

A Note on Modified Optimal Linear Multistep Methods*

By H. Brunner

Abstract. Modified optimal linear k -step methods (whose coefficients depend on the stepsize and on a parameter L) are used for the numerical integration of systems of nonlinear ordinary differential equations. It is shown that, by choosing L suitably (depending essentially on the growth parameters of the k -step method and on the logarithmic norm of the Jacobian of the given system), weak stability does no longer occur, and one of two types of stability (called asymptotical relative and asymptotical absolute stability) may be obtained.

1. Introduction. In [1], the author discussed a class of linear k -step methods (R, S) of the form

$$(1.1) \quad R(E, hL) Y_n = hS(E, hL) F_n,$$

where $EY_n = Y_{n+1}$, $x_n = nh$, $h > 0$, $F_n = F(x_n, Y_n)$,

$$(1.2) \quad R(w, hL) = \sum_{\nu=0}^k \tilde{\alpha}_\nu(hL) \cdot w^\nu = \rho(w) + \frac{1}{2}hL \cdot \rho^*(w),$$

$$(1.3) \quad S(w, hL) = \sum_{\nu=0}^k \tilde{\beta}_\nu(hL) \cdot w^\nu = \sigma(w) + \frac{1}{2}hL \cdot \sigma^*(w).$$

Here, L is a (nonnegative) parameter to be specified later. The polynomials

$$\rho(w) = \sum_{\nu=0}^k \alpha_\nu w^\nu, \quad \sigma(w) = \sum_{\nu=0}^k \beta_\nu w^\nu \quad (\alpha_\nu, \beta_\nu \text{ real})$$

are such that the corresponding linear k -step method (ρ, σ) is stable (in the sense of Dahlquist) and has (optimal) order $p = k + 2$ (see [6, p. 232]). The k -step method defined by the polynomials

$$(\rho^*, \sigma^*) \quad (\text{where } \rho^*(w) \equiv \rho'(w) \cdot (w - 1), \quad ' = d/dw)$$

has order $p^* \geq k + 1$ ([1], [6]).

The linear k -step method (R, S) , given by (1.1), is used to generate (for given initial values $\{Y_0, \dots, Y_{k-1}\}$) approximations $\{Y_k, \dots, Y_N\}$ to the exact solution $Y(x)$ of the initial-value problem

$$(1.4) \quad Y'(x) = F(x, Y(x)), \quad Y(0) = Y_0, \quad 0 \leq x \leq a,$$

at the points $x = x_n = nh$, $n = k, \dots, N$, $Nh = a$. Here, $Y \in R^m$, $m \geq 1$.

Received September 3, 1971.

AMS 1970 subject classifications. Primary 65L05.

Key words and phrases. System of nonlinear ordinary differential equations, optimal linear multistep methods, weak stability, logarithmic norm of the Jacobian, asymptotical relative and absolute stability.

* This research was supported by the National Research Council of Canada (Grant No. A4805).

Copyright © 1972, American Mathematical Society

It has been known for some time (see, for example, [3]) that optimal linear k -step methods (ρ, σ) as described above are subject to weak (or marginal) stability. The author has shown in [1] that, for the case $m = 1$, the modified k -step method (R, S) does no longer show weak stability if L is chosen suitably. In [2], these results were extended to the case $m > 1$. The purpose of this note is to simplify the results of [2] and to discuss their practical application, especially when the given system (1.4) is nonlinear. In this case, one is forced to use, in general, a "variable" L , i.e., L will be kept constant over a certain number of integration steps in order to avoid the evaluation of the Jacobian of the system at each step.

2. Elimination of Weak Stability. Let $G(x)$ be the Jacobian of $F(x, Y)$ in (1.4) along the exact solution $Y(x)$. Assume that $Y(x) \in C^{k+4}[0, a]$ (where k denotes the degree of the characteristic polynomials of the method (R, S) , $k \geq 2$). It was shown in [2] that under these conditions the discretization error $E_n = Y_n - Y(x_n)$ satisfies the asymptotic relation (see also [7, pp. 25–26])

$$E_n = E_n^H + E(x_n)h^{k+2} + O(h^{k+3}), \quad 0 \leq x_n \leq a,$$

$h \rightarrow 0$, $nh = x_n$ fixed, L fixed. Here (see [2, Theorem 3.1]), $E(x)$ is the solution of a system of linear differential equations involving $G(x)$ and the error constants of the methods (ρ, σ) and (ρ^*, σ^*) (but not the growth parameters of the method (R, S)). We have assumed that the initial errors satisfy $E_i = O(h^{k+2})$, $i = 0, 1, \dots, k-1$. The error term E_n^H is the solution of the homogeneous difference equation for E_n with given initial values E_i , $i = 0, 1, \dots, k-1$, and we have

$$(2.1) \quad \|E_n^H\| \leq \sum_{i=1}^k r_i(x_n, L) \cdot \|A_i\| + O(h^{k+3}), \quad 0 \leq x_n \leq a.$$

The vectors A_i are determined from the initial errors E_i , and one has $\|A_i\| = O(h^{k+2})$. The functions $r_i(x, L)$ are given by [2]

$$(2.2) \quad r_j(x, L) = \exp\left(\frac{L}{2}(\cos \varphi_j - 1)x + \int_0^x \mu_N[\lambda_j G(t)] dt\right), \quad j = 1, \dots, k.$$

Here, the quantities λ_j are the growth parameters of the method (ρ, σ) and are defined by [3]

$$(2.3) \quad \lambda_j = \sigma(z_j)/z_j \cdot \rho'(z_j), \quad j = 1, \dots, k,$$

(with $\rho(z_j) = 0$, $z_j = \exp(\varphi_j)$, $0 = \varphi_1 < \varphi_2 < \dots < \varphi_k < 2\pi$). The symbol μ_N denotes the logarithmic norm corresponding to a matrix norm N (induced by a given vector norm $\|\cdot\|$). Properties of the logarithmic norm may be found in [3, pp. 10–12].

Definition 1. The linear k -step method (R, S) is called (asymptotically) absolutely stable with respect to the system (1.4), if for $j = 2, \dots, k$ and for all $x_n \in [x_k, a]$,

$$(2.4) \quad r_j(x_n, L) \leq 1.$$

Definition 2. The linear k -step method (R, S) is called (asymptotically) relatively stable with respect to the system (1.4), if for $j = 2, \dots, k$ and for all $x_n \in [x_k, a]$,

$$(2.5) \quad r_j(x_n, L) < r_1(x_n, L) = r_1(x_n).$$

Before stating the easily derived criteria for the choice of the parameter L to achieve one of these types of stability, we recall some relevant properties of the growth parameters of the (optimal) method (ρ, σ) (see also [3, p. 40]).

(i) The growth parameters λ_j are real, $\lambda_1 = 1$.

(ii) $\lambda_j = \lambda_{k+2-j}$, $j = 2, \dots, k/2$, $\lambda_{k/2+1} \leq -\frac{1}{3}$ (with equality if and only if $k = 2$).

(iii) $\text{sign}(\lambda_j) = (-1)^{j-k/2}$, $j = 2, \dots, k/2 + 1$.

Define the set J_k by

$$J_k = \{2, 4, \dots, k/2 + 1\} \quad \text{if } k/2 \text{ is odd,}$$

$$= \{3, 5, \dots, k/2 + 1\} \quad \text{if } k/2 \text{ is even.}$$

For the following discussion, we shall assume that the spectral abscissa of $G(x)$ is negative, i.e.,

$$\omega(x) \equiv \max_{(i)} \text{Re } \omega_i(G(x)) < 0, \quad 0 \leq x \leq a,$$

where $\omega_i(G(x))$ denotes the i th eigenvalue of $G(x)$.

The following results are now obvious.

THEOREM 1. *The linear k -step method (R, S) is (asymptotically) absolutely stable with respect to the system (1.4), if the parameter L satisfies for all $x_n \in [x_k, a]$ the inequality*

$$(2.6) \quad L \geq \frac{2}{x_n} \cdot \max_{i \in J_k} \frac{-\lambda_i \cdot \int_0^{x_n} \mu_N[-G(t)] dt}{1 - \cos \varphi_i}.$$

Observe that the set J_k consists, for $k \leq 4$, of only one element: $J_k = \{2\}$ if $k = 2$, $J_k = \{3\}$ if $k = 4$.

THEOREM 2. *The linear k -step method (R, S) is (asymptotically) relatively stable with respect to the system (1.4), if L satisfies, for all $x_n \in [x_k, a]$, the inequality*

$$(2.7) \quad L > \frac{2}{x_n} \cdot \max_{i \in \{2, \dots, k/2+1\}} \frac{\int_0^{x_n} (\mu_N[\lambda_i G(t)] - \mu_N[G(t)]) dt}{1 - \cos \varphi_i}.$$

It follows from the assumption $\omega(x) < 0$, $0 \leq x \leq a$, that $\mu_N[-G(x)] > 0$, $0 \leq x \leq a$, for any matrix norm N , since

$$\begin{aligned} \mu_N[-G(x)] &\geq \max_{(i)} \text{Re } \omega_i(-G(x)) = -\min_{(i)} \text{Re } \omega_i(G(x)) \\ &\geq -\max_{(i)} \text{Re } \omega_i(G(x)) = -\omega(x) > 0. \end{aligned}$$

(On the other hand, $\omega(x) < 0$ does not imply in general that $\mu_N[G(x)] < 0$. But it can be shown that there exists a matrix norm N such that the corresponding logarithmic norm of $G(x)$ is arbitrarily close to the spectral abscissa $\omega(x)$.)

If, in (2.7), $\lambda_i < 0$ and if N is such that $\mu_N[G(x)] < 0$, $0 \leq x \leq a$, then $\mu_N[\lambda_i G(t)] - \mu_N[G(t)] > 0$.

But, in this case, we have also to consider those values of j for which $0 < \lambda_j < 1$, since

$$\mu_N[\lambda_i G(t)] - \mu_N[G(t)] = (\lambda_i - 1)\mu_N[G(t)] > 0$$

(see also the remark in [4, p. 155]).

3. Practical Application. If $F(x, Y)$ in (1.4) is a nonlinear function of Y , then the Jacobian $G(x)$ will not be known. It can be computed approximately, simultaneously with Y_n , at the points $\{(x_n, Y_n): n = 0, 1, \dots, N\}$. We set

$$\tilde{G}_n \equiv \partial F(x, Y) / \partial Y |_{(x_n, Y_n)}.$$

The integral in (2.6) (or (2.7)) will therefore have to be replaced by a discrete expression of the form

$$(3.1) \quad I_n(h) \equiv h \sum_{i=0}^n w_{n,i} \cdot \mu_N[-\tilde{G}_i].$$

An obvious choice is $w_{n,0} = w_{n,n-1} = 1/2$, $w_{n,i} = 1$ ($0 < i < n$). Since (2.6) (or (2.7)), with the integral replaced by (3.1), can, in general, no longer be satisfied a priori over the entire interval $[x_k, a]$, one will be forced to use a variable L . To avoid the evaluation of \tilde{G}_n for each $n = k - 1, \dots, N - 1$, the following procedure may be chosen. Consider the intervals $\{[x_{n_i}, x_{n_{i+1}}]: i = 0, \dots, l\}$, with $n_0 = k - 1, n_{l+1} = N$. The approximate Jacobian \tilde{G}_n will be computed only at the points $\{(x_{n_i}, Y_{n_i}): i = 0, \dots, l\}$. For $n_i < n \leq n_{i+1}$, we set, using (2.6),

$$(3.2) \quad L = L_n^{(i)} \geq \frac{2}{x_n} \cdot \max_{i \in J_k} \frac{-\lambda_j \cdot T_n^{(i)}(h)}{1 - \cos \varphi_j} \quad (i = 0, \dots, l),$$

where

$$T_n^{(i)}(h) = T^{(i-1)}(h) + h \sum_{\nu=n_i}^n w_{n,\nu} \cdot \mu_N[-\tilde{G}_{\nu}],$$

and

$$T^{(0)}(h) = I_{n_0}(h), \quad T^{(i)}(h) = T_{n_{i+1}}^{(i)}(h), \quad i = 1, \dots, l.$$

For practical purposes, it is convenient to write (3.2) in the form

$$(3.3) \quad L_n^{(i)} = \frac{2}{x_n} \cdot \max_{i \in J_k} \frac{-\lambda_j \cdot T_n^{(i)}(h)}{1 - \cos \varphi_j} + \delta_n^{(i)},$$

where $\delta_n^{(i)}$ takes the role of a parameter.

4. Numerical Illustration. We shall apply the above results to a system of moderately stiff nonlinear differential equations suggested by Gear [5, p. 218], namely

$$(4.1) \quad Y'(x) = F(x, Y(x)) \equiv -B \cdot Y(x) + U \cdot W(x),$$

where the matrix

$$U = \frac{1}{2} \begin{pmatrix} -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{pmatrix}$$

is unitary, $U^{-1} = U$, $B = U \cdot \text{diag}(b_1, b_2, b_3, b_4) \cdot U$ (with $b_i \neq 0$),

$$W(x) = (z_1^2(x), \dots, z_4^2(x))^T, \quad Z(x) = (z_1(x), \dots, z_4(x))^T = U \cdot Y(x).$$

The exact solution of (4.1), corresponding to the initial vector

$$Y(0) = (-1, -1, -1, -1)^T,$$

is given by $z_i(x) = b_i / (1 + c_i \exp(b_i x))$, $c_i = -(1 + b_i)$, $i = 1, \dots, 4$, and by using $Y(x) = UZ(x)$.

The eigenvalues of the Jacobian $G(x)$ of (4.1) along the exact solution are $\Lambda_i(x) = 2z_i(x) - b_i$, with $\lim_{x \rightarrow \infty} \Lambda_i(x) = -|b_i|$, $i = 1, \dots, 4$. System (4.1) was solved numerically by the modified Milne-Simpson method,

$$(4.2) \quad Y_{n+1} \cdot (1 + hL) - hL \cdot Y_n - Y_{n-1} = \frac{h}{12} [(4 + 5hL)F_{n+1} + (16 + 8hL)F_n + (4 - hL)F_{n-1}].$$

We chose

$$b_1 = \begin{Bmatrix} 40 \\ 50 \end{Bmatrix}, \quad b_2 = -0.1, \quad b_3 = 5, \quad b_4 = 0.001,$$

and (4.2) was applied with a step size $h = 0.1$. Tables 1 and 2 below contain a selection of numerical results. The logarithmic norm used in (2.6) and (3.1) was the one corresponding to the maximum norm.

TABLE 1. $b_1 = 40, h = 0.1$

| n | x_n | I | $\ E_n\ _\infty = \max_{(i)} e_{i,n} $ | III |
|-----|-------|-------------------------------|---|----------------------|
| | | | II | |
| 2 | 0.2 | 0.03782660 | 0.03504106 | 0.03504106 |
| | | ($L = 54.0 = \text{const}$) | ($L = 47.031$) | ($L = 47.031$) |
| 3 | 0.3 | 0.03171098 | 0.02994904 | 0.02994904 |
| . | | | ($L = 49.395$) | ($L = 49.395$) |
| . | | | | |
| 99 | 9.9 | 0.00020415 | 0.00021375 | 0.00023867 |
| | | | ($L = 53.981$) | ($L = 53.419$) |
| 100 | 10.0 | 0.00019380 | 0.00020292 | 0.00022479 |
| . | | | ($L = 53.983$) | ($L = 53.419$) |
| . | | | | |
| 500 | 50.0 | $7.23 \cdot 10^{-8}$ | $7.12 \cdot 10^{-8}$ | $7.11 \cdot 10^{-8}$ |
| | | | ($L = 54.095$) | ($L = 53.390$) |

Exact solution at $x = 50$:

$$Y(x) = (-0.05986120, 0.04074892, -0.05986120, -0.04074892)^T.$$

Column I: $L = \text{const} = \frac{1}{3} \cdot \mu_N[-\tilde{G}_0] + 40$ (using (2.6) with constant Jacobian).

Column II: $\mu_N[-\tilde{G}_n] = \mu[-\tilde{G}_1]$, $n \geq 2$ (using (3.3), trapezoidal rule, $\delta_n^{(i)} = 40$).

Column III: $n_{i+1} - n_i = 10$ (using (3.3), trapezoidal rule, $\delta_n^{(i)} = 40$).

TABLE 2. $b_1 = 50, h = 0.1$

| n | x_n | $\ E_n\ _\infty = \max_{(i)} e_{i,n} $ | | |
|-----|-------|---|---|---|
| | | I | II | III |
| 2 | 0.2 | 0.05167039 ($L = 100.33 = \text{const}$) | 0.05795966 ($L = 167.33 = \text{const}$) | 0.05743187 ($L = 158.699$) |
| 3 | 0.3 | 0.04597920 | 0.04870244 | 0.04841416 ($L = 161.619$) |
| . | . | | | |
| 99 | 9.9 | 0.04382924 | 0.00129796 | 0.00131064 ($L = 167.284$) |
| 100 | 10.0 | 0.04366285 | 0.00124998 | 0.00126221 ($L = 167.286$) |
| . | . | | | |
| 500 | 50.0 | 0.03869461 | $9.08 \cdot 10^{-8}$ | $9.06 \cdot 10^{-8}$ ($L = 167.426$) |

Column I: $L = \text{const} = \frac{1}{3}\mu_N[-\tilde{G}_0] + 83.0$ (using (2.6) with constant Jacobian).
 Column II: $L = \text{const} = \frac{1}{3}\mu_N[-\tilde{G}_0] + 150.0$ (using (2.6) with constant Jacobian).
 Column III: $\mu_N[-\tilde{G}_n] = \mu_N[-\tilde{G}_1], n \geq 2$ (using (3.3), trapezoidal rule, $\delta_n^{(i)} = 150.0$).

All the computations were performed on the CDC 6400 (single precision) at Dalhousie University Computer Centre.

As a first remark we note that the trace of the approximate negative Jacobian $-\tilde{G}_n$ of the system under consideration is equal to the sum of its eigenvalues. Hence, the logarithmic norm $\mu_N[-\tilde{G}_n]$ needs to be updated only if $\text{tr}(-\tilde{G}_n)$ increases by a significant amount over a number of integration steps. This remark becomes relevant when solving, for example, a nonlinear system of the form

$$y_1'(x) = -\frac{1}{2}(1+x)^{10} \cdot (1-y_1(x))^2 \cdot y_2(x), \quad y_1(0) = 3,$$

$$y_2'(x) = -\frac{10y_2(x)}{1+x} \cdot \left(y_1(x) - \frac{2}{1+x} \right), \quad y_2(0) = 1,$$

whose exact solution is $y_1(x) = 1 + 2/(1+x), y_2(x) = 1/(1+x)^{10}$. It is obvious that in this case the off-diagonal elements (and thus the logarithmic norm corresponding to the L_1 -norm or the L_∞ -norm) vary rapidly as x increases, whereas its trace decreases slowly from 12 to 0 as $x \rightarrow \infty$.

Table 2 also shows that the parameter L has to be chosen much larger than indicated by the right-hand side of (2.6), confirming the fact that the results of Section 2 have been derived for the asymptotic case $h \rightarrow 0$. In the following table, we list (for the modified Milne-Simpson method (4.2)) the smallest value L for which both roots, $w_1(hL)$ and $w_2(hL)$, of the characteristic equation $R(w, hL) + hq \cdot S(w, hL) = 0$ lie no longer outside the unit circle. The agreement between the results of this table and the results in Columns I and II of Table 2 is obvious.

| qh | L | $w_1(hL)$ | $w_2(hL)$ |
|------|--------|-----------|-----------|
| 1.0 | 4.0 | 0.36842 | -1.00000 |
| | 4.2 | 0.36852 | -0.98734 |
| 2.0 | 10.0 | 0.14286 | -1.00000 |
| | 10.2 | 0.14326 | -0.99340 |
| 3.0 | 20.0 | 0.07692 | -1.00000 |
| | 20.2 | 0.07783 | -0.99716 |
| 4.0 | 40.0 | 0.07692 | -1.00000 |
| | 40.2 | 0.07719 | -0.99905 |
| 5.0 | 100.0 | 0.10448 | -1.00000 |
| | 100.2 | 0.10455 | -0.99982 |
| 5.5 | 220.0 | 0.12292 | -1.00000 |
| 6.0 | 10^7 | 0.14286 | -1.000001 |

Acknowledgement. The author would like to thank the British Science Research Council for the support received during the visit with the Department of Mathematics, University of Dundee, July and August 1971, where part of this work was done.

Department of Mathematics
Dalhousie University
Halifax, Nova Scotia, Canada

1. H. BRUNNER, "Marginal stability and stabilization in the numerical integration of ordinary differential equations," *Math. Comp.*, v. 24, 1970, pp. 635-646. MR 42 #8697.
2. H. BRUNNER, "Stark stabile optimale k -Schrittverfahren," *Meth. Verf. Math. Phys.*, v. 4, 1970, pp. 140-163.
3. G. DAHLQUIST, "Stability and error bounds in the numerical integration of ordinary differential equations," *Kungl. Tekn. Högsk. Handl. Stockholm.*, no. 130, 1959, 87 pp. MR 21 #1706.
4. G. DAHLQUIST, "Stability questions for some numerical methods for ordinary differential equations," *Proc. Sympos. Appl. Math.*, vol. 15, Amer. Math. Soc., Providence, R.I., 1963, pp. 147-158. MR 28 #4679.
5. C. W. GEAR, *Numerical Initial Value Problems in Ordinary Differential Equations*, Prentice-Hall, Englewood Cliffs, N.J., 1971.
6. P. HENRICI, *Discrete Variable Methods in Ordinary Differential Equations*, Wiley, New York, 1962. MR 24 #B1772.
7. P. HENRICI, *Error Propagation for Difference Methods*, Wiley, New York, 1963. MR 27 #4365.