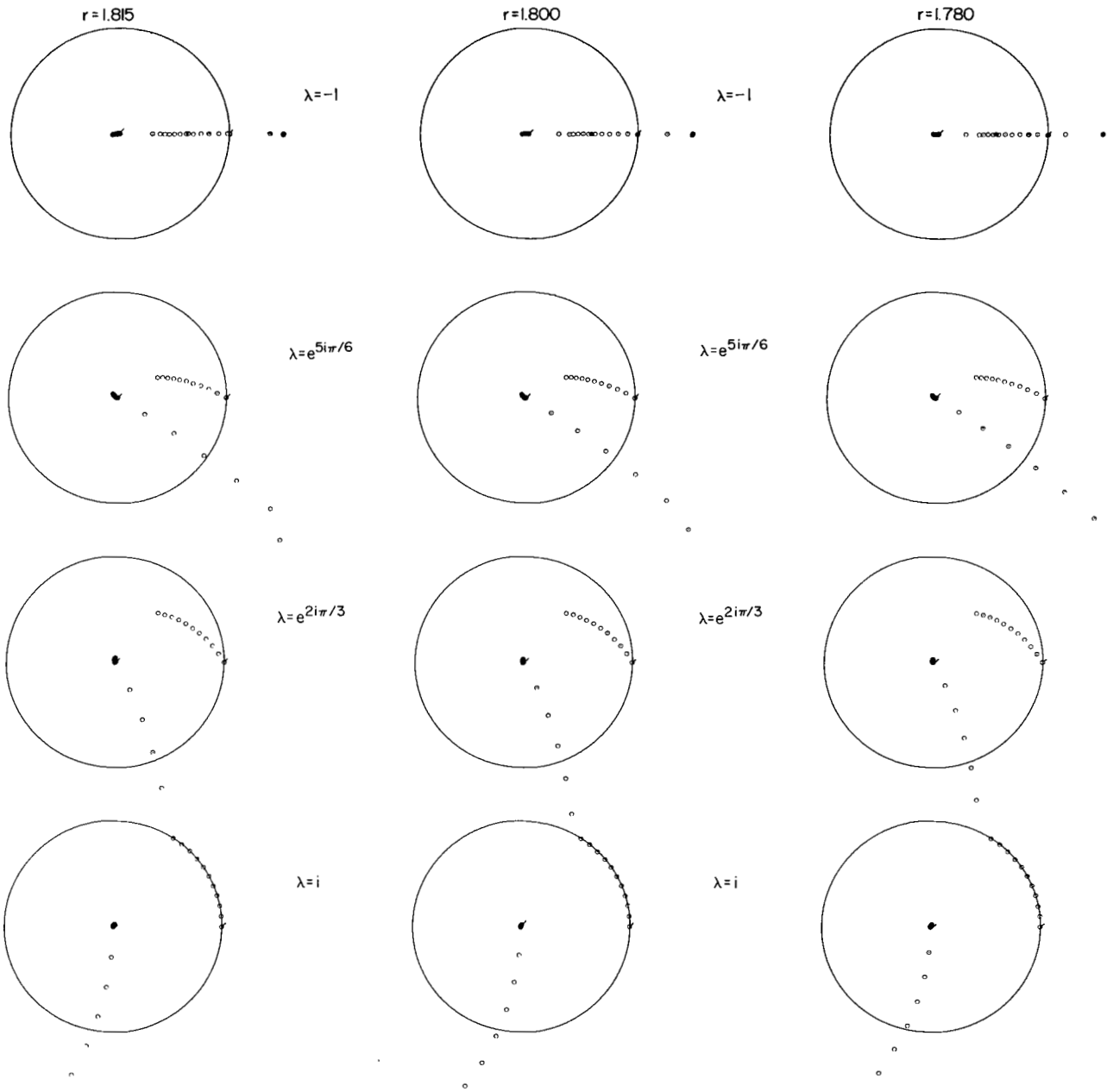


Figure 15.- Stability plots for equations (160).

$m_{max} = M + 1$. The maximum value of j is $k + 1$, where the interpretation of k varies according to the type of method. For incomplete, uncombined methods, such as Hamming's, Adams-Moulton, etc., k is the largest step number used in the predictor or the corrector. In complete, uncombined methods, k is twice the maximum step number. In combined methods, steps are not necessarily equispaced and k loses its connection with step number. The term r_i can be replaced with k in the uncombined methods. In the combined ones, however, r_i determines the location at which calculations of the function and/or its derivative are carried out and these locations are not necessarily spaced in integer multiples of h.

We refer to equation (164) as the operational form of a numerical method and to the coefficients R_{mj} and L_{mj} as the coefficients in the operational form. One can show:

1. All linear, difference-differential equations with constant coefficients have an operational form.

2. Any two such methods with the same operational form have, except for round-off considerations, the same accuracy and stability and give, therefore, except for round-off considerations, identical numerical results when applied to equations (11).

3. The coefficients $\alpha, \beta, \gamma, \ldots$ in the actual difference-differential equations affect the accuracy and stability of the method only as they affect the coefficients in the operational form and the correspondence is not unique.

Let us consider in more detail the above statement number 2. Strictly speaking, two methods that give identical numerical results when applied to linear equations cannot be classified as equivalent, since one can require more iterations than the other. As an example, the incomplete, four-step method.

$$\left.\begin{aligned} u^{(1)}_{n+4} &= 3.9u_{n+3} - 5.7u_{n+2} + 3.7u_{n+1} - 0.9u_n \\ u_{n+4} &= u_{n+3} + \frac{h}{12}\left(5u^{(1)'}_{n+4} + 8u'_{n+3} - u'_{n+2}\right) \end{aligned}\right\} \qquad (165)$$

when applied to equations (11) gives results[15] that are identical with those of the complete two-step method presented in (96). Notice that the number of multiplications and storage requirements of the two methods are the same. However, the use of equations (165) requires twice the number of iterations and for this reason, as a numerical method, is neither equivalent nor practical.

---

[15]To show this numerically, care must be taken with the initial conditions since neither method is self-starting.

Accuracy

The accuracy which a given operational form provides in the solution of a set of linear differential equations can be measured by the magnitude of the error terms er_μ, defined by equations (63) and (57); and by the error term er_λ where $er_\lambda = (er_\mu)_{\mu=\lambda}$. We are concerned only with polynomial approximation, so these terms are expanded in powers of h and the lowest power of h with a nonvanishing coefficient gives the order of the local polynomial fit.

Clearly, the parameters μ and λ depend on the differential equation, so the error of any method must be expressed as a function of μ, λ, and h. First, we notice that the error in the particular solution can always be expressed as a polynomial in λh. Collecting terms in this way, we next see that the coefficients of this polynomial are functions of μh and $e^{f(\mu h)}$. Finally, these coefficients are each expanded in powers of μh and made to vanish to the desired order. This leads to sets of conditions on the coefficients in the operational form, examples of which are given in equations (66), (67), and (131) through (134).

Stability

In this part we show that the Dahlquist stability theorem, derived for an implicit multistep equation, also holds for any multistep, predictor-corrector method contained in equation (164), provided $r_i = k$. That is, provided equispaced, predictor-corrector methods are not combined with Runge-Kutta techniques.

The argument starts by inspecting equations (3) and (4). As we have seen, the degree of the polynomial embedded in equation (3) depends upon how many of the terms $er_p(0)$, $er_p(1)$, . . ., $er_p(L)$ in equation (4) can be made identically zero. If, in fact, the β values are chosen so all terms through $er_p(L)$ are zero, $L + 1$ equations or constraints on β must be satisfied and the order of the embedded polynomial is L. On this basis, the maximum value of L is twice the step number. The stability, on the other hand, depends on the roots to the characteristic equation derived from equation (3), namely,

$$E^k - \sum_{j=1}^{k+1} (\beta_j + \lambda h \beta'_j) E^j = 0 \qquad (166)$$

The Dahlquist theorem is based on the hypothesis that, for a method to be stable, it is necessary that the roots to the characteristic equation lie on or inside the unit circle in a complex plane when $h = 0$. In our notation and analysis, these statements mean that L is to be made as large as possible in

$$\sum_{j=1}^{k+1} \left[l(k+1-j)^{l-1}\beta_j' + (k+1-j)^l \beta_j \right] = k^l , \qquad l = 0, 1, \ldots, L \tag{167a}$$

without having the absolute value of any of the roots to

$$E^k - \sum_{j=2}^{k+1} \beta_j E^{k+1-j} = 0 \tag{167b}$$

exceed unity. The Dahlquist theorem states that there are no combinations of β for which this is possible if $L > k + 2$ (k even) or $L > k + 1$ (k odd).

The extension of these results to uncombined methods constructed from an operational form with $r_i = k$ is quite simple. Consider, for example, the two-corrector method expressed by equations (111). The error, er_μ, for such a method is given by equations (117). For $h = 0$ we find a necessary condition for obtaining a method embedding a polynomial of order L is

$$\sum_{j=1}^{k+1} \left[l(k+1-j)^{l-1} R_{1j} + (k+1-j)^l L_{1j} \right] = k^l , \qquad l = 0, 1, \ldots, L \tag{168a}$$

and a necessary condition for stability is that the absolute value of the roots to

$$E^k - \sum_{j=2}^{k+1} L_{1j} E^{k+1-j} = 0 \tag{168b}$$

does not exceed unity.

Examining the discussion of equations (167), we see that the Dahlquist stability criterion is quite independent of the role of β in the difference-differential equations. Thus, although L and R of equations (168) can enter the difference-differential equations entirely differently, the Dahlquist theorem immediately tells us that only the first $k + 2$ (for even k, or $k+1$ for odd k) of equations (168a) can be satisfied if the absolute value of the roots to equation (168b) is to be no greater than one. Hence, the Dahlquist theorem still applies to combined predictor-corrector methods when any number of correctors, all of which may be different, are used in an uncombined, multistep, predictor-corrector method.

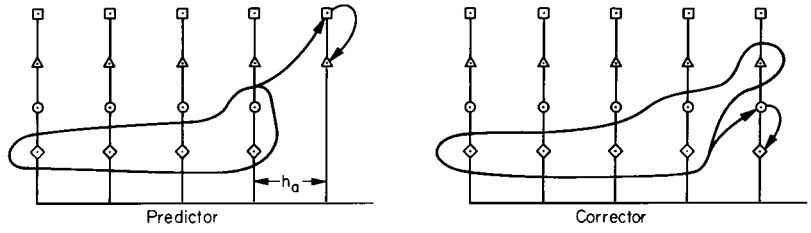
If combined predictor-corrector formulas are used (complete or incomplete), r_i is no longer equal to k. For example, equations (124) have, in place of (168a), the accuracy conditions

$$\sum_{j=2}^{k+1}\left[l(k+1-j)^{l-1}L_{1j}+(k+1-j)^{l}R_{1j}\right]+\sum_{i=1}^{3} lr_i^{l-1}R^{*}_{1i}=k^{l}\,,\qquad l=0,\,1,\,\ldots,\,L \tag{169}$$

although the stability equation remains identical to (168b). Using the same argument as above, we see that the Dahlquist theorem is no longer applicable to these cases.

The fact that the Dahlquist stability criterion must be modified if unequal steps are taken in advancing the integration of differential equations has been recorded by several authors, for example, references 13 through 16. But the variety of meanings given to the words "step number" in these and other references complicates a comparison of the stability capabilities of the various methods in a sense similar to that studied by Dahlquist. This problem, already discussed, is discussed here in light of its connection with the Dahlquist theorem.

Consider first the "conventional," four-step, method composed of an Adams-Bashforth predictor (line 5, table I(a)), followed by an Adams-Moulton corrector (line 4, table I(b)) as symbolized in sketch (g). (The symbols are defined in the previous section.)



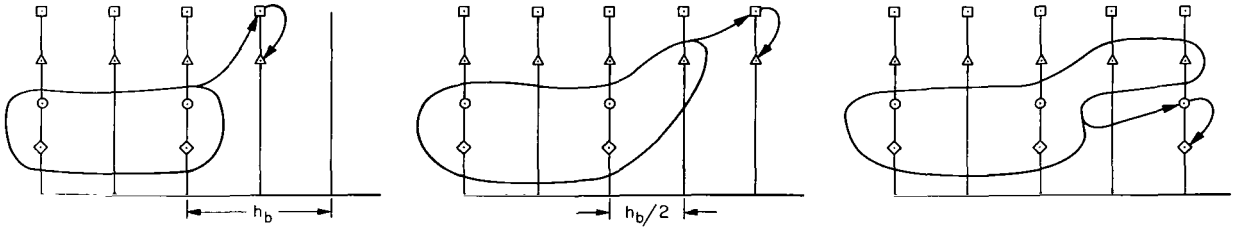
Sketch (g)

The data used to calculate the value of the predicted and corrected function are encircled; the remaining data are ignored. The step size is shown as h_a, the choice which coincides with the definition (123). It is equal to H, the distance advanced by two iterations. Six bits of data are weighted in the corrector and a stable method results with a local polynomial fit of order five. The error (see table II(j)) is around $-0.12\lambda\mu^5H^6$.

Consider next the combined, two-step method presented by Butcher in reference 15. In our notation it reads

$$
\left.
\begin{aligned}
u^{(1)}_{n+1.5} &= u_n + \frac{1}{8} h_b(9u'_{n+1} + 3u'_n) \\
u^{(2)}_{n+2} &= \frac{1}{5}(28u_{n+1} - 23u_n) + \frac{1}{15} h_b\left(32u^{(1)'}_{n+1.5} - 60u'_{n+1} - 26u'_n\right) \\
u_{n+2} &= \frac{1}{31}(32u_{n+1} - u_n) + \frac{1}{93} h_b\left(64u^{(1)'}_{n+1.5} + 15u^{(2)'}_{n+2} + 12u'_{n+1} - u'_n\right)
\end{aligned}
\right\} \quad (170)
$$

where h_b also coincides with the step size defined by (123). The process is symbolized in sketch (h). The error of the method is around $(\lambda h_b)^6/124$



Sketch (h)

or $\sim 0.06(\lambda H)^6$, since $h_b = 3H/2$. The increase in accuracy is compensated for, as usual, by a decrease in the stability boundary.

Although the method illustrated by sketch (h) is by definition a two-step method, it appears very similar to the four-step method shown in sketch (g). The real connection between the two is discovered by reducing each to its operational form. For simplicity, re-reference the indexing in equations (170) to a step size equal to h_a so that, for example, the first equation reads

$$
u^{(1)}_{n+3} = u_n + \frac{h_a}{16}(9u'_{n+2} + 3u'_n)
$$

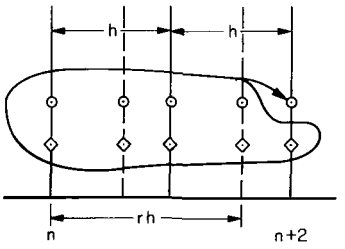
When this is done one can show that both methods are governed by equations (168) in which $k = 4$. Apply Dahlquist's theorem to the latter, and we find that a local polynomial fit of order 5 can certainly[16] be found which

[16]Actually the theorem states that a local polynomial fit of order 6 would be stable, but such methods would have a spurious root on the unit circle at $h = 0$, and are usually unstable for $h > 0$.

is stable for a variety of methods containing, over the length $2h_b$, the kind of data shown in sketch (h). Butcher showed that the particular choice of data encircled in sketch (h) are, in fact, stable for a polynomial of order 5 when weighted as in equations (170). If the number of steps, sized h_b, is increased to 3, a method similar to that given by equations (170) -- in that the values of the function and its derivative at the midpoints of steps behind the last are ignored -- is stable according to Butcher for a polynomial of order 7. However, Butcher also showed that when k_b, the number of h_b steps, is increased beyond three, this process is unstable for polynomials of order $2k_b + 1$. On the other hand, if we use the operational form, the Dahlquist theorem tells us that some choice of such equispaced data in the same interval can be used to derive a method that is stable for $2k_b + 1$ or even $2k_b + 2$ in the sense described in footnote 16.

Of course, the above example, when used in this perspective, is quite unfair as a true measure of the value of combined methods. This is due to the symmetrical choice of the sampling points. But it is quite useful in demonstrating that the words "step-number," as they are used in contemporary literature, are to be treated with great caution in comparing methods.

Perhaps the simplest way to present the basic issue involved is to consider sketch (i). The implicit method defined advances the solution a step h in one cycle of computation and, on this basis, it is a two-step method. The generalization of Dahlquist's stability theorem will provide an answer to the following question:



Sketch (i)

> Is there any value of r for which the data encircled in sketch (i) can be used to construct a stable method embedding a local polynomial of order 8?

Nine bits of data are now available so that a method having a polynomial fit of order 8 can easily be constructed. The Dahlquist theorem, when applied to the operational form, immediately tells us that the answer to the above question is negative when $r = 3/2$. If $r = 4/3$, however, the situation is not so simple. Then the data are spaced so as to be identical with operational forms of certain equispaced methods with six steps. The Dahlquist theorem tells us that there are some operational forms of such cases that are stable, but whether or not they can be obtained omitting data at two of the intervals is not known. Admittedly, questions such as this are approaching the academic, but their answer is a fundamental aspect to one area of numerical analysis.

Ames Research Center
National Aeronautics and Space Administration
Moffett Field, Calif., 94035, Jan. 24, 1967
129-04-03-02-00-21

REFERENCES

1. Dahlquist, Germund G.: Convergence and Stability in the Numerical Integration of Ordinary Differential Equations. Math. Scand., vol. 4, 1956, pp. 33-53.

2. Hamming, R. W.: Stable Predictor-Corrector Methods for Ordinary Differential Equations. J. Assoc. Comput. Mach., vol. 6, 1959, pp. 37-47.

3. Chase, P. E.: Stability Properties of Predictor-Corrector Methods for Ordinary Differential Equations. J. Assoc. Comput. Mach., vol. 9, 1962, pp. 457-468.

4. Hildebrand, Francis Begnaud: Introduction to Numerical Analysis. McGraw-Hill Book Co., Inc., 1956.

5. Henrici, Peter: Discrete Variable Methods in Ordinary Differential Equations. John Wiley and Sons, Inc., 1962.

6. Boole, George: Calculus of Finite Differences. Fourth ed. J. F. Moulton, ed., Chelsea Publishing Co., 1958.

7. Milne-Thomson, Louis Melville: The Calculus of Finite Differences. Macmillan, London, 1933.

8. Lewis, H. R., Jr.; and Stovall, E. J., Jr.: A FORTRAN Version of Nordsieck's Scheme for the Numerical Integration of Differential Equations. Rep. LA-3292, Los Alamos Scientific Lab., N. M., 22 June 1965.

9. Stetter, Hans J.: Stabilizing Predictors for Weakly Unstable Correctors. Math. Comp., vol. 19, 1965, pp. 84-89.

10. Milne, W. E.; and Reynolds, R. R.: Stability of a Numerical Solution of Differential Equations. Pt II, J. Assoc. Comput. Mach., vol. 7, pp. 46-56.

11. Crane, R. L.; and Klopfenstein, R. W.: A Predictor-Corrector Algorithm With an Increased Range of Absolute Stability. J. Assoc. Comput. Mach., vol. 12, 1965, pp. 227-241.

12. Dahlquist, G. G.: Stability Questions for Some Numerical Methods for Ordinary Differential Equations. Proc. Symp. Appl. Math., vol. XV, 1963, pp. 147-158. Am. Math. Soc., Providence, R.I.

13. Butcher, J. C.: On the Convergence of Numerical Solutions to Ordinary Differential Equations. Math. Comp., vol. 20, no. 73, 1966, pp. 1-10.

14. Gragg, William B.; and Stetter, Hans J.: Generalized Multistep Predictor-Corrector Methods. J. Assoc. Comput. Mach., vol. 11, 1964, pp. 188-209.

15. Butcher, J. C.: A Modified Multistep Method for the Numerical Integration of Ordinary Differential Equations. J. Assoc. Comput. Mach., vol. 12, 1965, pp. 124-135.

16. Gear, C. W.: Hybrid Methods for Initial Value Problems in Ordinary Differential Equations. J. SIAM Numer. Anal., Ser. B., vol. 2, no. 1, 1964, pp. 69-86.

17. Ralston, Anthony; and Wilf, Herbert S., eds.: Mathematical Methods for Digital Computers. John Wiley and Sons, Inc., 1960, pp. 110-120.

TABLE I.- COEFFICIENTS IN DIFFERENCE-DIFFERENTIAL EQUATIONS FOR CERTAIN PREDICTOR-CORRECTOR FORMULAS

(a) Predictor formulas

		α_1	α_1'	α_2	α_2'	α_3	α_3'	α_4	α_4'	α_5	α_5'	α_6	α_p
1	Euler			1	1								$1/2h^2u^{ii}$
2	Nystrom				2	1							$1/3h^3u^{iii}$
3	A-B* two-step			1	3/2		-1/2						$5/12h^3u^{iii}$
4	A-B* three-step			1	23/12		-16/12		5/12				$3/8h^4u^{iv}$
5	A-B* four-step			1	55/24		-59/24		37/24		-9/24		$251/720h^5u^{v}$
6	Milne-Hamming (No mod.)				8/3		-4/3		8/3	1			$14/45h^5u^{v}$
7	Hamming (mod.)			1	8/3		-4		4	1	-8/3	-1	$-11/72h^5u^{v}$
8	Crane -K			1.547652	2.002247	-1.867503	-2.031690	2.017204	1.818609	-0.697353	-0.714320		$0.4016h^5u^{v}$
9	(0.01, -0.01)			-59/17	127/34	76/17	59/34						$3/68h^3u^{iii}$
10	Stetter			-4	4	5	2						$1/6h^4u^{iv}$

*Adams-Bashforth method

(b) Corrector formulas

		β_1	β_1'	β_2	β_2'	β_3	β_3'	β_4	β_4'	β_5	β_5'	β_6	β_p
1	Modified Euler		1/2	1	1/2								$-1/12h^3u^{iii}$
2	A-M* two-step		5/12	1	8/12		-1/12						$-1/24h^4u^{iv}$
3	A-M* three-step		9/24	1	19/24		-5/24		1/24				$-19/720h^5u^{v}$
4	A-M* four-step		251/720	1	646/720		-264/720		106/720		-19/720		$-3/160h^6u^{vi}$
5	Milne		1/3		4/3	1	1/3						$-1/90h^5u^{v}$
6	Hamming (No mod.)		3/8	9/8	6/8		-3/8	-1/8					$-1/40h^5u^{v}$
7	Hamming (mod.)		42/121	126/121	108/121	0	-54/121	-14/121	24/121	9/121			$-21/1210h^6u^{vi}$
8	(0.01, -0.01)		34/93	12/31	100/93	19/31	16/93						$-1/62h^4u^{iv}$
9			3/11	-27/11	27/11	27/11	27/11	1	3/11				$3/1540h^7u^{vii}$

*Adams-Moulton method

TABLE II.- COEFFICIENTS IN THE OPERATIONAL FORM OF A NUMBER OF METHODS

(a) Predictor, row 6 of table I(a) (Milne-Hamming (no mod.)); corrector, row 5 of table I(b) (Milne)

m \ j	L_{mj}					R_{mj}					
	2	3	4	5	6	1	2	3	4	5	6
1	0	9	0	0		3	12	3	0	0	
2	12	3	0	3		0	8	-4	8	0	
3	8	-4	8	0							

Divide by 9

$$er_{\mu} = (0.0056 - 0.052\lambda H)(\mu H)^5$$

$$er_{\lambda} = 0.0056(\lambda H)^5$$

(b) Predictor, row 6 of table I(a) (Milne-Hamming (no mod.)); corrector, row 6 of table I(b) (Hamming (no mod.))

m \ j	L_{mj}					R_{mj}					
	2	3	4	5	6	1	2	3	4	5	6
1	9	0	-1	0		3	6	-3	0	0	
2	6	-3	0	3		0	8	-4	8	0	
3	8	-4	8	0							

Divide by 8

$$er_{\mu} = (0.033 - 0.155\lambda H)(\mu H)^5$$

$$er_{\lambda} = 0.033(\lambda H)^5$$

(c) Predictor, row 7 of table I(a) (Hamming (mod.)); corrector, row 7 of table I(b) (Hamming (mod.))

m \ j	L_{mj}					R_{mj}					
	2	3	4	5	6	1	2	3	4	5	6
1	126	0	-14	9	0	42	108	-54	24	0	0
2	150	-54	24	42	-42	0	112	-168	168	-112	0
3	112	-168	168	-112	0						

Divide by 121

$$er_{\mu} = (0.0175 - 0.109\lambda H)(\mu H)^6$$

$$er_{\lambda} = 0.0175(\lambda H)^6$$

TABLE II.- COEFFICIENTS IN THE OPERATIONAL FORM OF A NUMBER OF METHODS - Continued

(d) Predictor, row 3 of table I(a) (A-B two step); corrector, row 2 of table I(b) (A-M two step)

m \ j	L_{mj}					R_{mj}					
	2	3	4	5	6	1	2	3	4	5	6
1	24	0				10	16	-2			
2	26	-2				0	15	-5			
3	15	-5									

Divide by 24

$$er_{\mu} = (0.042\mu - 0.174\lambda)\mu^3 H^4$$

$$er_{\lambda} = -0.132(\lambda H)^4$$

(e) Predictor, row 2 of table I(a) (Nystrom); corrector, row 2 of table I(b) (A-M two step)

m \ j	L_{mj}					R_{mj}					
	2	3	4	5	6	1	2	3	4	5	6
1	12	0				5	8	-1			
2	8	4				0	10	0			
3	10	0									

Divide by 12

$$er_{\mu} = (0.042\mu - 0.138\lambda)\mu^3 H^4$$

$$er_{\lambda} = -0.097(\lambda H)^4$$

(f) Predictor, row 9 of table I(a); corrector, row 8 of table I(b)

m \ j	L_{mj}					R_{mj}					
	2	3	4	5	6	1	2	3	4	5	6
1	1224	1938				1156	3400	544			
2	-612	5712				0	4318	2006			
3	4318	2006									

Divide by 3162

$$er_{\mu} = 0.01(\mu - \lambda)\mu^3 H^4$$

$$er_{\lambda} = -0.027(\lambda H)^5$$

TABLE II.- COEFFICIENTS IN THE OPERATIONAL FORM OF A NUMBER OF METHODS - Continued

(g) Predictor, row 4 of table I(a) (A-B three step); corrector, row 2 of table I(b) (A-M two step)

m \ j	L_{mj}					R_{mj}					
	2	3	4	5	6	1	2	3	4	5	6
1	144	0	0			60	96	-12	0		
2	156	-12	0			0	115	-80	25		
3	115	-80	25								

Divide by 144

$$er_{\mu} = (0.042 - 0.156\lambda H)(\mu H)^4$$

$$er_{\lambda} = 0.042(\lambda H)^4$$

(h) Predictor, row 4 of table I(a) (A-B three step); corrector, row 3 of table I(b) (A-M three step)

m \ j	L_{mj}					R_{mj}					
	2	3	4	5	6	1	2	3	4	5	6
1	288	0	0			108	228	-60	12		
2	336	-60	12			0	207	-144	45		
3	207	-144	45								

Divide by 288

$$er_{\mu} = (0.026\mu - 0.141\lambda)\mu^4 H^5$$

$$er_{\lambda} = -0.114(\lambda H)^5$$

(i) Predictor, row 5 of table I(a) (A-B four step); corrector, row 3 of table I(b) (A-M three step)

m \ j	L_{mj}					R_{mj}					
	2	3	4	5	6	1	2	3	4	5	6
1	192	0	0	0		72	152	-40	8	0	
2	224	-40	8	0		0	165	-177	111	-27	
3	165	-177	111	-27							

Divide by 192

$$er_{\mu} = (0.026 - 0.131\lambda H)(\mu H)^5$$

$$er_{\lambda} = 0.026(\lambda H)^5$$

TABLE II.- COEFFICIENTS IN THE OPERATIONAL FORM OF A NUMBER OF METHODS - Concluded

(j) Predictor, row 5 of table I(a) (A-B four step); corrector, row 4 of table I(b) (A-M four step)

m \ j	L_{mj}					R_{mj}					
	2	3	4	5	6	1	2	3	4	5	6
1	17280	0	0	0		6024	15504	-6336	2544	-456	
2	21528	-6336	2544	-456		0	13805	-14809	9287	-2259	
3	13805	-14809	9287	-2259							

Divide by 17280

$$er_{\mu} = (0.019\mu - 0.122\lambda)\mu^{5}H^{6}$$

$$er_{\lambda} = -0.103(\lambda H)^{6}$$

(k) Predictor, row 10 of table I(a) (Stetter); corrector, row 5 of table I(b) (Milne)

m \ j	L_{mj}					R_{mj}					
	2	3	4	5	6	1	2	3	4	5	6
1	0	3				1	4	1			
2	0	6				0	4	2			
3	4	2									

Divide by 3

$$er_{\mu} = (0.0056\mu - 0.028\lambda)\mu^{4}H^{5}$$

$$er_{\lambda} = -0.022(\lambda H)^{5}$$

TABLE III.- COEFFICIENTS OF L AND R FOR USE IN THE CALCULATION OF er_μ FOR ONE- THROUGH FIVE-STEP METHOD

(a) Equation (66)

	k = 1					R_{11}	R_{12}					L_{12}					
	2				R_{11}	R_{12}	R_{13}				L_{12}	L_{13}					
	3			R_{11}	R_{12}	R_{13}	R_{14}			L_{12}	L_{13}	L_{14}					
	4		R_{11}	R_{12}	R_{13}	R_{14}	R_{15}		L_{12}	L_{13}	L_{14}	L_{15}	k^l				
l	5	R_{11}	R_{12}	R_{13}	R_{14}	R_{15}	R_{16}	L_{12}	L_{13}	L_{14}	L_{15}	L_{16}	k = 1	2	3	4	5
0		0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
1		1	1	1	1	1	1	4	3	2	1	0	1	2	3	4	5
2		10	8	6	4	2	0	16	9	4	1	0	1	4	9	16	25
3		75	48	27	12	3	0	64	27	8	1	0	1	8	27	64	125
4		500	256	108	32	4	0	256	81	16	1	0	1	16	81	256	625
5		3125	1280	405	80	5	0	1024	243	32	1	0	1	32	243	1024	3125
6		18750	6144	1458	192	6	0	4096	729	64	1	0	1	64	729	4096	15625

TABLE III.- COEFFICIENTS OF L AND R FOR USE IN THE CALCULATION OF er_μ FOR ONE- THROUGH FIVE-STEP METHOD - Concluded

(b) Equation (67)

	k = 1					$-R_{11}$	$-R_{12}$					R_{22}					L_{22}
	2				$-R_{11}$	$-R_{12}$	$-R_{13}$				R_{22}	R_{23}				L_{22}	L_{23}
	3			$-R_{11}$	$-R_{12}$	$-R_{13}$	$-R_{14}$			R_{22}	R_{23}	R_{24}			L_{22}	L_{23}	L_{24}
	4		$-R_{11}$	$-R_{12}$	$-R_{13}$	$-R_{14}$	$-R_{15}$		R_{22}	R_{23}	R_{24}	R_{25}		L_{22}	L_{23}	L_{24}	L_{25}
l	5	$-R_{11}$	$-R_{12}$	$-R_{13}$	$-R_{14}$	$-R_{15}$	$-R_{16}$	R_{22}	R_{23}	R_{24}	R_{25}	R_{26}	L_{22}	L_{23}	L_{24}	L_{25}	L_{26}
0		1	1	1	1	1	1	0	0	0	0	0	1	1	1	1	1
1		5	4	3	2	1	0	1	1	1	1	1	4	3	2	1	0
2		25	16	9	4	1	0	8	6	4	2	0	16	9	4	1	0
3		125	64	27	8	1	0	48	27	12	3	0	64	27	8	1	0
4		625	256	81	16	1	0	256	108	32	4	0	256	81	16	1	0
5		3125	1024	243	32	1	0	1280	405	80	5	0	1024	243	32	1	0
6		15625	4096	729	64	1	0	6144	1458	192	6	0	4096	729	64	1	0

TABLE IV.- COEFFICIENTS OF L FOR USE IN THE CALCULATION OF er_λ ONE- THROUGH FIVE-STEP METHOD
(see eq. (72))

	k = 1												L_{12}	L_{22}	L_{32}					
	2									L_{12}	L_{22}	L_{32}	L_{13}	L_{23}	L_{33}					
	3						L_{12}	L_{22}	L_{32}	L_{13}	L_{23}	L_{33}	L_{14}	L_{24}	L_{34}					
	4				L_{12}	L_{22}	L_{32}	L_{13}	L_{23}	L_{33}	L_{14}	L_{24}	L_{34}	L_{15}	L_{25}	L_{35}	k^l				
l	5	L_{12}	L_{22}	L_{32}	L_{13}	L_{23}	L_{33}	L_{14}	L_{24}	L_{34}	L_{15}	L_{25}	L_{35}	L_{16}	L_{26}	L_{36}	k = 1	2	3	4	5
0		1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	1	1	1	1
1		4	1	0	3	1	0	2	1	0	1	1	0	0	1	0	1	2	3	4	5
2		16	8	2	9	6	2	4	4	2	1	2	2	0	0	2	1	4	9	16	25
3		64	48	24	27	27	18	8	12	12	1	3	6	0	0	0	1	8	27	64	125
4		256	256	192	81	108	108	16	32	48	1	4	12	0	0	0	1	16	81	256	625
5		1024	1280	1280	243	405	540	32	80	160	1	5	20	0	0	0	1	32	243	1024	3125
6		4096	6144	7680	729	1458	2430	64	192	480	1	6	30	0	0	0	1	64	729	4096	15625

"The aeronautical and space activities of the United States shall be conducted so as to contribute . . . to the expansion of human knowledge of phenomena in the atmosphere and space. The Administration shall provide for the widest practicable and appropriate dissemination of information concerning its activities and the results thereof."

—NATIONAL AERONAUTICS AND SPACE ACT OF 1958

NASA SCIENTIFIC AND TECHNICAL PUBLICATIONS

TECHNICAL REPORTS: Scientific and technical information considered important, complete, and a lasting contribution to existing knowledge.

TECHNICAL NOTES: Information less broad in scope but nevertheless of importance as a contribution to existing knowledge.

TECHNICAL MEMORANDUMS: Information receiving limited distribution because of preliminary data, security classification, or other reasons.

CONTRACTOR REPORTS: Scientific and technical information generated under a NASA contract or grant and considered an important contribution to existing knowledge.

TECHNICAL TRANSLATIONS: Information published in a foreign language considered to merit NASA distribution in English.

SPECIAL PUBLICATIONS: Information derived from or of value to NASA activities. Publications include conference proceedings, monographs, data compilations, handbooks, sourcebooks, and special bibliographies.

TECHNOLOGY UTILIZATION PUBLICATIONS: Information on technology used by NASA that may be of particular interest in commercial and other non-aerospace applications. Publications include Tech Briefs, Technology Utilization Reports and Notes, and Technology Surveys.

Details on the availability of these publications may be obtained from:

SCIENTIFIC AND TECHNICAL INFORMATION DIVISION

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Washington, D.C. 20546