THE PROPAGATION OF ERROR IN NUMERICAL INTEGRATIONS

MARK LOTKIN

1. Introduction. The numerical integration of differential equations is generally performed by replacing the differential equations by approximate difference equations whose solutions are expected to approach those of the associated differential equations as the step size approaches zero. The replacement of differential by difference equations may clearly be carried out in a variety of ways; the actual choice will depend on particular circumstances, accuracy requirements, computational facilities, etc.

It is now a well known fact that whenever the order of the difference equations exceeds that of the original differential equations there are introduced certain numerical solutions that are extraneous to the original differential equations. The behavior of these extraneous solutions in general determines the usefulness of the integration method. For such a method to be effective it must be “stable” in the sense that the extraneous solutions always remain of negligible size as compared with the actual solutions.

Thus it is of interest to distinguish first between stable and unstable methods. In addition, it is also of interest to determine, in either case, the growth of error in the large, since the knowledge of this quantity permits an estimation of the accuracy obtained.

This paper, then, deals with a number of standard methods of integration, and investigates their stability and propagation of error. Round-off is considered to a certain extent, but not completely; see footnote 1. While some of the results have been obtained previously, mainly by L. H. Thomas [1] and H. Rutishauser [2], others do not seem to be as well known.

The propagation of error was already treated previously by Rademacher [3] and others. While the method of adjoint differential equations employed there seems to be capable of general application, it was used, in [3] especially, for Heun’s method only.

Finally there are carried out a few illustrative examples; they show that the theoretical expressions obtained frequently lead to good estimates.

2. Solutions of linear difference equations. Let us assume that the

Presented to the Society, September 1, 1953; received by the editors November 11, 1953 and, in revised form, May 3, 1954.

869
integration of the $n$th order differential equation has proceeded from a starting point $x_0$ to a point $x_k$, and that the original differential equation has been replaced by a difference equation of order $s$:

$$\alpha_{ks}v_{k+s} + \alpha_{k,s-1}v_{k+s-1} + \cdots + \alpha_{k,0}v_k + \alpha_k = 0,$$

where the coefficients $\alpha_{kj}$, $\alpha_k$ are known, with $\alpha_{ks} \neq 0$ for each $k$, and initial values $v_0, v_1, \cdots, v_{s-1}$ have been supplied. For our purposes it is now convenient to use matrix notation. Introducing, then, the column matrices

$$u_k = \begin{bmatrix} v_{k+s-1} \\ v_{k+s-2} \\ \vdots \\ v_k \end{bmatrix}, \quad b_k = \begin{bmatrix} \alpha_k \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad u_0 = \begin{bmatrix} v_{s-1} \\ \vdots \\ v_0 \end{bmatrix}$$

and the square matrices of order $s$:

$$J_k = \begin{bmatrix} -\alpha_{ks} & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}, \quad A_k = \begin{bmatrix} \alpha_{k,s-1} & \alpha_{k,s-2} & \cdots & \alpha_{k,0} \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

we may express (2.1) as a system of difference equations of the first order:

$$J_k E u_k = A_k u_k + b_k,$$

with $E$ denoting the displacement operator

$$Ev_i = v_{i+1}.$$

Since $J_k$ is nonsingular, (2.2) may be written as

$$u_{k+1} = J_k^{-1} A_k u_k + J_k^{-1} b_k = C_k u_k + d_k,$$

with

$$C_k = J_k^{-1} A_k, \quad d_k = J_k^{-1} b_k.$$

The general solution of (2.3) is

$$u_k = P_{k-1} \left( u_0 + \sum_{t=0}^{k-1} P_t^{-1} d_t \right),$$

where
1954] PROPAGATION OF ERROR IN NUMERICAL INTEGRATIONS 871

\[ P_k = C_k P_{k-1}, \quad P_0 = C_0, \]

or

\[ P_{k-1} = \prod_{t=1}^{k} C_{k-t}. \]

In particular, if \( C_k \) is actually independent of \( k \), then \( P_{k-1} = C_k \) and

\[ u_k = C^k \left( u_0 + \sum_{t=0}^{k-1} C^{-t-1} d_t \right). \tag{2.5} \]

The use of Sylvester's theorem [4] now permits us to express the solution \( u_k \) in slightly different form, as follows: let \( G(\lambda) = \lambda I - C \), \( G_a(\lambda) \) denote the adjoint of \( G(\lambda) \), \( \delta(\lambda) \) the determinant of \( G(\lambda) \), and \( \delta'(\lambda) = d\delta/d\lambda \). If the characteristic roots \( \lambda_m, m = 1, 2, \ldots, s \), of \( C \) are distinct, then

\[ C^k = \sum_{m=1}^{s} \lambda_m^k H_m \tag{2.6} \]

with

\[ H_m = H(\lambda_m) = G_a(\lambda_m)/\delta'(\lambda_m). \]

Consequently, from (2.5),

\[ u_k = \sum_{m=1}^{s} \lambda_m^k H_m \left( u_0 + \sum_{t=0}^{k-1} \lambda_m^{-t-1} d_t \right), \]

or, if \( u_0 = 0 \),

\[ u_k = \sum_{m=1}^{s} H_m \sum_{t=0}^{k-1} \lambda_m^{-t-1} d_t. \tag{2.7} \]

If the distinct roots \( \lambda_i, i = 1, 2, \ldots, q \), of \( C \) have multiplicities \( \mu_i \), then (2.6) must be replaced by

\[ C^k = \sum_{i} \frac{1}{(\mu_i - 1)!} \left[ \frac{d^{\mu_i-1}}{d\lambda^{\mu_i-1}} \left( \frac{\delta(\lambda)}{\delta(\lambda)} \right) \right]_{\lambda = \lambda_i} \]

\[ \delta_{\mu_i}(\lambda) = (\lambda - \lambda_i)^{-\mu_i}\delta(\lambda). \tag{2.8} \]

In particular, if the only multiple root of \( C \) has the value zero, then clearly (2.8) reverts to the form (2.6) with the summation to be extended over all the nonvanishing roots. This observation will be put to use in the subsequent discussion.

3. The variational difference equations. The foregoing treatment
of the difference equation will now be applied to the numerical integration of the \( n \)th order differential equation

\[
y^{(n)} = f_n(x, y, y', \ldots, y^{(n-1)})
\]

subject to suitable boundary conditions. The exact solution \( y(x) \) of (3.1), assumed to exist uniquely, may be expressed in the form [2]

\[
y(x_{k+1}) = \sum_{j=0}^{r} a_{0j}y(x_{k-j}) + h \sum_{j=1}^{r} a_{1j}y'(x_{k-j}) + \cdots
\]

\[+ h^N \sum_{j=1}^{r} a_{nj}y^{(N)}(x_{k-j}) + T_k,
\]

(3.2)

\[
y^{(i)}(x_{k+1}) = \sum_{j=0}^{r} a_{ij}^{(i)} y^{(i-j)}(x_{k-j}) + h \sum_{j=1}^{r} a_{ij+1}^{(i)} y^{(i+j-1)}(x_{k-j}) + \cdots
\]

\[+ h^{N-i} \sum_{j=1}^{r} a_{nj}^{(N)} y^{(N)}(x_{k-j}) + T_{ki},
\]

for \( i = 1, 2, \ldots, n-1 \). Here \( h = x_{k+1} - x_k \) is the step used in the integration, the \( a_{ij}^{(i)} \) are constants that will in general depend on \( r \), where \( r \) itself indicates a certain range of points \( x_{k-j} \) to the left of \( x_k \); to each numerical method of integration there is associated a fixed value of \( r \). The functions \( T_k, T_{ki} \) are truncation errors of orders \( h^{N+1}, h^{N-i+1} \), respectively, with \( N \geq n \) denoting a positive integer. In case \( N \) exceeds the order \( n \) of the equation (3.1) the derivatives of orders \( n+1, n+2, \ldots, N \) occurring in (3.2) may be obtained by \( N-n \) successive differentiations of (3.1):

\[
y^{(i)}(x) = f_i(x, y(x), y'(x), \ldots, y^{(i-1)}(x)),
\]

\( i = n+1, \ldots, N \). The coefficients \( a_{ij}^{(i)} \) are not arbitrary but normally depend on certain conditions (4.5) derived below.

In solving the integration problem numerically (3.1) is actually replaced by

\[
* y^{(n)} = * f_n(x, * y, * y', \ldots, * y^{(n-1)}),
\]

where the asterisks indicate rounded values, and the method of integration actually employed may be of the form

\[
* y_{k+1}^{(i)} = \sum_{j=0}^{r} * a_{ij}^{(i)} \ominus * y_{k-j}^{(i)} \ominus \cdots \ominus h^{N-i} \sum_{j=1}^{r} * a_{nj}^{(N)} y_{k-j}^{(N)}
\]

for \( i = 0, 1, 2, \ldots, n-1 \), and

\[
* y_{k+1}^{(i)} = f_i(x_{k+1}, * y_{k+1}, \ldots, * y_{k+1}^{(i-1)}), \quad n \leq i \leq N.
\]
In (3.2a) the circled symbols indicate pseudo-addition and pseudo-multiplication, i.e. certain digital operations by which the corresponding arithmetical operations must be replaced whenever numerical calculations (which of necessity involve rounding) are carried out.\textsuperscript{1}

In practice the numerical solution of the sets (3.2a) and (3.3a) is obtained iteratively in the following manner: Extrapolation, or some other means, permits the determination of a first set of values for \( *y_{k+1}^{(0)} \), \( n \leq i \leq N \). Then (3.2a), with \( i = n - 1, n - 2, \ldots, 0 \), lead to a first set of values for \( *y_{k+1}^{(0)} \), \( 0 \leq i \leq n - 1 \). Next an improved set \( *y_{k+1}^{(0)} \), \( n \leq i \leq N \), is computed by means of (3.3a), etc., this cycle being repeated until duplication occurs.

Our main interest is now the determination of the errors

\[
\eta_i^{(i)} = *y_i^{(i)} - y_i(x_i),
\]

and of the associated property of "numerical stability" in the sense that all the \( \eta_i^{(i)} \) remain small throughout the entire region of integration. By (3.2a) and (3.2),

\[
\eta_i^{(i)} = \sum_{j=0}^{r} \left[ *a_{ij}^{(i)} \odot *y_{k-j}^{(i)} - a_{ij} y_i(x_{k-j}) + \cdots \right] - T_{ki}.
\]

However, \( *a \odot *y = a * y + (*a \odot *y - a * y) + (a - a) * y \), and \( *a \odot *y = a * y = \rho, a - a = \sigma \), with \( |\rho| \leq \mu, |\sigma| \leq \mu, \mu = 2^{-1} \beta^{-1/3} \), denoting the basic rounding error of a computation carried out to \( \gamma \) places in a number system of base \( \beta \).

Consequently

\[
\eta_i^{(i)} = \sum_{j=0}^{r} a_{ij}^{(i)} \eta_{k-j} + h \sum_{j=1}^{r} a_{i+1,j}^{(i)} \eta_{k-j} + \cdots + h^{N-1} \sum_{j=1}^{r} a_{N,j}^{(i)} \eta_{k-j} - T_{ki} + \tau_{ki},
\]

(3.4)

\[
\tau_{ki} = \sum_i (\rho_{ijk} + \sigma_{ijk} y_{k-j}).
\]

Further, by (3.3) and (3.3a),

\textsuperscript{1} Note that there is no provision in formula (3.2a) for rounding the product involving \( h^{N-i} \). Thus the formula and later special cases of it in §4 are based on the assumption that the independent variable \( x \) and the step size \( h \) may be chosen exactly, and that multiplication by powers of \( h \) does not necessitate rounding. These assumptions could be dropped at the expense of including additional rounding items. As written, however, the conclusions of the paper may not be precisely applicable to numerical integrations in which the term in question is rounded.
\[ \eta_{k+1}^{(i)} = f_i(x_{k+1}, y_{k+1}, \ldots) - f_i(x_{k+1}, y(x_{k+1}), \ldots) \]
for \( i = n, n + 1, \ldots, N. \)

But

\[ f_i(x, y, \ldots) = f_i(x, y, \ldots) + \phi_i, \]

with \( \phi_i \) denoting certain quantities that depend on the procedure employed in calculating \( f_i \) from its arguments \( x, y, \ldots. \) Thus, correct to terms of the first order in \( \eta, \)

\[ \eta_{k+1}^{(i)} = \sum_{j=0}^{i-1} (\partial f_i / \partial y^{(j)}) \eta_{k+1}^{(j)} + \phi_i, \quad n \leq i \leq N, \]

the partial derivatives to be evaluated at \( x_{k+1}, y_{k+1}, \ldots, \)

The system (3.4) and (3.5) of difference equations is transformed into the form (2.2) by the introduction of the column matrix \( U_k, \)

\[ U_k^T = [\eta_k, \eta_{k-1}, \ldots, \eta_k^{(N)}], \]

the superscript \( T \) denoting transposition. Then our system becomes

\[ J_k E U_k = A U_k + b_k, \]

where \( J_k, A \) are square matrices of order \( s = (r + 1)(N + 1), \) and \( b_k \) is a column matrix of the same order. These matrices are composed as follows:

(i) The elements \( J_{ij} \) of \( J_k \) are square matrices of order \( r + 1, \)

\[ J_{ii} = I \] for \( i = 0, 1, \ldots, N, \)

\[ J_{ij} = 0 \] for \( 0 \leq i \leq n - 1, j < i, \) and for \( n \leq i \leq N, j > i, \)

\[ J_{ij} = \begin{bmatrix} -h^{j-i} & a_{j-i-1}^{(i)} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ 0 & \ddots & \ddots & \ddots & \ddots \\ 0 & \cdots & 0 & \ddots & \ddots \\ \end{bmatrix} \]

for \( 0 \leq i \leq n - 1, 1 \leq j \leq N, i < j, \)

\[ J_{ij} = \begin{bmatrix} -\partial f_i / \partial y^{(i)} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots \\ \vdots & \ddots & \ddots & \ddots \\ 0 & \ddots & \ddots & \ddots \\ 0 & \cdots & 0 & \ddots \\ \end{bmatrix} \]

for \( n \leq i \leq N, 0 \leq j \leq N - 1, i > j; \)

(ii) the elements \( A_{ij} \) of \( A \) are square matrices of order \( r + 1, \)
A_{ii} = \begin{bmatrix}
    a_{i0}^{(i)} & a_{i1}^{(i)} & \cdots & a_{ir}^{(i)} \\
    1 & 0 & 0 \\
    0 & \cdot & \cdot & \cdot \\
    \cdot & \cdot & \cdot & \cdot \\
    0 & 0 & 1 & 0 \\
\end{bmatrix}
\quad \text{for } 0 \leq i \leq n - 1,

A_{ii} = \begin{bmatrix}
    0 & 0 & \cdots & 0 \\
    1 & 0 & \cdots & 0 \\
    0 & \cdot & \cdot & \cdot \\
    \cdot & \cdot & \cdot & \cdot \\
    0 & 0 & 1 & 0 \\
\end{bmatrix}
\quad \text{for } n \leq i \leq N,

A_{ij} = h^{i-j} \begin{bmatrix}
    a_{j0}^{(i)} & a_{j1}^{(i)} & \cdots & a_{jr}^{(i)} \\
    0 & 0 & \cdots & 0 \\
    \cdot & \cdot & \cdots & \cdot \\
    \cdot & \cdot & \cdots & \cdot \\
    0 & \cdots & 0 \\
\end{bmatrix}
\quad \text{for } 0 \leq i \leq n - 1, 1 \leq j \leq N, i < j,

A_{ii} = 0 \quad \text{for } n \leq i \leq N, j \neq i, \text{ and also for } 0 < i \leq n - 1, j < i;

(iii) the column matrix \( b_k \) has the transpose

\[ b_k^T = \left[ -T_k + \tau, 0, \cdots, 0, \cdots, -T_k, 0, 0; \phi_0, 0, \cdots, 0; \cdots, \phi_N, 0, \cdots, 0 \right]. \]

Clearly the determinant \( D(J) \) of \( J \) is of the form

\[ D(J) = 1 + C_1 h + C_2 h^2 + \cdots, \]

i.e. \( D(J) \neq 0 \) for sufficiently small \( h \).

Let us now assume that the partial derivatives \( (\partial f_i/\partial y^{(j)}) \) have the property that there exist points \( \alpha_0, \beta_0^{(j)} \) in a suitably defined space \( |x-x_0| \leq \alpha, \ |y^{(j)}-y_0^{(j)}| \leq \beta^{(j)}, \ j = 0, 1, \cdots, N, \) such that \( \partial f_i/\partial y^{(i)} \) is approximately equal to a constant \( f_{ij} (\alpha_0, \beta_0, \beta_0', \cdots, \beta_0^{(i-1)}) \). In such a case let \( J_0 \) denote the matrix \( J \) whose elements are evaluated at \( \alpha_0, \beta_0^{(j)} \). If the characteristic roots \( \lambda \) of \( C_0 = J_0^{-1}A \) are distinct, then by (2.7)

\[ U_k = \sum_{m} H_m \sum_{t=0}^{k-1} \lambda_{m}^{k-t-1} J_0^{-1} b_t. \]

Now \( A \) has at least \( N-n+1 \) rows of zeros. There are thus
(N - n + 1) (r + 1) artificial characteristic roots of \( C_0 \) of value zero. There must be, further, the \( n \) roots associated with the \( n \) independent solutions of the variational equations; these roots are of the form
\[
\lambda_m = 1 + \gamma_m h + \cdots, \quad m = 1, 2, \cdots, n.
\]

Thus there remain
\[
(r + 1)(N + 1) - (r + 1)(N - n + 1) - n = nr
\]
additional roots. These are "extraneous," introduced by the method of integration. "Extraneous" solutions of the integration problem are, consequently, solutions belonging to extraneous roots of \( C_0 \).

Now in the solution vector \( U_k \) the only components of interest are those of \( \Omega_k \), where
\[
\Omega_k^T = [\eta_k, \eta_k', \cdots, \eta_k^{(N)}].
\]

These may be calculated obviously from (3.7) by simply replacing \( b_k \) in (3.7) by \( c_k \) where
\[
c_k^T = [-T + \tau, -T_{k1} + \tau_{k1}, \cdots, -T_{k,n-1} + \tau_{k,n-1}; \phi_n, \cdots, \phi_N],
\]
accompanied by a similar contraction of the matrices \( H(\lambda) \) and \( J_0^{-1} \).

The determination of the characteristic values \( \lambda \) of \( C_0 = J_0^{-1}A \) and the construction of the error vector may be simplified somewhat, as follows:

Let us define
\[
(3.8) \quad \Gamma(\Lambda) = \Lambda I - J_0 A,
\]
where \( J, A \) are square matrices of orders \( s \), and \( J_a \) denotes the adjoint of \( J \). Similarly, let \( G_a \) denote the adjoint of \( G(\lambda) = \lambda I - J^{-1}A \). Let, further,
\[
\Delta(\Lambda) = \det \Gamma(\Lambda),
\]
and, as before, \( \delta(\lambda) = \det G(\lambda) \).

Then we have the following

**Lemma.** \( G_a/\delta' = \Gamma_a/\Delta' \).

**Proof.** Clearly
\[
(3.9) \quad \Gamma(\Lambda) = D[\Delta D^{-1} - J^{-1}A] = DG(\Lambda/D),
\]
where again \( D = \det J \), \( G(\lambda) = \lambda I - J^{-1}A \). Consequently,
\[
\Delta(\Lambda) = D^*\delta(\Lambda/D).
\]
To each root $\Lambda$ of $\Delta(\Lambda) = 0$ there is then associated a root
$$\lambda = \Lambda/D$$
of $\delta(\lambda) = 0$. Further,
$$(3.10) \quad \delta'(\lambda) = D^{-t+1} \Delta'(\Lambda).$$
However, by (3.9),
$$(3.11) \quad \Gamma_a(\Lambda) = D^{r-1} \Gamma_a(\lambda)$$
which, together with (3.9), proves the lemma.

We have thus obtained the following general

**Propagation Theorem.**

$$(3.12) \quad \Omega_k = \sum_m \left[ \Gamma_a(\Lambda_m) J_a(D\Lambda(\Lambda_m)) \right] \sum_t \lambda_m^{k-t-1} c_t.$$  

In this theorem the index $m$ is to be extended over all distinct non-zero characteristic roots $\Lambda_m$ of $\Delta(\Lambda) = 0$, $\lambda_m = \Lambda_m/D(J)$, and the elements of the vector $c_k$ are due to truncation error and rounding. The theorem shows again that for a method to be stable for sufficiently small $h$ it is sufficient that all characteristic roots $\lambda_m$ be of absolute value not exceeding unity.

The actual computation of $\Omega_k$ would then proceed in obvious fashion from the construction of $J_0 A$ and $\Delta(\Lambda)$ to the calculation of $D(J), \Delta_m,$ and $\Gamma_a(\Lambda_m) \cdot J_a$, and could be carried out concurrently with the integration.

4. **The propagation of error in the case** $n = N = 1$. The deductions of the previous sections will be applied now to a number of well known methods of numerical integration. We shall start by considering the general first order differential equation
$$y' = f(x, y).$$

In this case the associated homogeneous variational equation is
$$\eta' = f_y \eta;$$
it has the fundamental solution
$$\eta(x) = \exp \left( \int_{x_0}^x f_y dt \right).$$

Among the solutions $\lambda$ of the characteristic equation $\delta(\lambda) = 0$ inherent in any useful method of integration there must be one, $\lambda = \lambda_1$, that approaches this fundamental solution $\eta(x)$ as the step size $h$
goes to zero. It will be seen that this root $\lambda_1$ is of the form

$$\lambda_1 = 1 + hf_v + \cdots \approx \exp hf_v \equiv \exp \rho, \rho = hf_v;$$

it gives rise in (3.12) to the principal term of the form

$$\Omega_k(\lambda_1) = M_1\lambda_1^{k-1-1} \approx M_1 \exp \left( \int_{z'_{+1}}^z f_v dt \right).$$

In order to prevent the error in the large from increasing rapidly it is thus sufficient to carry out the integration in the direction $Ax$ in which $f_v\Delta x$ is nonpositive.

Let us consider first the important case $n = N = 1$. Here we have

$$J = \begin{bmatrix} I & -J_{01} \\ -J_{10} & I \end{bmatrix}, \quad J_{01} = \begin{bmatrix} h & a_{1, -1} & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 \end{bmatrix}, \quad J_{10} = \begin{bmatrix} f_v & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix},$$

$$A = \begin{bmatrix} A_{00} & A_{01} \\ 0 & A_{11} \end{bmatrix}, \quad A_{00} = \begin{bmatrix} a_{00} & a_{01} & \cdots & a_{0r} \\ 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{bmatrix},$$

$$A_{01} = \begin{bmatrix} ha_{10} & ha_{11} & \cdots & ha_{1r} \\ 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix}, \quad A_{11} = \begin{bmatrix} 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 \end{bmatrix}.$$  

It follows that

$$D = 1 - \rho a_1, \quad a_1 \equiv a_{1, -1},$$

$$J_a = \begin{bmatrix} K & J_{01} \\ J_{10} & K \end{bmatrix}, \quad K = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & D & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 \end{bmatrix},$$

$$\Gamma(\Lambda) = \Lambda I - J_a J = \begin{bmatrix} \Gamma_{00} & \Gamma_{01} \\ \Gamma_{10} & \Gamma_{11} \end{bmatrix},$$

$$\Gamma_{00} = \begin{bmatrix} \Lambda - a_{00} & -a_{01} & \cdots & -a_{0,r-1} & -a_{0r} \\ -D & \Lambda & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -D & \Lambda \end{bmatrix}.$$
The characteristic equation $\Delta(\Lambda) = 0$ may be expressed in the form

$$\Delta(\Lambda) = \Lambda^{r+1} - \Delta_0(\Lambda) = 0,$$

$$\Delta_0(\Lambda) = \Lambda^{r+1} - e_0\Lambda^r - e_1\Lambda^{r-1} - \cdots - e_r,$$

where

$$e_\rho = D^\rho(a_{00} + \rho a_{10}), \quad \rho = 0, 1, \cdots, r.$$

The equation for the extraneous roots $\Lambda_m$, if any, is now quite easily obtained from (4.2).

It is convenient to write the characteristic expression in the form

$$\Delta_0(\Lambda) = \Lambda^{r+1} - \Delta_0(\Lambda) - p\Delta_{c1}(\Lambda),$$

where

$$\Delta_{c1}(\Lambda) = \Lambda^{r+1} - \Delta_0(\Lambda) - \Lambda^r a_{i0} + \Lambda^{r-1} Da_{i1} + \cdots + D^r a_{ir}, \quad i = 0, 1.$$

Since $\Lambda_1 = D\Lambda_1 = (1 + \rho)D$ is a solution of $\Delta_0(\Lambda) = 0$, there are obtained for the coefficients $a_{ij}$ the relationships

$$\sum_{j=0}^r a_{0j} = 1,$n

$$a_1 + \sum_{j=0}^r a_{1j} - \sum_{j=0}^r (j) a_{0j} = 1.$$
\[
\Gamma_{\alpha}(\Lambda) = \Lambda^r \begin{bmatrix} \Delta c_0(\Lambda) & h\Delta c_1(\Lambda) \\ f_y \Delta c_0(\Lambda) & \rho \Delta c_1(\Lambda) \end{bmatrix},
\]
or
\[
\Gamma_{\alpha}(\Lambda) = \Lambda^r \begin{bmatrix} 1 \\ f_y \end{bmatrix} \begin{bmatrix} \Delta c_0, \Delta c_1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & h \end{bmatrix}.
\]
Therefore,
\[
\Gamma_{\alpha}(\Lambda) J_{\alpha} = \Lambda^r \begin{bmatrix} 1 \\ f_y \end{bmatrix} \begin{bmatrix} \Lambda^{r+1}, h(a_1 \Delta c_0(\Lambda) + \Delta c_1(\Lambda)) \end{bmatrix},
\]
and, finally,
\[
(4.6) \quad \begin{bmatrix} \eta_k \\ \eta'_k \end{bmatrix} = \frac{1}{D} \sum_m \frac{1}{\Lambda_m \Delta'_c(\Lambda_m)} \begin{bmatrix} \Lambda^{r+1}_m, h(a_1 \Delta c_0 + \Delta c_1) \end{bmatrix} \sum_t \lambda_m^{k-t-1} c_t.
\]
The error \(\eta'_k\) may thus be obtained quite simply from \(\eta_k\) by multiplication with \(f_y\).

The contribution of the root \(\Lambda_1 = (1+p)D\) to the error \(\eta_k\) may now be written down at once; it is
\[
\eta_k(\Lambda_1) = \frac{1}{\Delta'_c(\Lambda_1)} \begin{bmatrix} 1 + p(r - (r - 1)a_1), h \left( a_1 + \sum_0^r a_{1j} \right) \end{bmatrix}
\cdot \sum_t \exp \left( \int f_y dx \right) c_t.
\]
\[
(4.7)
\]
It is of interest to apply above deductions to some specific cases.

I. Case \(r = 0\). As was pointed out above no extraneous solutions arise in this case, so that the methods are stable in the direction in which \(p < 0\). Techniques falling into this class are due to Euler, Heun, Runge-Kutta, Milne, and others.

Since now
\[
\Delta_c(\Lambda) = \Lambda - (a_{00} + pa_{10}),
\]
there is obtained from (4.7) and (4.5)
\[
\eta_k = [1 + pa_1, h] \sum_t \exp \left( \int_{x_{t+1}}^{x_t} f_y dx \right) c_t,
\]
\[
c_t = \begin{bmatrix} -T_t + \tau_t \\ \phi \end{bmatrix},
\]
or
(4.8) \[ \eta_k \approx \int_{x_0}^{x_k} \left[ (-T_t + \tau_t)/h + a_1 f + \phi \right] \exp \left( \int_{x_{t+1}}^{x_k} f \, dx \right) \, dt. \]

I, 1. Euler's method.

\[ *y_{k+1} = \eta_k + h * y_k, \]

\[ (a_{00} \ a_1 \ a_{10}) = (1, 0, 1), \quad T = (h^2/2)y''. \]

\[ \eta_k \approx \int \left[ -(h/2)y'' + \tau_t/h + \phi \right] \exp \left( \int f \, dx \right) \, dt. \]

I, 2. Heun's (modified Euler) method.

\[ *y_{k+1} = \eta_k + (h/2)(*y_k + *y_{k+1}), \]

\[ (a_{00} \ a_1 \ a_{10}) = (1, 1/2, 1/2), \quad T = - (h^3/12)y'''. \]

\[ \eta_k \approx \int \left[ - (h/2) + \tau_t/h \right] \sum_i \exp \left( \int f \, dx \right) \, c_i. \]

II. Case \( r=1 \). There is one extraneous root \( \Lambda_2 \); it satisfies the equation

(4.9) \[ \Delta_c(\Lambda) \equiv \Lambda^2 - \Delta_{c0} - \rho \Lambda_{c1} = 0 \]

where

\[ \Delta_{c0}(\Lambda) = \Lambda a_{00} + Da_{01}, \quad \Delta_{c1}(\Lambda) = \Lambda a_{10} + Da_{11}. \]

Since \( \Lambda_1 + \Lambda_2 = a_{00} + \rho a_{10} \), and \( a_{ij} \) satisfy (4.5), it follows that

(4.10) \[ \Lambda_2 = - a_{01} + \rho (a_{01} - a_{11}). \]

Further

\[ \Delta'_{c}(\Lambda_1) = (2 - a_{00}) + \rho [2(1 - a_1) - a_{10}], \quad \Delta'_{c}(\Lambda_2) = - \Delta'_{c}(\Lambda_1). \]

II, 1. Simple central difference method.

\[ *y_{k+1} = *y_{k-1} + 2h * y_k', \]

\[ (a_{00} \ a_0 \ a_1 \ a_{10} \ a_{11}) = (0, 1, 0, 2, 0), \quad T = (h^3/3)y'''. \]

Thus

\[ D(J) = 1, \]

\[ \Lambda_m = \pm 1 + \rho, \quad m = 1, 2, \]
\[ \Delta_c(\Lambda_1) = 2, \]

and, consequently,

\[ \eta_k = \frac{1}{2} \left\{ [1 + p, 2h] \sum_i \exp \left( \int f_y \, dx \right) \right. \]

\[ + \left. \left[ -1 + p, 2h \right] \sum_i (-1)^{k-i} \exp \left( -\int f_y \, dx \right) \right \} c_i. \]  

(4.11)

The extraneous root \( \Lambda_2 \) may thus give rise to an oscillating term of increasing magnitude whenever the integration is applied in a direction in which \( hf_y < 0 \). However, it is entirely possible that the actual increase of this term is choked off by the rounding procedure itself. See footnote 1. Used for integrations in the opposite direction, over short ranges, the method may give useful results.

II, 2. Simpson’s method. Here

\[ *y_{k+1} = *y_{k-1} + \left( h/3 \right) (*y_{k+1}' + 4*y_k' + *y_{k-1}'), \]

whence

\[ \left( a_{00} \ a_{01} \ a_1 \ a_{10} \ a_{11} \right) = (0, 1, 1/3, 4/3, 1/3), \]

\[ T = - \left( h^5/90 \right) y'. \]

It follows that

\[ D(J) = 1 - p/3, \quad \Lambda_m = \pm 1 + 2p/3, \quad \Delta_c'(\Lambda_1) = 2. \]

Therefore,

\[ \eta_k = \frac{1}{2} \left\{ [1 + p, 2h] \sum_i \exp \left( \int f_y \, dx \right) \right. \]

\[ \left. + \left[ -1 + p/3, 2h/3 \right] \sum_i (-1)^{k-i} \exp \left( -\int f_y \, dx/3 \right) \right \} c_i. \]  

(4.12)

The second root \( \Lambda_2 \) may thus again make this integration method unsuitable.

III. Case \( r = 2 \). The three roots \( \Lambda_m, m = 1, 2, 3 \), satisfy the characteristic equation

\[ \Delta_c(\Lambda) = \Lambda^3 - \Delta_{e0}(\Lambda) - p\Delta_{e1}(\Lambda) = 0, \]

with

\[ \Delta_{ei}(\Lambda) = \Lambda^2 a_{i0} + \Lambda Da_{i1} + D^2 a_{i2}, \quad i = 0, 1. \]
The 2 extraneous roots $\Lambda_2, \Lambda_3$ may thus be obtained from
\[(4.13) \quad \Lambda^2 - \Lambda [(a_{00} - 1) + p(a_1 + a_{10} - 1)]
- \frac{D}{- a_{02} + p(a_{02} - a_{12})} = 0.\]

III, 1. Adams method.
\[\ast y_{k+1} = y_k + h \left[ y_k' + \frac{1}{2} \nabla y_k' + \frac{5}{12} y_k'' \right] + \cdots\]
with
\[\nabla^{(i+1)} q_k = \nabla^{(i)} q_k - \nabla^{(i)} q_{k-1}.\]
Thus, alternately,
\[\ast y_{k+1} = y_k + \left( \frac{h}{12} \right) \left[ 23 y_k' - 16 y_{k-1}' + 5 y_{k-2}' \right] + \cdots.\]
Then,
\[(a_{00}, a_{01}, a_{02}, a_1, a_{10}, a_{11}, a_{12}) = \left( \frac{1}{12} \right)(12, 0, 0, 23, -16, 5).\]
The 2 extraneous roots $\Lambda_2, \Lambda_3$ satisfy
\[\Lambda^2 - \left( \frac{11p}{12} \right) \Lambda + \left( \frac{5p}{12} \right) = 0.\]
Therefore,
\[\Lambda_m = \pm \left( - \frac{5p}{12} \right)^{1/2} + \frac{11p}{24}, \quad m = 2, 3.\]

For sufficiently small $h$ this method, then, is stable, and the propagation of error depends essentially on $\Lambda_1$.
Since
\[\Delta'_c(\Lambda_1) = 1 + \frac{3p}{2},\]
it is found that
\[(4.14) \quad \eta_k = \left[ 1 + \frac{p}{2}, h \right] \sum_i \exp \left( \int f_y dx \right) c_i.\]

III, 2. Gregory’s method. Here
\[\ast y_{k+1} = y_k + h \left[ \ast y_{k+1} - \frac{1}{2} \nabla y_{k+1} - \left( \frac{1}{12} \right) \nabla'' \ast y_{k+1}
- \left( \frac{1}{24} \right) \nabla''' \ast y_{k+1} \right] + \cdots
= y_k + \left( \frac{h}{24} \right) \left[ 9 y_{k+1}' + 19 y_k' - 5 y_{k-1}' + \ast y_{k-2}' \right] + \cdots.\]
Thus
\[(a_{00} a_{01} a_{02} a_1 a_{10} a_{11} a_{12}) = \left( \frac{1}{24} \right)(24, 0, 0, 9, 19, -5, 1),
T = - \left( \frac{19}{720} \right) h^8 y^2.\]
By (4.13), then, $D = 1 - 3p/8$,

$$\Lambda^2 - (p/6)\Lambda + p/24 = 0,$$

so that

$$\Lambda_m = \pm (- p/24)^{1/2} + p/12, \quad m = 2, 3.$$

The method thus has the same stability properties as Adams' method.

Since again

$$\Delta_c'(\Lambda) = 1 + 3p/2,$$

one obtains from (4.7)

(4.15) \[ \eta_k = [1 + p/8, h] \sum \exp \left( \int f_u dx \right) c_i. \]

5. Numerical example. To test the propagation theorem let us integrate the differential equation

(5.1) \[ y' = y - 2x/y \]

by means of Simpson's method II, 2. Taking $h = 0.5$ and starting at $x = 0$ with values computed from the exact solution

$$y(x) = (2x + 1)^{1/2},$$

there are obtained the "solutions" shown in columns (2) and (3) of Table I. At each step a sufficient number of iterations is carried out in order to achieve agreement to five decimals (column (2)), or four decimals (column (3)). Due to the instability of the method the five-decimal "solution" diverges more and more from the four-decimal "solution."

The fourth column (4) contains the exact solution $y(x)$, and the fifth column (5) the error $\eta = y_\ast - y(x)$, the solution $y_\ast$ taken from column (3).

The growth of error may be inferred from (4.12), or, somewhat more accurately, from

(5.2) \[ \eta_k(\lambda_1) \approx \frac{1}{2} \sum_{i=0}^{k-1} \left[ (1 + p)((h^8/90)y_i^\ast + \tau_i) + 2h\phi \right] \lambda_1^{k-t-1}, \]

(5.3) \[ \eta_k(\lambda_2) \approx - \frac{1}{2} \sum_{i=0}^{k-1} \left[ (-1 + p/3)((h^8/90)y_i^\ast + \tau_i) + (2/3)h\phi \right] \lambda_2^{k-t-1}. \]
Table I. Integration of \( y' = y - 2x/y \), \( h = 0.5 \)

<table>
<thead>
<tr>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x )</td>
<td>( *y )</td>
<td>( *y )</td>
<td>( y(x) )</td>
<td>( \eta )</td>
</tr>
<tr>
<td>0</td>
<td>1.00000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0</td>
</tr>
<tr>
<td>0.5</td>
<td>1.41421</td>
<td>1.4142</td>
<td>1.4142</td>
<td>0</td>
</tr>
<tr>
<td>1.0</td>
<td>1.73516</td>
<td>1.7352</td>
<td>1.7320</td>
<td>.0032</td>
</tr>
<tr>
<td>1.5</td>
<td>2.00529</td>
<td>2.0053</td>
<td>2.0000</td>
<td>.0053</td>
</tr>
<tr>
<td>2.0</td>
<td>2.25064</td>
<td>2.2507</td>
<td>2.2361</td>
<td>.0146</td>
</tr>
<tr>
<td>2.5</td>
<td>2.48438</td>
<td>2.4845</td>
<td>2.4495</td>
<td>.0350</td>
</tr>
<tr>
<td>3.0</td>
<td>2.73397</td>
<td>2.7342</td>
<td>2.6458</td>
<td>.0884</td>
</tr>
<tr>
<td>3.5</td>
<td>3.04775</td>
<td>3.0483</td>
<td>2.8284</td>
<td>.220</td>
</tr>
<tr>
<td>4.0</td>
<td>3.53706</td>
<td>3.5383</td>
<td>3.0000</td>
<td>.538</td>
</tr>
<tr>
<td>4.5</td>
<td>4.42054</td>
<td>4.4232</td>
<td>3.1623</td>
<td>1.26</td>
</tr>
<tr>
<td>5.0</td>
<td>6.07815</td>
<td>6.0834</td>
<td>3.3166</td>
<td>2.77</td>
</tr>
<tr>
<td>5.5</td>
<td>9.08576</td>
<td>9.0953</td>
<td>3.4641</td>
<td>5.63</td>
</tr>
<tr>
<td>6.0</td>
<td>14.31274</td>
<td>14.3292</td>
<td>3.6056</td>
<td>10.7</td>
</tr>
<tr>
<td>6.5</td>
<td>23.14506</td>
<td>23.1727</td>
<td>3.7417</td>
<td>19.4</td>
</tr>
<tr>
<td>7.0</td>
<td>37.86292</td>
<td>37.9089</td>
<td>3.8730</td>
<td>34.0</td>
</tr>
<tr>
<td>7.5</td>
<td>62.23708</td>
<td>62.3131</td>
<td>4.0000</td>
<td>58.3</td>
</tr>
<tr>
<td>8.0</td>
<td>102.49977</td>
<td>102.6253</td>
<td>4.1231</td>
<td>98.5</td>
</tr>
<tr>
<td>8.5</td>
<td>168.93852</td>
<td>169.1431</td>
<td>4.2426</td>
<td>165</td>
</tr>
<tr>
<td>9.0</td>
<td>278.52584</td>
<td>278.8552</td>
<td>4.3589</td>
<td>274</td>
</tr>
<tr>
<td>9.5</td>
<td>459.25450</td>
<td>459.8020</td>
<td>4.4721</td>
<td>455</td>
</tr>
<tr>
<td>10.0</td>
<td>757.28847</td>
<td>758.1877</td>
<td>4.5826</td>
<td>754</td>
</tr>
</tbody>
</table>

For our example,

\[
\begin{align*}
k &= 20, \quad h = 1/2, \quad p \equiv hf_y = 1 - [2(2x + 1)]^{-1}, \\
y^\nu &= 105(2x + 1)^{-9/2}, \\
|\tau|, \quad |\phi| &= c_110^{-5}, \quad c_1 < 10,
\end{align*}
\]

and

\[
\lambda_1 = 1 + p.
\]

Now \( p \) increases from 0.5 at \( x = 0 \) to 0.98 at \( x = 20 \), so that \( p \approx 0.7 \) could be taken as average value. Furthermore, for the terms corresponding to the low values of \( x \), which contribute most to \( \eta_k(\lambda_1) \), \( \tau \) and \( \phi \) are negligible. Thus, by (5.2),

\[
(5.4) \quad \eta_k(\lambda_1) \approx \frac{1}{2} \sum_{t=0}^{19} (1/2880)y^\nu_t(1.7)^{20-t}.
\]

One obtains
$\eta_k(\lambda_1) \approx 762.$

Since $\lambda_2 = -1 + \rho/3$, so that

$$\eta_k(\lambda_2) \approx -\frac{1}{2} \sum_{t=0}^{19} (1/2880) y_i^\rho (-0.77)^{20-t},$$

and consequently negligible, we get

$$\eta_k \approx \eta_k(\lambda_1) \approx 762,$$

which compares very favorable indeed with the actual value of 754 for the total error.

In order to examine the oscillating term $\eta_k(\lambda_2)$, let us integrate again (5.1), this time starting at $x = 60$, and taking $h = -1$. The "solution" is shown in column (2) of Table 2.

Now,

$$k = 34, \quad h = -1, \quad \rho = -2 + (2x + 1)^{-1} \approx -2, \quad |r|, \quad |\phi| = c_2 \cdot 10^{-6}, \quad c_2 \approx 2,$$

$$\lambda_2 = -1 + \rho/3 \approx -1.66.$$

Since for low values of $t$ the term $(h^5/90)y_i^t$ is less than $1.10^{-7}$, in absolute value, there is obtained the expression

$$(5.5) \quad |\eta_k(\lambda_2)| \approx 10^{-6} \sum_{t=0}^{33} (-1 + \rho/3)^{34-t}.$$

This formula leads to

$$|\eta_k(\lambda_2)| \approx 19.0,$$

which is very close indeed to the exact error given in Table 2.

The exhibited expressions, then, lead to quite useful estimates of the error.

**References**

Table 2. Integration of $y' = y - 2x/y$, $h = -1$

<table>
<thead>
<tr>
<th>$x$</th>
<th>$y$</th>
<th>$y(x)$</th>
<th>$10^8 \eta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>11.00000</td>
<td>11.00000</td>
<td>0</td>
</tr>
<tr>
<td>59</td>
<td>10.90871</td>
<td>10.90871</td>
<td>0</td>
</tr>
<tr>
<td>58</td>
<td>10.81665</td>
<td>10.81665</td>
<td>0</td>
</tr>
<tr>
<td>57</td>
<td>10.72381</td>
<td>10.72381</td>
<td>0</td>
</tr>
<tr>
<td>56</td>
<td>10.63014</td>
<td>10.63015</td>
<td>-1</td>
</tr>
<tr>
<td>55</td>
<td>10.53566</td>
<td>10.53565</td>
<td>1</td>
</tr>
<tr>
<td>54</td>
<td>10.44030</td>
<td>10.44031</td>
<td>-1</td>
</tr>
<tr>
<td>53</td>
<td>10.34409</td>
<td>10.34408</td>
<td>1</td>
</tr>
<tr>
<td>52</td>
<td>10.24694</td>
<td>10.24695</td>
<td>-1</td>
</tr>
<tr>
<td>51</td>
<td>10.14891</td>
<td>10.14889</td>
<td>2</td>
</tr>
<tr>
<td>50</td>
<td>10.04984</td>
<td>10.04988</td>
<td>-4</td>
</tr>
<tr>
<td>49</td>
<td>9.94994</td>
<td>9.94987</td>
<td>7</td>
</tr>
<tr>
<td>48</td>
<td>9.84875</td>
<td>9.84886</td>
<td>-11</td>
</tr>
<tr>
<td>47</td>
<td>9.74698</td>
<td>9.74679</td>
<td>19</td>
</tr>
<tr>
<td>46</td>
<td>9.64333</td>
<td>9.64365</td>
<td>-32</td>
</tr>
<tr>
<td>45</td>
<td>9.53994</td>
<td>9.53939</td>
<td>55</td>
</tr>
<tr>
<td>44</td>
<td>9.43304</td>
<td>9.43398</td>
<td>-94</td>
</tr>
<tr>
<td>43</td>
<td>9.32899</td>
<td>9.32738</td>
<td>161</td>
</tr>
<tr>
<td>42</td>
<td>9.21679</td>
<td>9.21954</td>
<td>-275</td>
</tr>
<tr>
<td>41</td>
<td>9.11515</td>
<td>9.11043</td>
<td>472</td>
</tr>
<tr>
<td>40</td>
<td>8.99193</td>
<td>9.00000</td>
<td>-807</td>
</tr>
<tr>
<td>39</td>
<td>8.90203</td>
<td>8.88819</td>
<td>1384</td>
</tr>
<tr>
<td>38</td>
<td>8.75130</td>
<td>8.77496</td>
<td>-2366</td>
</tr>
<tr>
<td>37</td>
<td>8.70087</td>
<td>8.66025</td>
<td>4062</td>
</tr>
<tr>
<td>36</td>
<td>8.47474</td>
<td>8.54400</td>
<td>-6926</td>
</tr>
<tr>
<td>35</td>
<td>8.54560</td>
<td>8.42615</td>
<td>11945</td>
</tr>
<tr>
<td>34</td>
<td>8.10464</td>
<td>8.30662</td>
<td>-20198</td>
</tr>
<tr>
<td>33</td>
<td>8.53865</td>
<td>8.18535</td>
<td>35330</td>
</tr>
<tr>
<td>32</td>
<td>7.47987</td>
<td>8.06226</td>
<td>-58239</td>
</tr>
<tr>
<td>31</td>
<td>9.00031</td>
<td>7.93725</td>
<td>106306</td>
</tr>
<tr>
<td>30</td>
<td>6.19044</td>
<td>7.81025</td>
<td>-161981</td>
</tr>
<tr>
<td>29</td>
<td>11.03789</td>
<td>7.68115</td>
<td>335674</td>
</tr>
<tr>
<td>28</td>
<td>3.61153</td>
<td>7.54983</td>
<td>-393830</td>
</tr>
<tr>
<td>27</td>
<td>19.42204</td>
<td>7.41620</td>
<td>1200584</td>
</tr>
<tr>
<td>26</td>
<td>-12.03925</td>
<td>7.28011</td>
<td>-1931936</td>
</tr>
</tbody>
</table>

RCA Service Company, Patrick Air Force Base