

# Linear Multistep Methods with Mildly Varying Coefficients

By J. D. Lambert

**Abstract.** Consideration of a common assumption in the theory of weak stability of linear multistep methods for ordinary differential equations leads to the study of a class of linear multistep methods with mildly varying coefficients. It is well known that, in the case of constant-coefficient methods, optimal stable methods suffer from weak instability. Corresponding methods of the new class, of step-number 2 and 4, which do not suffer from weak instability are derived.

**1. Introduction.** The general linear multistep method for the numerical solution of the initial-value problem

$$(1.1) \quad y' = f(x, y); \quad y(a) = \eta$$

on an interval  $a \leq x \leq b$  may be written as

$$(1.2) \quad \sum_{j=0}^k \alpha_j y_{n+j} = h \sum_{j=0}^k \beta_j f_{n+j}$$

where  $y_j = y(x_j)$ ,  $f_j = f(x_j, y_j)$ ,  $x_j = a + jh$  and  $h$ ,  $\alpha_j$  and  $\beta_j$  are constants. The theory of the convergence of the solution of (1.2), with appropriate starting values to the solution of (1.1), is well understood [1], [2]. In particular, the necessary and sufficient conditions for convergence are that the linear multistep method be consistent, and stable in the sense that the associated polynomial  $\rho(\zeta) = \sum_{j=0}^k \alpha_j \zeta^j$  has all its roots of modulus less than or equal to unity, and has no multiple roots of modulus unity. The main effect of this important theorem is that, in general, we reject out of hand any method of class (1.2) which is not stable and consistent. It is well known, [1], that the maximum order that a stable linear  $k$ -step method can attain is  $k + 2$  and that this order can be achieved only when  $k$  is even and all the roots of  $\rho(\zeta)$  have modulus unity. In this paper we shall refer to stable linear  $k$ -step methods of order  $k + 2$  as *optimal stable methods*.

The theorem of Dahlquist does not, however, guarantee that a stable consistent method will give satisfactory numerical results when we calculate with a fixed nonzero value for  $h$ ; nor does it help us to choose a value for  $h$  which can be expected to give acceptable results for a given initial value problem. To answer this last point, theories of weak, conditional or numerical stability have been developed by many authors [3], [4], [5]. In particular, it is known that optimal stable methods exhibit weak instability.

It is the purpose of this paper to show that, by replacing the constant coefficients  $\alpha_j$ ,  $\beta_j$  of (1.2) by variable coefficients  $\alpha_j + ha_j(x_n)$ ,  $\beta_j + hb_j(x_n)$ , where

---

Received July 16, 1968, revised June 2, 1969.

AMS Subject Classifications. Primary 6560, 6561; Secondary 6570.

Key Words and Phrases. Linear multistep methods, weak stability, initial value problem, variable coefficient linear difference equations.

the  $a_j(x)$  and  $b_j(x)$  are functions specified by the particular differential equation under consideration, it is possible to control the weak stability characteristics of the method. In particular, the weak instability associated with optimal stable methods can be removed, and stabilized versions of these methods are derived in the cases  $k = 2, k = 4$ . It is hoped that the analysis, which is necessarily restricted to the case of a scalar differential equation, will throw light on the mechanism of weak stability.

The investigation starts from a consideration of the assumptions commonly made in theories of weak stability. A typical argument runs as follows. Let the method (1.2) be stable and consistent, let the exact solution of (1.1) be  $Y(x)$  and let  $Y_j$  denote  $Y(x_j)$ . Then

$$\sum_{j=0}^k \alpha_j Y_{n+j} = h \sum_{j=0}^k \beta_j f(x_{n+j}, Y_{n+j}) + T_n$$

where  $T_n$  is the local truncation error. If  $\{y_n\}$  now denotes the computed solution of the difference equation obtained by substituting for  $f$  from (1.1) in (1.2), then

$$\sum_{j=0}^k \alpha_j y_{n+j} = h \sum_{j=0}^k \beta_j f(x_{n+j}, y_{n+j}) + R_n$$

where  $R_n$  is the local round-off error. Subtracting these two equations, and applying the Mean Value Theorem, gives the following equation for the accumulated error,  $\epsilon_n = Y_n - y_n$ :

$$(1.3) \quad \sum_{j=0}^k \alpha_j \epsilon_{n+j} = h \sum_{j=0}^k \beta_j \frac{\partial f}{\partial y}(x_{n+j}, \xi_{n+j}) \epsilon_{n+j} + E_n$$

where  $\xi_{n+j} \in (y_{n+j}, Y_{n+j})$  and  $E_n = T_n - R_n$ . Two assumptions are now made:

*Assumption (a).*  $E_n = E$ , constant.

*Assumption (b).*  $\partial f(x, y)/\partial y = -q$ , constant.

The error equation now reduces to the linear constant-coefficient difference equation

$$(1.4) \quad \sum_{j=0}^k (\alpha_j + hq\beta_j) \epsilon_{n+j} = E.$$

If  $r_m$  ( $m = 1, 2, \dots, k$ ) are the roots (assumed distinct) of the polynomial  $\sum_{j=0}^k (\alpha_j + hq\beta_j) r^j$ , then the solution of (1.4) is

$$\epsilon_n = \sum_{j=0}^k B_j r_j^n + E / \left( hq \sum_{j=0}^k \beta_j \right),$$

where  $B_j$  ( $j = 0, 1, \dots, k$ ) are arbitrary constants. Obvious modifications are made in the case when the roots  $r_m$  are not all distinct. As  $h \rightarrow 0$ , the roots  $r_m$  tend to the roots  $\zeta_m$  of  $\rho(\zeta)$  and, in particular, there will exist a root  $r_1$  which tends to  $\zeta_1 = +1$ , since, by consistency,  $\rho(\zeta)$  must have a root at  $+1$ . By stability, the remaining roots  $\zeta_m$  ( $m = 2, 3, \dots, k$ ) of  $\rho(\zeta)$  all lie in or on the unit circle. Weak instability in a stable method arises when, for some range of  $hq$ , any of the roots  $r_m$  ( $m = 2, 3, \dots, k$ ) lie outside the unit circle (or outside the circle  $|z| = |r_1|$  if the weak instability is defined in a relative rather than an absolute sense). The most familiar example of this phenomenon arises with Simpson's rule, for which we

find  $\zeta_1 = 1, \zeta_2 = -1, r_1 = +1 - hq + O(h^2), r_2 = -1 - \frac{1}{3}hq + O(h^2)$ , whence the well-known fact that the method exhibits weak instability (both absolute and relative) for all positive  $h$ , when  $\partial f/\partial y$  is negative.

Of the two assumptions (a) and (b) made in the above argument, (b) is the harder to justify. It is strictly valid only in the case  $f(x, y) = -qy, q$  constant, and it is not difficult to fabricate differential equations for which the assumption is difficult to accept.

In this paper, we propose replacing Assumption (b) by

*Assumption (b')*.  $\partial f(x_{n+j}, \xi_{n+j})/\partial y = -q(x_n), (j = 0, 1, \dots, k)$ .

Thus we ignore variations in  $\partial f/\partial y$  throughout the interval spanned by the method, but permit variation of  $\partial f/\partial y$  with  $n$ . [In the analysis which follows, we do not need to make Assumption (a).] The error equation (1.3) now reduces to

$$(1.5) \quad \sum_{j=0}^k \{\alpha_j + hq(x_n)\beta_j\}\epsilon_{n+j} = E_n,$$

a linear difference equation with variable coefficients. We do not make the analysis of this equation any harder, and may well make it more profitable, if we replace  $\alpha_j$  by  $\alpha_j + ha_j(x_n)$  and  $\beta_j$  by  $\beta_j + hb_j(x_n)$ . We are thus motivated to study a class of linear multistep methods with variable coefficients.

**2. Linear Multistep Methods with Mildly Varying Coefficients.** The class of methods is defined by

$$(2.1) \quad \sum_{j=0}^k \tilde{\alpha}_j(x_n)y_{n+j} = h \sum_{j=0}^k \tilde{\beta}_j(x_n)f_{n+j}$$

$$\tilde{\alpha}_j(x_n) = \alpha_j + ha_j(x_n); \tilde{\beta}_j(x_n) = \beta_j + hb_j(x_n),$$

where  $\alpha_j, \beta_j$  are constants and  $|a_j(x)| \leq A, |b_j(x)| \leq B$  for all  $x \in [a, b], (j = 0, 1, \dots, k)$ , where  $A$  and  $B$  are finite constants.

We cannot proceed with confidence to study the application of method (2.1) with a fixed nonzero value for  $h$  until we have established the conditions under which the solution of (2.1), with appropriate starting values, will converge to the solution of (1.1).

*Order and Consistency.* Define the linear difference operators  $L$  and  $M$  by

$$L[y(x); h] = \sum_{j=0}^k \alpha_j y(x + jh) - h \sum_{j=0}^k \beta_j y'(x + jh),$$

$$M[y(x); x; h] = \sum_{j=0}^k a_j(x)y(x + jh) - h \sum_{j=0}^k b_j(x)y'(x + jh).$$

$L$  is the operator associated with the constant-coefficient linear multistep method (1.2), and its order  $p$  is defined in the usual way. If  $M$  is applied to a sufficiently differentiable function  $y(x)$ , we obtain

$$M[y(x); x; h] = D_0(x)y(x) + hD_1(x)y'(x) + \dots + h^s D_s(x)y^{(s)}(x) + \dots$$

If  $D_0(x) = D_1(x) = \dots = D_r(x) = 0$  for all  $x \in [a, b]$  and  $D_{r+1}(x) \neq 0$  for some  $x \in [a, b]$ , then we say  $M$  is of order  $r$ .

Let the operator  $\tilde{L}$  associated with (2.1) be defined by

$$\tilde{L}[y(x); x; h] = L[y(x); h] + hM[y(x); x; h].$$

Then the order of  $\tilde{L}$  is defined to be  $\tilde{p} = \min(p, r + 1)$ . Method (2.1) is defined to be consistent if  $\tilde{L}$  has order  $\tilde{p} \geq 1$ .

*Stability.* The method (2.1) associated with the operator  $\tilde{L}$  is said to be stable if the operator  $L$  is stable.

**THEOREM 1.** *A sufficient condition for the method (2.1) to be convergent is that it be stable and consistent.*

Proof is by a direct extension of the proof given by Henrici [2, pp. 242–246], for the corresponding theorem in the case of constant coefficient methods. That the extension is possible is a consequence of the dependence of Henrici's proof on a lemma (Lemma 5.6 of [2]) concerning the growth of solutions of a linear difference equation whose coefficients are mildly varying in just the sense that the coefficients of (2.1) are mildly varying. Details of the proof are given in the Appendix.

It is of interest to note that without further conditions on  $a_j(x)$  and  $b_j(x)$  we cannot show that stability and consistency are necessary for the convergence of the method (2.1). However, in later sections of this paper we shall be concerned with the case where the variable parts  $a_j(x)$  and  $b_j(x)$  of the coefficients in (2.1) are chosen proportional to  $q(x)$ . The following result holds.

**THEOREM 2.** *For the subclass of methods (2.1) for which  $a_j(x) = \hat{a}_j q(x)$ ,  $b_j(x) = \hat{b}_j q(x)$ , where  $q(x) = -\partial f(x, y(x))/\partial y$  and  $\hat{a}_j, \hat{b}_j$  ( $j = 0, 1, \dots, k$ ) are constants, stability is necessary for convergence.*

*Proof.* If the method is convergent, then it is so for the initial-value problem  $y' = f(x); y(a) = \eta$ . For this problem  $q(x) = 0$ , and hence  $a_j(x) = b_j(x) = 0$  for all  $x$ . Method (2.1) is then equivalent to method (1.2) whose stability is necessary for convergence, by the theorem of Dahlquist [1], [2]. The result follows, since stability of (1.2) implies stability of (2.1).

Similar reasoning applied to the question of the necessity of the consistency condition breaks down, since consistency of (1.2) does not imply consistency of (2.1). However, the existence of a convergent but inconsistent method of the subclass described in Theorem 2 would be of little practical interest, since its truncation error would be of order  $h$ .

**3. The Choice of Variable Coefficients.** We assume that method (2.1) is stable and consistent. Under Assumption (b') the error equation which arises when we apply (2.1) to (1.1) with a fixed nonzero value for  $h$  is

$$(3.1) \quad \sum_{j=0}^k \{\alpha_j + ha_j(x_n) + hq(x_n)[\beta_j + hb_j(x_n)]\} \epsilon_{n+j} = E_n.$$

Since we would like to choose  $a_j(x)$  and  $b_j(x)$  in order to minimise the chances of (2.1) developing a weak instability, we might be tempted to make the choice

$$(3.2) \quad a_j(x) = -\beta_j q(x), \quad (j = 0, 1, \dots, k).$$

This is a practical possibility, since  $q(x_n) = -\partial f(x_n, y_n)/\partial y$  may be computed step-by-step along with the solution. However, consistency demands that the order of the operator  $L$  be at least one, and that of the operator  $M$  be at least zero. The first

of these two requirements implies

- (i)  $\alpha_0 + \alpha_1 + \dots + \alpha_k = 0$ , and
- (ii)  $\alpha_1 + 2\alpha_2 + \dots + k\alpha_k = \beta_0 + \beta_1 + \dots + \beta_k$ , while the second implies
- (iii)  $a_0(x) + a_1(x) + \dots + a_k(x) = 0$  for all  $x \in [a, b]$ . Condition (3.2) together with (iii) can hold for all  $q(x)$  only if  $\sum_{j=0}^k \beta_j = 0$ , which, with (i) and (ii) implies that the polynomial  $\rho(\zeta)$  has a double root at  $\zeta = 1$ , contradicting the assumption that (2.1) is stable.

It is of interest to observe that if, despite the resulting instability, we were to make the choice (3.2) and use the remaining degrees of freedom in choosing coefficients in such a way that the order of (2.1) be as high as possible, we would find that the  $k + 1$  coefficients  $b_j(x)$  would be defined by a set of  $k + 1$  linear constant-coefficient equations whose inhomogeneous terms were all proportional to  $q(x)$ . Thus the  $b_j(x)$  as well as the  $a_j(x)$  would turn out to be proportional to  $q(x)$  and, by Theorem 2, the instability of (2.1) would necessarily cause divergence.

Since direct attempts to remove the terms of order  $h$  in (3.1) fail, we must find a solution for (3.1) in a form which indicates how an advantageous choice for the variable coefficients may be made. Let

$$(3.3) \quad \phi_j(n) = a_j(x_n) + q(x_n)\{\beta_j + hb_j(x_n)\}, \quad (j = 0, 1, \dots, k).$$

The error equation (3.1) may now be written as

$$(3.4) \quad \sum_{j=0}^k \alpha_j \epsilon_{n+j} = E_n - h \sum_{j=0}^k \phi_j(n) \epsilon_{n+j},$$

where there is no loss of generality in taking  $\alpha_k = +1$ . We now formally regard (3.4) as a *constant-coefficient* linear difference equation with an inhomogeneous term represented by the whole of the right-hand side. The complete solution of such an equation may be expressed as the sum of the complementary function and a particular solution; moreover, the latter may be represented as the convolution of the inhomogeneous term with the Green's function. Let  $\{\eta_n\}$  be the complete solution of the reduced equation

$$(3.5) \quad \sum_{j=0}^k \alpha_j \epsilon_{n+j} = 0.$$

The Green's function  $\{z_n\}$  is then the particular solution of (3.5) satisfying initial conditions  $z_0 = z_1 = \dots = z_{k-2} = 0, z_{k-1} = 1$ . We then have

$$(3.6) \quad \epsilon_n = \eta_n + \sum_{t=1}^n z_{n-t} d_t = \eta_n + \sum_{t=1}^{n-1} z_{n-t} d_t,$$

where

$$d_t = E_{t-1} - h \sum_{j=0}^k \phi_j(t-1) \epsilon_{t+j-1}.$$

Since the right-hand side of (3.4) contains the dependent variable  $\epsilon_n$ , (3.6) is not, of course, a solution of (3.4), but represents an alternative form of the Eq. (3.4). Direct substitution establishes that (3.6) and (3.4) are indeed equivalent.

Define

$$(3.7) \quad d_t^* = E_{t-1} - h \sum_{j=0}^k \phi_j(t-1) \eta_{t+j-1}.$$

From (3.6) and (3.7) it follows that

$$\begin{aligned} d_t^* - d_t &= h \sum_{j=0}^k \phi_j(t-1) \left\{ \sum_{s=1}^{t+j-2} z_{t+j-s-1} d_s \right\} \\ &= h \sum_{s=1}^{t+k-2} \left\{ \sum_{j=0}^k \phi_j(t-1) z_{t+j-s-1} \right\} d_s \end{aligned}$$

if we take  $z_n$  to be zero, when  $n$  is a negative integer. From the definition of  $z_n$ , we see that the expression in braces in the last equation vanishes when  $t < s$ , and hence we may write, for  $k \geq 2$ ,

$$(3.8) \quad d_t^* = d_t + h \sum_{s=1}^t \Phi_{ts} d_s, \quad t = 1, 2, \dots, T$$

where

$$\Phi_{ts} = \sum_{j=0}^k \phi_j(t-1) z_{t+j-s-1}.$$

For any fixed  $T$  let  $\mathbf{d}^*$  be the vector with components  $d_1^*, d_2^*, \dots, d_T^*$  and  $\mathbf{d}$  be the vector with components  $d_1, d_2, \dots, d_T$ . Then, since  $\Phi_{ts} = 0$  when  $s > t$ , we may write (3.8) in the form

$$\mathbf{d}^* = (I + h\Phi)\mathbf{d},$$

where  $\Phi$  is a lower triangular matrix. Moreover, if the original method (2.1) is stable,  $z_n$  is bounded as  $n \rightarrow \infty$ ,  $h \rightarrow 0$ ,  $nh$  fixed, and thus, recalling the definition (3.3) of  $\phi_j(t-1)$ , we see that  $\Phi_{ts} = O(1)$  in this limit. Hence for sufficiently small  $h$  the inverse of  $I + h\Phi$  exists, and we may write

$$\mathbf{d} = (I + h\Phi)^{-1} \mathbf{d}^*$$

or, equivalently,

$$(3.9) \quad d_t = \sum_{s=1}^t \psi_{ts} d_s^*, \quad t = 1, 2, \dots, T,$$

where  $\psi_{ts} = 0$  when  $s > t$ . Since (3.9) holds for any  $T$ , we may, from (3.6) express the solution of the error Equation (3.4) in the form

$$(3.10) \quad \epsilon_n = \eta_n + \sum_{t=1}^{n-1} z_{n-t} \left\{ \sum_{s=1}^t \psi_{ts} d_s^* \right\}.$$

If we now consider the case when  $\rho(\zeta) = \sum_{j=0}^k \alpha_j \zeta^j$  has distinct roots  $\zeta_1, \zeta_2, \dots, \zeta_n$  satisfying the condition of stability, we may write  $\eta_n = \sum_{p=1}^k A_p \zeta_p^n$ , where  $A_p$ ,  $p = 1, 2, \dots, k$  are arbitrary constants. (If  $\zeta_q, \zeta_{q+1}$  are a conjugate pair of complex roots, then  $A_q, A_{q+1}$  will be taken to be complex conjugates.) From (3.7), we may write

$$(3.11) \quad d_t^* = E_{t-1} - h \sum_{p=1}^k A_p \zeta_p^{t-1} F_p(t-1)$$

where

$$F_p(t-1) = \sum_{j=0}^k \phi_j(t-1) \zeta_p^j.$$

On substituting for  $\eta_n$  and  $d_s^*$  in (3.10) and rearranging, we find

$$(3.12) \quad \begin{aligned} \epsilon_n = & \sum_{p=1}^k A_p \left\{ \zeta_p^n - h \sum_{t=1}^{n-1} z_{n-t} \left( \sum_{s=1}^t \psi_{ts} \zeta_p^{s-1} F_p(s-1) \right) \right\} \\ & + \sum_{t=1}^{n-1} z_{n-t} \left( \sum_{s=1}^t \psi_{ts} E_{s-1} \right). \end{aligned}$$

This equation displays the solution for  $\epsilon_n$  in the form of complementary function plus particular solution, the former being a linear combination of fundamental solutions of the form  $\zeta_p^n + P_p(n)$ , where

$$P_p(n) = -h \sum_{t=1}^n z_{n-t} \left( \sum_{s=1}^t \psi_{ts} \zeta_p^{s-1} F_p(s-1) \right), \quad p = 1, 2, \dots, k.$$

$P_p(n)$  may be interpreted as a perturbation of the corresponding fundamental solution  $\zeta_p^n$  of the equation  $\sum_{j=0}^k \alpha_j \epsilon_{n+j} = 0$ .

We shall now show that if  $F_p(s-1) = O(h)$  for all  $s$  as  $h \rightarrow 0$ ,  $n \rightarrow \infty$ ,  $nh$  fixed, then the perturbation  $P_p(n) = O(h)$ . To do this, we must find the order of magnitude of the coefficients  $\psi_{ts}$ . From (3.8) it is clear that

$$\psi_{tt} = \frac{1}{1 + h\Phi_{tt}} = O(1), \quad t = 1, 2, \dots$$

since  $\Phi_{tt} = O(1)$ . Assume  $\psi_{ts} = O(h)$  for  $s < t$ ,  $t = 2, 3, \dots, q$ . From (3.8) and (3.9)

$$d_{q+1}^* = (1 + h\Phi_{q+1,q+1})d_{q+1} + h \sum_{s=1}^q \Phi_{q+1,s} \left\{ \sum_{u=1}^s \psi_{su} d_u^* \right\}.$$

Hence

$$d_{q+1}^* = (1 + h\Phi_{q+1,q+1})d_{q+1} + h \sum_{u=1}^q \left\{ \sum_{j=u}^q \Phi_{q+1,j} \psi_{ju} \right\} d_u^*.$$

Comparing with  $d_{q+1}^*$  as given by (3.9), we have

$$\psi_{q+1,q+1} = 1/(1 + h\Phi_{q+1,q+1}); \quad \psi_{q+1,u} = h \left( \sum_{j=u}^q \Phi_{q+1,j} \psi_{ju} \right) / (1 + h\Phi_{q+1,q+1}),$$

$$u = 1, 2, \dots, q.$$

Since  $\Phi_{q+1,j} = O(1)$ ,  $\psi_{ju} = O(h)$  for  $u < j \leq q$  by hypothesis, and since  $q \leq n$ , we conclude  $\psi_{q+1,u} = O(h)$ ,  $u = 1, 2, \dots, q$ . Direct computation establishes that  $\psi_{21} = O(h)$  and so, by induction, it follows that  $\psi_{ts} = O(h)$  for  $s < t$  as  $h \rightarrow 0$ ,  $n \rightarrow \infty$ ,  $nh$  fixed.

Since  $t \leq n$  and  $|\zeta_p| \leq 1$  by stability, then, if  $F_p(s-1) = O(h)$  for all  $s$ ,

$$\sum_{s=1}^t \psi_{ts} \zeta_p^{s-1} F_p(s-1) = O(h)$$

and, since  $z_{n-i}$  is bounded as  $n \rightarrow \infty$ ,  $P_p(n) = O(h)$ .

Thus, by choosing  $a_j(x)$  and  $b_j(x)$  such that  $F_p(s-1) = O(h)$  we can control the perturbation in the corresponding fundamental solution to remain  $O(h)$  as  $h \rightarrow 0$ ,  $n \rightarrow \infty$ ,  $nh$  fixed. For a consistent method, one of the roots  $\zeta_p$  will be  $+1$ ; let  $\zeta_1$  be this root. Since weak instability is due to uncontrolled growth of the perturbations of the solutions corresponding to the *other* roots, we are motivated to choose  $a_j(x)$  and  $b_j(x)$  such that

$$(3.13) \quad F_p(s-1) = O(h) \text{ for all } s, \quad p = 2, 3, \dots, k.$$

(It is of interest to note that to achieve  $F_p(s-1) = O(h)$  for all  $s$ ,  $p = 1, 2, 3, \dots, k$ , it turns out that we have to make the very choice for  $a_j(x)$ ,  $b_j(x)$  which we found abortive in (3.2).)

We now take as our specific target the removal of the weak instability associated with optimal stable constant-coefficient methods. Since these occur only when the step-number is even, we consider the cases  $k = 2$ ,  $k = 4$ .

*Case  $k = 2$ .* From (3.13) the appropriate choice for  $a_j(x)$  and  $b_j(x)$  will cause  $F_2(s-1) = O(h)$  for  $s = 1, 2, \dots$ . From (3.3) and (3.11), we see that this can be achieved if we take

$$a_j(x_n) = \hat{a}_j q(x_n), \quad (j = 0, 1, 2)$$

and

$$\hat{a}_0 + \beta_0 + \zeta_2(\hat{a}_1 + \beta_1) + \zeta_2^2(\hat{a}_2 + \beta_2) = 0.$$

If at the same time we seek the highest possible order for the method, we find, taking  $b_j(x_n) = \hat{b}_j q(x_n)$ , ( $j = 0, 1, 2$ )

$$(3.14) \quad \begin{aligned} \alpha_2 &= 1, & \alpha_1 &= 0, & \alpha_0 &= -1, \\ \beta_2 &= \frac{1}{3}, & \beta_1 &= \frac{4}{3}, & \beta_0 &= \frac{1}{3}, \\ \hat{a}_2 &= \mathcal{K}, & \hat{a}_1 &= -\frac{1}{3}, & \hat{a}_0 &= \frac{1}{3} - \mathcal{K}, \\ \hat{b}_2 &= \frac{1}{36} + \frac{\mathcal{K}}{3}, & \hat{b}_1 &= -\frac{2}{9} + \frac{4\mathcal{K}}{3}, & \hat{b}_0 &= -\frac{5}{36} + \frac{\mathcal{K}}{3}. \end{aligned}$$

It is readily seen that the method defined by these coefficients is stable and consistent; it is, indeed, of order four, its principal truncation error being

$$h^5 \left[ \frac{-1}{90} y^{(5)}(x_n) - \frac{1}{72} q(x_n) y^{(4)}(x_n) \right].$$

The parameter  $\mathcal{K}$  is free, and from the form of the principal truncation error it is clear that we cannot utilize this freedom to increase the order beyond four.

In a recent paper Brunner [6], using a different approach, has succeeded in removing the weak instability from optimal stable linear multistep methods in the case when they are applied to specific linear differential equations. In the case  $k = 2$ , Brunner's formula is



$$(3.15) \quad \begin{aligned} &(1 + hL)y_{n+2} - hLy_{n+1} - y_n \\ &= \frac{h}{12} [(4 + 5hL)f_{n+2} + 4(4 + 2hL)f_{n+1} + (4 - hL)f_n], \end{aligned}$$

where  $L$  is a stability parameter. If in (3.14) we take  $\mathcal{K} = \frac{1}{3}$ , and put  $q(x_n) = 3L$  the resulting formula is identical with (3.15). However, in Brunner's formula the stability parameter  $L$  is a constant, which is chosen according to the specific problem. Thus, for a fixed value of  $h$ , the coefficients in Brunner's formula are constant, whereas those of (3.14) will vary with  $q(x_n)$  as the computation proceeds. Thus the two methods coincide only when  $\mathcal{K} = \frac{1}{3}$  and the differential equation is linear with constant coefficients.

Returning to (3.14) we still have the choice for the parameter  $\mathcal{K}$  at our disposal. From (3.3) and (3.11) it is seen that  $F_2(x_n)$  can be made identically zero rather than  $O(h)$ , if we satisfy the additional condition

$$(3.16) \quad \hat{b}_0 + \zeta_2 \hat{b}_1 + \zeta_2^2 \hat{b}_2 = 0.$$

From (3.12) we see that this choice will cause the perturbation in the fundamental solution corresponding to the nonessential root  $\zeta_2$  to be completely annihilated.

We find, from (3.14), that (3.16) is satisfied if  $\mathcal{K}$  is given the value  $1/6$ , and with this value Eqs. (3.14) define the following method:

$$(3.17) \quad \begin{aligned} &\left[ 1 + \frac{1}{6} hq(x_n) \right] y_{n+2} - \frac{1}{3} hq(x_n) y_{n+1} + \left[ -1 + \frac{1}{6} hq(x_n) \right] y_n \\ &= h \left\{ \left[ \frac{1}{3} + \frac{1}{12} hq(x_n) \right] f_{n+2} + \frac{4}{3} f_{n+1} + \left[ \frac{1}{3} - \frac{1}{12} hq(x_n) \right] f_n \right\}. \end{aligned}$$

This may be regarded as a stabilized modification of Simpson's rule.

In passing, it is of interest to observe that if (3.17) is applied to the differential equation  $y' = Qy$ ,  $Q$  constant, the resulting difference equation is

$$\left( 1 - \frac{1}{2} hQ + \frac{1}{12} h^2 Q^2 \right) y_{n+2} - hQ y_{n+1} - \left( 1 + \frac{1}{2} hQ + \frac{1}{12} h^2 Q^2 \right) y_n = 0$$

whose *exact* solution is

$$(3.18) \quad y_n = A_1 \left( \frac{1 + \frac{1}{2} hQ + \frac{1}{12} h^2 Q^2}{1 - \frac{1}{2} hQ + \frac{1}{12} h^2 Q^2} \right)^n + A_2 (-1)^n,$$

where  $A_1$  and  $A_2$  are arbitrary constants. Hence if  $Q < 0$ ,  $y_n \rightarrow A_2(-1)^n$  as  $n \rightarrow \infty$  for any fixed positive  $h$ . Thus, all solutions of the difference equation are bounded as  $n \rightarrow \infty$  and, moreover, if the initial conditions  $y_0 = C$ ,  $y_1 = Ce^{Qh} + O(h^5)$  are satisfied, the bound will be of order  $h^5$ . This convergence property, besides corroborating that (3.14) does not suffer the weak instability that Simpson's rule would when applied to the equation  $y' = Qy$ ,  $Q < 0$ , is close to the concept of  $A$ -stability defined by Dahlquist [7];  $A$ -stability, however, is defined in terms of a system of linear differential equations, and requires that all solutions of the resulting system of difference equations should not only be bounded, but should tend to zero as  $n \rightarrow \infty$  for any fixed positive value of  $h$ . Dahlquist has shown that the most accurate linear multistep method which is  $A$ -stable is the Trapezoidal Rule; if this

method is applied to the equation  $y' = Qy$  the exact solution of the resulting difference equation is

$$y_n = A_1 \left( \frac{1 + \frac{1}{2} hQ}{1 - \frac{1}{2} hQ} \right)^n.$$

The expressions

$$\frac{1 + \frac{1}{2} hQ}{1 - \frac{1}{2} hQ} \quad \text{and} \quad \frac{1 + \frac{1}{2} hQ + \frac{1}{12} h^2 Q^2}{1 - \frac{1}{2} hQ + \frac{1}{12} h^2 Q^2}$$

are both Padé approximations to  $e^{hQ}$ . The price that has to be paid for the more accurate approximation employed in (3.18) is the introduction of the spurious second solution  $A_2(-1)^n$ ; the spurious solution is, however, bounded as  $n \rightarrow \infty$ , for all fixed positive  $h$ , and by appropriate choice of initial conditions the bound can be made of the order of the local truncation error.

Case  $k = 4$ . By a similar analysis, whose details we omit, we can find a set of coefficients for which

- (i) the roots of the polynomial  $\sum_{j=0}^4 \alpha_j \zeta^j$  satisfy the condition of stability;
- (ii)  $F_p(s-1) = O(h)$  for all  $s$ ,  $p = 2, 3, 4$ ; and
- (iii) the method has order six.

These coefficients, which correspond to (3.14) in the case of  $k = 2$ , are functions of *two* free parameters, and it is not possible to satisfy the three additional conditions which would make  $F_p(s-1) \equiv 0$ . Thus for  $k = 4$ , there exists no formula analogous to (3.17). The underlying reason for this is that in the case  $k = 2$  it happens that the constant-coefficient method of maximal order, namely Simpson's rule, is stable. This is not so in the case  $k = 4$ , where the method of maximal order is unstable. Thus, in demanding that the roots of  $\rho(\zeta)$  satisfy the condition of stability, we have effectively used up parameters which would otherwise have been at our disposal.

However, the set of coefficients does define a class of formulae analogous to that defined by (3.14). Choosing the two free parameters to give the least cumbersome values for the coefficients yields the following formula:

$$\begin{aligned} & \left[ 1 + \frac{17}{90} hq(x_n) \right] y_{n+4} - \frac{19}{45} hq(x_n) y_{n+3} + \frac{7}{15} hq(x_n) y_{n+2} - \frac{19}{45} hq(x_n) y_{n+1} \\ & \qquad \qquad \qquad + \left[ -1 + \frac{17}{90} hq(x_n) \right] y_n \\ (3.23) \quad & = h \left\{ \left[ \frac{14}{45} + \frac{29}{360} hq(x_n) \right] f_{n+4} + \left[ \frac{64}{45} + \frac{1}{180} hq(x_n) \right] f_{n+3} + \frac{8}{15} f_{n+2} \right. \\ & \qquad \qquad \qquad \left. + \left[ \frac{64}{45} - \frac{1}{180} hq(x_n) \right] f_{n+1} + \left[ \frac{14}{45} - \frac{29}{360} hq(x_n) \right] f_n \right\}. \end{aligned}$$

The principal truncation error of this formula is

$$h^7 \left[ \frac{-8}{945} y^{(7)}(x_n) - \frac{23}{408} q(x_n) y^{(6)}(x_n) \right].$$

**4. Numerical Example.** We consider the initial-value problem

$$y' = -10(y - 1)^2; \quad y(0) = 2,$$

which we integrate over the range [0, 5]. The theoretical solution is

$$y = 1 + \frac{1}{10x + 1}.$$

Thus  $\partial f/\partial y$  is negative throughout the range of integration, and we expect Simpson's rule to exhibit weak instability. The equation is integrated with  $h = 0.1$ , first using Simpson's rule, and then using (3.17), with  $q(x_n) = -20(y_n - 1)$ . The results are shown in Table 1, in which starting values are underlined.

TABLE 1

$x$	<i>Theoretical Solution</i>	<i>Simpson's rule</i>	<i>Formula (3.17)</i>
0.0	2.000 000	<u>2.000 000</u>	<u>2.000 000</u>
0.1	1.500 000	<u>1.500 000</u>	<u>1.500 000</u>
0.2	1.333 333	1.302 776	1.333 333
0.3	1.250 000	1.270 115	1.249 579
0.4	1.200 000	1.165 775	1.200 050
.	.	.	.
.	.	.	.
3.8	1.025 641	0.867 153	1.025 815
3.9	1.025 000	0.953 325	1.024 819
4.0	1.024 390	0.850 962	1.024 565
.	.	.	.
.	.	.	.
4.8	1.020 408	0.040 686	1.020 583
4.9	1.020 000	-5.990 968	1.019 820
5.0	1.019 608	-394.086	1.019 782

*Appendix. Proof of Theorem 1.* Using the notation and partial results of Henrici's proof, we have

$$\tilde{L}[y(x_m); x_m; h] = L[y(x_m); h] + hM[y(x_m); x_m; h],$$

where

$$|L[y(x_m); h]| \leq K_{\chi}(kh)h$$

and

$$\begin{aligned} M[y(x_m); x_m; h] = & [a_0(x_m) + a_1(x_m) + \cdots + a_k(x_m)]y(x_m) \\ & + [a_1(x_m) + 2a_2(x_m) + \cdots + ka_k(x_m)]y'(x_m)h \\ & + \phi' [|a_1(x_m)| + 2|a_2(x_m)| + \cdots + k|a_k(x_m)|] \chi(kh)h \\ & - [b_0(x_m) + b_1(x_m) + \cdots + b_k(x_m)]y'(x_m)h \\ & - \phi [|b_0(x_m)| + |b_1(x_m)| + \cdots + |b_k(x_m)|] \chi(kh)h, \end{aligned}$$

where  $|\phi| \leq 1$ ,  $|\phi'| \leq 1$ .

For all  $x \in [a, b]$ , we have

(i)  $|a_j(x)| \leq A$ ,  $|b_j(x)| \leq B$ , ( $j = 0, 1, \dots, k$ ), by (2.1),

(ii)  $a_0(x) + a_1(x) + \dots + a_k(x) = 0$ , by consistency, and

(iii)  $|y'(x)| \leq F$ , where  $F$  is a finite constant, since  $y'(x) = f(x, y(x))$ , being a continuous function of  $x$  on the closed interval  $[a, b]$ , is necessarily bounded on  $[a, b]$ .

Thus  $|M[y(x_m); x_m; h]| \leq (k + 1)h(\frac{1}{2}kA + B)[F + \chi(kh)]$ , and hence  $|\tilde{L}[y(x_m); x_m; h]| \leq K'\chi(kh)h + Gh^2$ , where

$$K' = K + (k + 1)h(\frac{1}{2}kA + B), \quad G = (k + 1)F(\frac{1}{2}kA + B).$$

On subtracting  $\tilde{L}[y(x_m); x_m; h]$  from the corresponding relation

$$\tilde{\alpha}_k(x_m)y_{m+k} + \dots + \tilde{\alpha}_0(x_m)y_m - h[\tilde{\beta}_k(x_m)f_{m+k} + \dots + \tilde{\beta}_0(x_m)f_m] = 0$$

we obtain

$$(A.1) \quad \tilde{\alpha}_k(x_m)e_{m+k} + \dots + \tilde{\alpha}_0(x_m)e_m - h[\tilde{\beta}_k(x_m)g_{m+k}e_{m+k} + \dots + \tilde{\beta}_0(x_m)g_me_m] \\ = \theta_m[K'\chi(kh)h + Gh^2],$$

where  $e_m = y_m - y(x_m)$  and

$$g_m = \begin{cases} [f(x_m, y_m) - f(x_m, y(x_m))]e_m^{-1}, & e_m \neq 0, \\ 0, & e_m = 0 \end{cases}$$

and where  $|\theta_m| \leq 1$ .

On rewriting (A.1) in the form

$$(\alpha_k e_{m+k} + \dots + \alpha_0 e_m) \\ - h\{[\tilde{\beta}_k(x_m)g_{m+k} - a_k(x_m)]e_{m+k} + \dots + [\tilde{\beta}_0(x_m)g_m - a_0(x_m)]e_m\} \\ = \theta_m[K'\chi(kh)h + Gh^2],$$

we see that Henrici's Lemma 5.6 applies with

$$\Lambda = [K'\chi(kh) + Gh]h$$

and

$$B^* = \mathfrak{L}[|\beta_0| + \dots + |\beta_k| + h(k + 1)B] + (k + 1)A,$$

where  $\mathfrak{L}$  is the Lipschitz constant of  $f(x, y)$ . Since  $K'\chi(kh) + Gh \rightarrow 0$  as  $h \rightarrow 0$ , the proof is concluded exactly as in Henrici.

**Acknowledgement.** The author wishes to acknowledge the extremely useful comments made by the referees.

Department of Mathematics  
University of Dundee  
Dundee, Scotland

2. P. HENRICI, *Discrete Variable Methods in Ordinary Differential Equations*, Wiley, New York, 1962. MR 24 #B1772.
3. T. E. HULL & A. C. R. NEWBERRY, "Corrector formulas for multi-step integration methods," *J. Soc. Indust. Appl. Math.*, v. 10, 1962, pp. 351-369. MR 27 #2130.
4. A. RALSTON, "Relative stability in the numerical solution of ordinary differential equations," *SIAM Rev.*, v. 7, 1965, pp. 114-125. MR 31 #2831.
5. R. W. HAMMING, "Stable predictor-corrector methods for ordinary differential equations," *J. Assoc. Comput. Mach.*, v. 6, 1959, pp. 37-47. MR 21 #973.
6. H. BRUNNER, "Stabilization of optimal difference operators," *Z. Angew. Math. Phys.*, v. 18, 1967, pp. 438-444. MR 36 #1111.
7. G. DAHLQUIST, "A special stability problem for linear multistep methods," *Nordisk Tidskr. Informations-Behandling*, v. 3, 1963, pp. 27-43. MR 30 #715.