Asymptotic ($h \to \infty$) Absolute Stability for BDFs Applied to Stiff Differential Equations

FRED T. KROGH and KRIS STEWART
California Institute of Technology

Methods based on backward differentiation formulas (BDFs) for solving stiff differential equations require iterating to approximate the solution of the corrector equation on each step. One hope for reducing the cost of this is to make do with iteration matrices that are known to have errors and to do no more iterations than are necessary to maintain the *stability* of the method. This paper, following work by Klopfenstein, examines the effect of errors in the iteration matrix on the stability of the method. Application of the results to an algorithm is discussed briefly.

Categories and Subject Descriptors: G.1.7 [**Numerical Analysis**]: Ordinary Differential Equations—*stiff equations*

General Terms: Algorithms, Theory

Additional Key Words and Phrases: Stability, BDF methods

1. INTRODUCTION

Klopfenstein [2] introduces the concept of asymptotic ($h \to \infty$) absolute stability, and computes asymptotic stability regions for a family of backward differentiation formulas (BDFs). In this paper, we give a simplified derivation of his results and compute stability regions using a slightly different characterization of asymptotic stability which we believe is slightly better for use in a code.

The analysis is for the linear constant coefficient differential equation

$$\mathbf{y}' = \frac{d\mathbf{y}}{dt} = J\mathbf{y}, \text{ where } J \text{ is a matrix,} \tag{1}$$

with the understanding that we will use the analysis to guide us in the solution of general nonlinear differential equations. We write the family of BDFs as

The research described in this paper was carried out at the Jet Propulsion Laboratory at the California Institute of Technology under contract with the National Aeronautics and Space Administration.
Authors' addresses: F. T. Krogh, Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109; K. Stewart, Dept. of Computer Science, Univ. of New Mexico, Albuquerque, NM 87131.

follows

$$\begin{aligned}
\mathbf{p}_{n+1} &= \sum_{r=0}^{k} \nabla^r \mathbf{y}_n, \\
\mathbf{p}'_{n+1} &= \frac{1}{h} \sum_{r=1}^{k} d_r \nabla^r \mathbf{y}_n, \\
\mathbf{y}_{n+1}^{(j+1)} &= \mathbf{y}_{n+1}^{(j)} + \delta \mathbf{y}_{n+1}^{(j)}, \qquad \mathbf{y}_{n+1}^{(0)} = \mathbf{p}_{n+1}, \\
\hat{G}\delta \mathbf{y}_{n+1}^{(j)} &= -\mathbf{r}_{n+1}^{(j)} = -[J\mathbf{y}_{n+1}^{(j)} - \mathbf{p}'_{n+1} - \alpha(\mathbf{y}_{n+1}^{(j)} - \mathbf{p}_{n+1})], \\
\mathbf{y}_{n+1} &= \mathbf{y}_{n+1}^{(m)},
\end{aligned} \tag{2}$$

where h is the step size, ∇ is the backward difference operator ($\nabla y_n = \mathbf{y}_n - \mathbf{y}_{n-1}$), the d_r are coefficients which give a correct value for the derivative of a polynomial of degree k, α is a free parameter, and $\hat{G}$ is an approximation to

$$G = \frac{\partial \mathbf{r}_{n+1}^{(j)}}{\partial \mathbf{y}_{n+1}^{(j)}} = J - \alpha I.$$

(When solving a more general differential equation, the $J\mathbf{y}_{n+1}^{(j)}$ is replaced by the expression for the derivative evaluated at $(t_{n+1}, \mathbf{y}_{n+1}^{(j)})$.)

When a new factorization of the iteration matrix is needed we form

$$G_f = \hat{J} - \hat{\alpha} I = LU, \tag{3}$$

where $\hat{J}$ is an approximation to J, L and U are lower and upper triangular matrices, and the iteration matrix coefficient $\hat{\alpha}$ may either be the current value of α or some estimate for a future value of α. The iteration matrix is

$$\hat{G} = cG_f, \tag{4}$$

where c is a scalar selected to minimize the effect that the difference $\alpha - \hat{\alpha}$ has on the iteration. (Note that if G_f is written $(1/\hat{\alpha}\ \hat{J} - I)$, as it sometimes is, then defining $\hat{G} = (\hat{\alpha}/\alpha\ c)G_f$, where the c is used as in eq. (4), gives an equivalent method. Thus if c is not introduced the two approaches give different results. But with c, one can get identical results, and presumably the form we have used for G_f is to be preferred since it requires less work to form the matrix.) We also define the absolute error matrix

$$\Delta = G - \hat{G} = J - c\hat{J} - \alpha I + c\hat{\alpha} I. \tag{5}$$

Methods based on BDFs can be partially characterized by three attributes: (A) the formulas used when the stepsize is constant, (B) the way the method is modified when the stepsize is varied, and (C) the criteria used for terminating the corrector iteration.

For attribute A, where the stepsize is constant, we have $\hat{\alpha} = \alpha$, and can choose α with different objectives: (1) to optimize stability characteristics as in [2]; (2) to give a corrector with order one greater than the predictor, in which case $\alpha = d_{k+1}/h$; (3) to give a corrector with the same order as the predictor, in which case $\alpha = d_k/h$; or (4) to do something else. Any reasonable choice has the characteristic that $\alpha \to 0$ as $h \to \infty$.

Although our stability analysis is for a constant h, we are really interested in a variable step implementation using modified divided differences as described in Krogh [3]. Thus when considering attribute B the coefficients d_r change for k steps after a change in h, and one can adjust the method in one of three ways (in all cases $\hat{\alpha}$ is fixed until a new iteration matrix is formed): (1) using the constant step formula chosen in attribute A and interpolating in the stored data whenever the step size changes; (2) trying to model the constant step formula chosen in attribute A by replacing the constant step coefficients used there with the corresponding variable step coefficients; or (3) using a variable step formula to model the formula chosen in attribute A but modifying the coefficient of the highest order difference in the corrector to reduce errors in the iteration matrix as in [1] by setting $\alpha = \hat{\alpha}$ (frequently called a fixed leading coefficient method). Note: if the second choice is made, then one can compensate when $\hat{\alpha} \neq \alpha$ by modifying the iteration matrix.

With regard to attribute C, one can terminate the corrector iteration: (1) when it "converges" in some sense; or (2) when some fixed number of iterations have been completed. In this latter case, the number of iterations required depends on how accurately $\hat{G}$ approximates G. Asymptotic absolute stability is a way of characterizing, in the fixed stepsize case, how accurately $\hat{G}$ must approximate G in order that the sequence y_n remain bounded as $h \to \infty$. As one might expect, asymptotic absolute stability is more restrictive than the condition on the spectral radius, $\rho(\hat{G}^{-1}(G - \hat{G})) < 1$, required for the convergence of the iteration.

We are currently inclined toward the choice (A2, B2, C2), that is, choosing the constant step method to give a corrector of order $k + 1$; using the variable step coefficient to define α and modifying the iteration matrix to compensate; and doing a fixed number of iterations where the number done depends on the estimated error in the iteration matrix.

2. DERIVATION OF THE CHARACTERISTIC EQUATION

Following Klopfenstein, we observe

$$\begin{aligned} \mathbf{r}_{n+1}^{(j)} &= -\alpha(\mathbf{y}_{n+1}^{(j)} - \mathbf{p}_{n+1}) + J\mathbf{y}_{n+1}^{(j)} - \mathbf{p}'_{n+1} + J\mathbf{p}_{n+1} - J\mathbf{p}_{n+1}, \\ &= G(\mathbf{y}_{n+1}^{(j)} - \mathbf{p}_{n+1}) + \mathbf{r}_{n+1}^{(0)} = G(\mathbf{y}_{n+1}^{(j)} - \mathbf{p}_{n+1}) - \hat{G}(\mathbf{y}_{n+1}^{(1)} - \mathbf{p}_{n+1}), \end{aligned} \tag{6}$$

and since

$$\mathbf{r}_{n+1}^{(j)} = -\hat{G}\delta y_{n+1}^{(j)} = \hat{G}\mathbf{y}_{n+1}^{(j)} - \hat{G}\mathbf{y}_{n+1}^{(j+1)},$$

we have

$$\begin{aligned} \hat{G}\mathbf{y}_{n+1}^{(j+1)} &= \hat{G}\mathbf{y}_{n+1}^{(j)} - G(\mathbf{y}_{n+1}^{(j)} - \mathbf{p}_{n+1}) + \hat{G}(\mathbf{y}_{n+1}^{(1)} - \mathbf{p}_{n+1}) \\ &= -(G - \hat{G})(\mathbf{y}_{n+1}^{(j)} - \mathbf{p}_{n+1}) + \hat{G}\mathbf{y}_{n+1}^{(1)}, \end{aligned}$$

yielding

$$\mathbf{y}_{n+1}^{(j+1)} = -\hat{G}^{-1}\Delta(\mathbf{y}_{n+1}^{(j)} - \mathbf{p}_{n+1}) + \mathbf{y}_{n+1}^{(1)}, \tag{7}$$

which displays the immediate convergence when $\Delta = 0$. Repeated application of eq. (7) gives

$$\begin{aligned}\mathbf{y}_{n+1} = \mathbf{y}_{n+1}^{(m)} &= -\hat{G}^{-1}\Delta(\mathbf{y}_{n+1}^{(m-1)} - \mathbf{p}_{n+1}) + \mathbf{y}_{n+1}^{(1)} - \mathbf{p}_{n+1} + \mathbf{p}_{n+1} \\ &= -\hat{G}^{-1}\Delta[-\hat{G}^{-1}\Delta(\mathbf{y}_{n+1}^{(m-2)} - \mathbf{p}_{n+1}) + \mathbf{y}_{n+1}^{(1)} - \mathbf{p}_{n+1}] \\ &\quad + \mathbf{y}_{n+1}^{(1)} - \mathbf{p}_{n+1} + \mathbf{p}_{n+1} \\ &\quad \vdots \\ \mathbf{y}_{n+1} &= \sum_{r=0}^{m-1} (-\hat{G}^{-1}\Delta)^r(\mathbf{y}_{n+1}^{(1)} - \mathbf{p}_{n+1}) + \mathbf{p}_{n+1}.\end{aligned} \tag{8}$$

Since

$$\hat{G}(\mathbf{y}_{n+1}^{(1)} - \mathbf{p}_{n+1}) = -J\mathbf{p}_{n+1} + \mathbf{p}'_{n+1} = -\mathbf{r}_{n+1}^{(0)},$$

eq. (8) can be written in the form

$$\mathbf{y}_{n+1} = \sum_{r=0}^{m-1} (-\hat{G}^{-1}\Delta)^r\hat{G}^{-1}[\mathbf{p}'_{n+1} - J\mathbf{p}_{n+1}] + \mathbf{p}_{n+1}. \tag{9}$$

Since

$$\left[\sum_{r=0}^{m-1} (-\hat{G}^{-1}\Delta)^r\right][I + \hat{G}^{-1}\Delta] = [I - (-\hat{G}^{-1}\Delta)^m],$$

we have

$$\begin{aligned}\sum_{r=0}^{m-1} (-\hat{G}^{-1}\Delta)^r &= [I - (-\hat{G}^{-1}\Delta)^m](I + \hat{G}^{-1}\Delta)^{-1} \\ &= [I - (-\hat{G}^{-1}\Delta)^m](\hat{G} + \Delta)^{-1}\hat{G} \\ &= [I - (-\hat{G}^{-1}\Delta)^m]G^{-1}\hat{G}.\end{aligned}$$

And thus eq. (9) can be written

$$\begin{aligned}\mathbf{y}_{n+1} &= [I - (-\hat{G}^{-1}\Delta)^m]G^{-1}[\mathbf{p}'_{n+1} - J\mathbf{p}_{n+1}] + \mathbf{p}_{n+1} \\ &= [\mathrm{I} - (-\hat{\mathrm{G}}^{-1}\Delta)^m][-\mathbf{p}_{n+1} + G^{-1}(\mathbf{p}'_{n+1} - \alpha\mathbf{p}_{n+1})] + \mathbf{p}_{n+1} \\ &= (-\hat{G}^{-1}\Delta)^m\mathbf{p}_{n+1} + [I - (-\hat{G}^{-1}\Delta)^m]G^{-1}(\mathbf{p}'_{n+1} - \alpha\mathbf{p}_{n+1}).\end{aligned} \tag{10}$$

Substitution of the expressions for $\mathbf{p}_{n+1}$ and $\mathbf{p}'_{n+1}$ from eq. (2) (with $d_0 \equiv 0$) then gives

$$\mathbf{y}_{n+1} = \sum_{r=0}^{k} \left[(-\hat{G}^{-1}\Delta)^m + (I - (-\hat{G}^{-1}\Delta)^m)G^{-1}\left(\frac{d_r}{h} - \alpha\right)\right]\nabla^r\mathbf{y}_n \tag{11}$$

$$= \sum_{r=0}^{k} \left[B_m + C_m\left(\frac{d_r}{h} - \alpha\right)\right]\nabla^r\mathbf{y}_n, \tag{12}$$

where B_m is the matrix $(-\hat{G}^{-1}\Delta)^m$ and C_m is the matrix $(I - (-\hat{G}^{-1}\Delta)^m)G^{-1}$.

When k, B_m, C_m, α, h, and d_r are all constant we can write eq. (12) in the form

$$\mathbf{y}_{n+1} = \sum_{r=0}^{k} A_r \mathbf{y}_{n-r}, \tag{13}$$

which in turn can be written

$$\begin{bmatrix} \mathbf{y}_{n+1} \\ \mathbf{y}_n \\ \\ \vdots \\ \\ \mathbf{y}_{n-k+1} \end{bmatrix} = \begin{bmatrix} A_0 & A_1 & A_2 & \cdots & A_k \\ I & 0 & 0 & \cdots & 0 \\ 0 & I & 0 & \cdots & 0 \\ \vdots & & \ddots & & \\ & & & \ddots & \\ 0 & 0 & \cdots & I & 0 \end{bmatrix} \begin{bmatrix} \mathbf{y}_n \\ \mathbf{y}_{n+1} \\ \\ \vdots \\ \\ \mathbf{y}_{n-k} \end{bmatrix}$$

$$= W \begin{bmatrix} \mathbf{y}_n \\ \mathbf{y}_{n-1} \\ \\ \vdots \\ \\ \mathbf{y}_{n-k} \end{bmatrix} = W^{n+1} \begin{bmatrix} \mathbf{y}_0 \\ \mathbf{y}_{-1} \\ \\ \vdots \\ \\ \mathbf{y}_{-k} \end{bmatrix} \tag{14}$$

$$= T^{-1} \begin{bmatrix} z_1^{n+1} & & & 0 \\ & z_2^{n+1} & & \\ & & \ddots & \\ 0 & & & \end{bmatrix} T \begin{bmatrix} \mathbf{y}_0 \\ \\ \vdots \\ \\ \mathbf{y}_{-k} \end{bmatrix},$$

if W, the block companion matrix associated with eq. (14), can be diagonalized with a similarity transformation. Thus the solution to eq. (12) is given by

$$\mathbf{y}_n = \sum_i \mathbf{c}_i z_i^n, \tag{15}$$

where the z_i are eigenvalues of W, and the $\mathbf{c}_i$ are constant vectors which depend on the initial conditions and the eigenvectors of W.

To find regions of absolute stability, one must find the conditions necessary for $|z_i| \leq 1$, for all i. To investigate this, substitute $\sum_i \mathbf{c}_i z_i^n$ for $\mathbf{y}_n$ in eq. (12), obtaining

$$\sum_i \mathbf{c}_i z_i^{n+1} = \sum_{r=0}^{k} \left[B_m + C_m \left(\frac{d_r}{h} - \alpha \right) \right] \nabla^r \left(\sum_i \mathbf{c}_i z_i^n \right). \tag{16}$$

Since eq. (16) is true for all n, and z_i^n and z_j^n are linearly independent functions of n for $z_i \neq z_j$, eq. (16) must be satisfied for the individual terms in the i sums. Thus we replace $\sum_i \mathbf{c}_i z_i^n$ with $\mathbf{c}z^n$, and since $\nabla^r z^n = (1 - z^{-1})^r z^n$, we have

$$\mathbf{c} = \sum_{r=0}^{k} \left[B_m + C_m \left(\frac{d_r}{h} - \alpha \right) \right] z^{-1}(1 - z^{-1})^r \, \mathbf{c} = \sum_{r=0}^{k} E_{r,m} z^{-1}(1 - z^{-1})^r \, \mathbf{c}. \tag{17}$$

(If $z_i = z_{i+1} = \cdots = z_I$ then replace z_{i+j} with $n^j z_i$ for $j = 1, 2, \ldots, I - i$ in eq. (16) and a slight generalization of the above applies.)

Equation (17) is the general characteristic equation. Its solution in this general form for anything but a single first order differential equation is impractical. But in the special case where the $E_{r,m}$ are simultaneously diagonalizable, we can write

$$1 = \sum_{r=0}^{k} v_{j,r,m} z^{-1}(1 - z^{-1})^r, \tag{18}$$

where $v_{j,r,m}$ is the jth diagonal of $E_{r,m}$ after it has been diagonalized.

There are two cases of interest where eq. (17) can be put in the form of eq. (18). If $\Delta = 0$, then $B_m = 0$, $C_m = G^{-1}$, and

$$v_{j,r,m} = \frac{[(d_r/h) - \alpha]}{(\lambda - \alpha]},$$

where λ is an eigenvalue of J. Using the identities $(1 - z^{-1})^r = z[(1 - z^{-1})^r - (1 - z^{-1})^{r+1}]$ and $d_r - d_{r-1} = 1/r$, eq. (18) can be transformed (in the case $\Delta = 0$) to

$$h\lambda = \sum_{r=1}^{k} \frac{1}{r} (1 - z^{-1})^r + (h\alpha - d_k)(1 - z^{-1})^{k+1}, \tag{19}$$

which for $\alpha = d_{k+1}/h$ (or d_k/h) is the more familiar characteristic equation for the backward differentiation formula of order $k + 1$ (or k) when the corrector is solved exactly.

The other case is for $h \to \infty$. In this case the coefficient of $C_m \to 0$, $v_{j,r,m} \to \mu^m$ where μ is an eigenvalue of $-\hat{G}^{-1}\Delta$, and eq. (18) takes the form

$$1 = \mu^m \sum_{r=0}^{k} z^{-1}(1 - z^{-1})^r = \mu^m[1 - (1 - z^{-1})^{k+1}], \tag{20}$$

$$\mu^{-m} = 1 - (1 - z^{-1})^{k+1}. \tag{21}$$

In order to get Klopfenstein's result, note that μ is an eigenvalue of $-\hat{G}^{-1}\Delta = -\hat{G}^{-1}G + I$; $1 - \mu$ is an eigenvalue of $\hat{G}^{-1}G$; $1/(1 - \mu)$ is an eigenvalue of $G^{-1}\hat{G}$; and $g = 1 - (1 - \mu)^{-1} = -\mu/(1 - \mu)$ is an eigenvalue of $I - G^{-1}\hat{G} = G^{-1}(G - \hat{G})$. Solving for μ, we have Klopfenstein's eq. (3.23):

$$(1 - g^{-1})^m = 1 - (1 - z^{-1})^{k+1}. \tag{22}$$

It should perhaps be emphasized that the only assumption about the problem that is made to derive eq. (21) is that $\hat{G}^{-1}\Delta$ can be diagonalized with a similarity transformation; no assumptions on J are required.

3. CHOOSING BETWEEN CHARACTERIZATIONS OF ASYMPTOTIC STABILITY

Define

$$R_\mu = \{\mu: |\, z(\mu)| \leq 1\}, \tag{23}$$

where $|\, z(\mu)|$ is the magnitude of the largest z which satisfies eq. (21) for a given μ. Similarly, define R_g based on eq. (22). Since both $\mu = 0$ and $g = 0$ map into a $(k + 1)$ fold 0 of z, the origin is in both regions.

Table I Minimum $|\mu|$ for μ on the Boundary of the Region of Asymptotic Absolute Stability for a $P(EC)^m$ Algorithm with a Predictor of Order k

	Values of k						
m	0	1	2	3	4	5	6
1	1 0	0.333	0.143	0.067	0.032	0.016	0.008
2	1.0	0.577	0.378	0.258	0.179	0.126	0.088
3	1 0	0.693	0.523	0.405	0.318	0.251	0.199
4	1.0	0.760	0.615	0.508	0.424	0.355	0.298
.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
∞	1.0	1.0	1.0	1.0	1.0	1.0	1.0

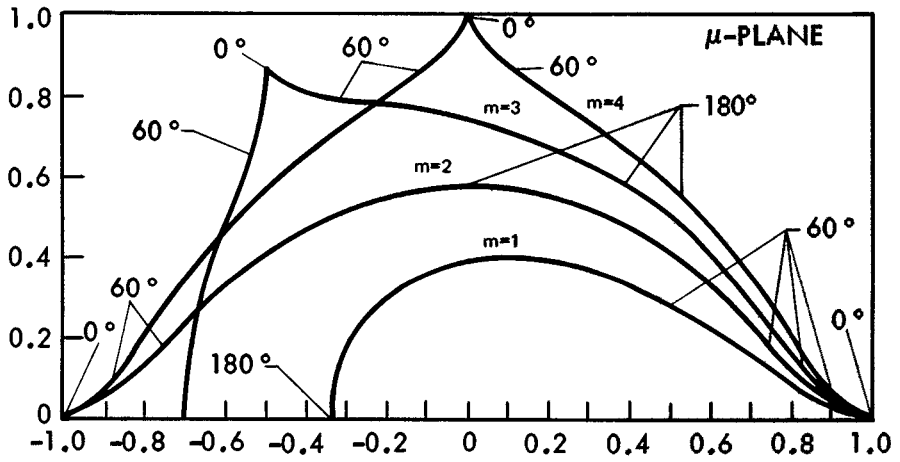

Fig. 1. Asymptotic ($h \to \infty$) stability regions for $k = 1$ and $m = 1 - 4$.

With Klopfenstein's characterization, R_g is the region of asymptotic absolute stability where g is an eigenvalue of $G^{-1}(G - \hat{G})$ as $h \to \infty$. We suggest using R_μ in this definition where μ is an eigenvalue of $-\hat{G}^{-1}(G - \hat{G})$ as $h \to \infty$. Note that these two characterizations are mathematically equivalent.

Our primary reason for suggesting this change is that μ is more intimately connected with the rate of convergence of the corrector iteration. From eq. (7) for successive values of j,

$$\delta \mathbf{y}_{n+1}^{(j)} = -\hat{G}^{-1}\,\Delta\delta \mathbf{y}_{n+1}^{(j-1)} = -\hat{G}^{-1}(G - \hat{G})\,\delta \mathbf{y}_{n+1}^{(j-1)}. \tag{24}$$

In Table I we have given the min $|\mu|$ for μ on the boundary of R_μ. Note that this value occurs at $z = -1$ giving a μ of

$$\mu(-1) = [1 - 2^{k+1}]^{-1/m}. \tag{25}$$

A secondary benefit of using R_μ is that the numbers in Table I are uniformly larger than the corresponding values of g given in Table 2 of [2] (a benefit when a code approximates the region by storing a single number, the minimum distance

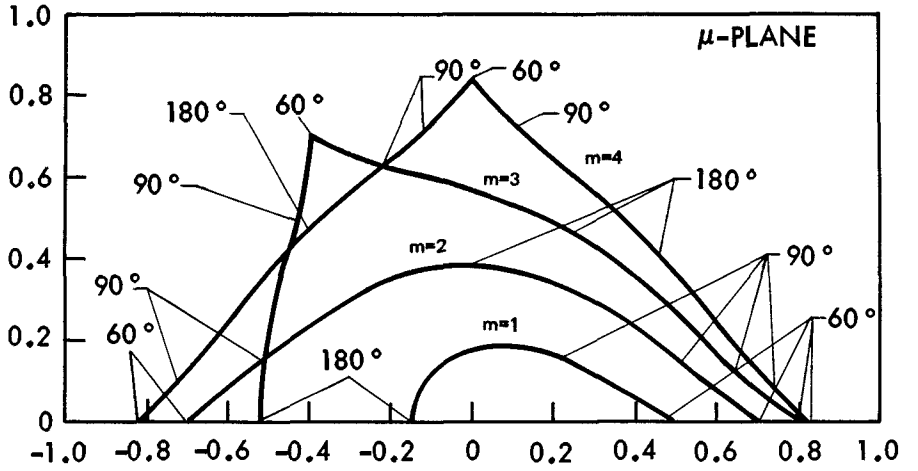

Fig. 2. Asymptotic ($h \to \infty$) stability regions for $k = 2$ and $m = 1 - 4$.

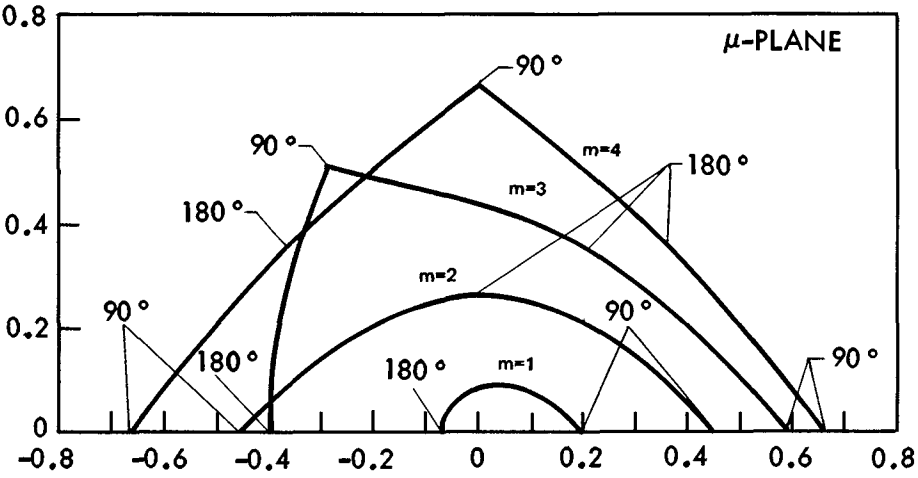

Fig. 3. Asymptotic ($h \to \infty$) stability regions for $k = 3$ and $m = 1 - 4$.

to the boundary from the origin, for each order). We henceforth take R_μ as the definition of the region of asymptotic absolute stability.

The regions R_μ are obtained by mapping $z = e^{i\theta}$ (the unit circle) into the μ-plane using eq. (21). We give the results in Figures 1-6. Results are not given for $k = 0$, since in this case the result is obviously a unit circle. In the figures we have also given the value of $|\theta|$ associated with some of the points on the boundary of R_μ.

This value of θ is useful in understanding the behavior of the differences as a function of their order. It is easy to see that $\nabla^k z^n = (1 - z^{-1})^k z^n$. When instability is first noticed, z should be reasonably close to the unit circle, and thus we have

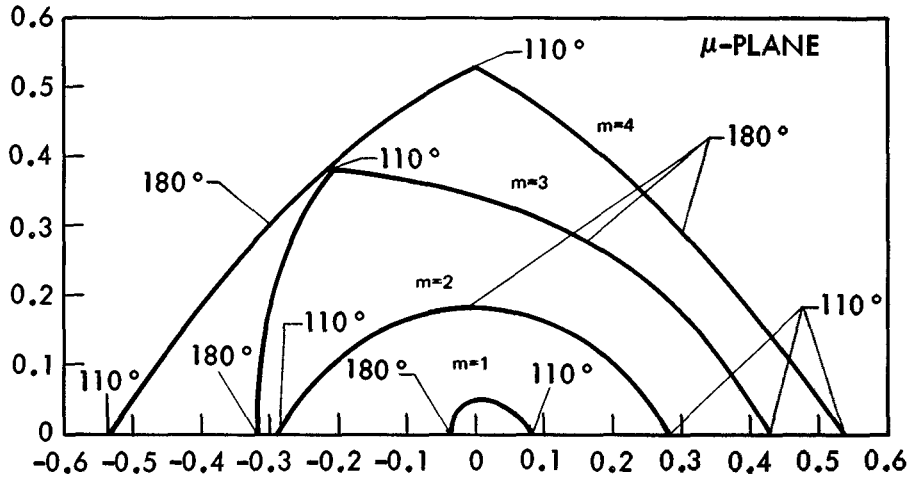

Fig. 4. Asymptotic ($h \to \infty$) stability regions for $k = 4$ and $m = 1 - 4$.

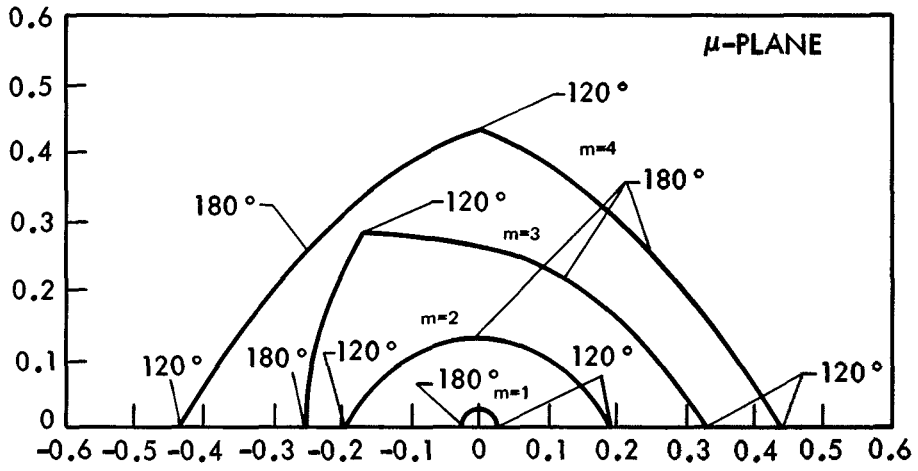

Fig. 5. Asymptotic ($h \to \infty$) stability regions for $k = 5$ and $m = 1 - 4$.

the approximation

$$|\nabla^k z^n| \approx |(1 - e^{-i\theta})^k z^n| = |(1 - \cos\theta) + i \sin\theta\,|^k |z^n|$$
$$= |2(1 - \cos\theta)|^{k/2} |z^n|$$
$$= [D(\theta)]^k |z^n|.$$

Note that $D(\theta)$ increases monotonically with $|\theta|$ in the interval $[0, \pi]$, $D(0) = 0$, $D(\pi/3) = 1$, $D(\pi/2) = \sqrt{2}$, $D(\pi) = 2$. For $|\theta| \gtrsim 90°$, instability gives a roughness in the solution that tends to induce a reduction in k. (This kind of natural protection happens to be present for most methods, including the Adams

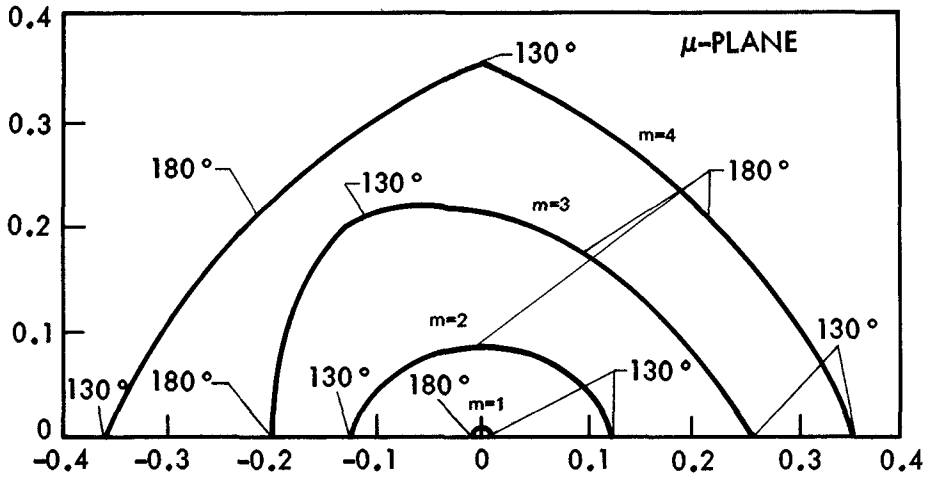

Fig. 6. Asymptotic ($h \to \infty$) stability regions for $k = 6$ and $m = 1 - 4$.

method and the BDF corrector solved exactly.) For $|\theta| \lesssim 60°$, the undesired root does not provide this warning and thus instability can occur without increasing the error estimate or leading to a reduced integration order. Since such values of $|\theta|$ can occur for $k = 0$ or 1, some provision for detecting the instability, other than checking the convergence of differences, scaled derivatives, or error estimators, is necessary at low order. (Note: the case $k = 0$ arises if a code allows the use of a zeroth-order predictor with a first-order corrector. We believe there are cases where this is in fact the best choice.)

4. THE CHOICE OF c IN EQUATION (4)

We have found that introducing the scalar c to minimize the spectral radius of $\hat{G}^{-1}(G - \hat{G})$ results in a significant improvement in the performance of a code. Although the results given earlier have guided us in what we present here, the theory only applies rigorously as $h \to \infty$. Thus performance of our code is the primary justification for what we present here [5]. We believe that a more complete study of eq. (17) would provide a stronger justification.

To derive a value of c, we assume that $\hat{J} = J$, in which case

$$\hat{G}^{-1}(G - \hat{G}) = c^{-1}(J - \hat{\alpha}I)^{-1}\,[(1 - c)J - (\alpha - c\hat{\alpha})I]. \tag{26}$$

Using arguments similar to those used earlier, one can show that if λ is an eigenvalue of J, then

$$\mu(c, \lambda) = -\frac{(1 - c)\lambda - (\alpha - c\hat{\alpha})}{c(\lambda - \hat{\alpha})} = 1 - \frac{1}{c}\left(\frac{\alpha - \lambda}{\hat{\alpha} - \lambda}\right) \tag{27}$$

is an eigenvalue of the error matrix $-\hat{G}^{-1}(G - \hat{G})$. Note there are situations with zero matrix error: (a) if all eigenvalues equal 0 and $c = \alpha/\hat{\alpha}$; (b) if $\alpha = \hat{\alpha}$ or the magnitude of all eigenvalues $\to \infty$, and $c = 1$. We have found the algebraic

manipulations which follow to be simplified if we assume c has the form

$$c = 1 + \frac{(\alpha - \hat{\alpha})}{m}, \tag{28}$$

and then determine m. With this substitution,

$$\mu = 1 - \frac{m}{m + (\alpha - \hat{\alpha})} \frac{\alpha - \lambda}{\hat{\alpha} - \lambda} = \frac{(\alpha - \hat{\alpha})}{(m + \alpha - \hat{\alpha})} \frac{\hat{\alpha} - m - \lambda}{\hat{\alpha} - \lambda}.$$

We are concerned with the sizes of $|\mu|$ which correspond to the values of λ in the spectrum of J, $\sigma(J)$, and therefore want to examine the quantity

$$\frac{|\mu|^2}{(\alpha - \hat{\alpha})^2} = \frac{1}{(m + \alpha - \hat{\alpha})^2} \frac{|\hat{\alpha} - m - \lambda|^2}{|\hat{\alpha} - \lambda|^2}. \tag{29}$$

Differentiating the right member of eq. (29) with respect to m, setting the result to 0, and solving for m gives

$$m = \hat{\alpha} - \text{Re } \lambda + \frac{(\text{Im } \lambda)^2}{\alpha - \text{Re } \lambda}. \tag{30}$$

By checking the second derivative, one can verify that $|\mu|$ is minimized for a particular λ by the m given in eq. (30). Observe that $m \geq \hat{\alpha} > 0$ if Re $\lambda \leq 0$, which we assume to be the case here and below.

For a given $|\lambda|$, we now show that $|\mu|$ takes its maximum value on the imaginary axis. With $\psi = [(m + \alpha - \hat{\alpha})^2/(\alpha - \hat{\alpha})^2]|\mu|^2$, $\lambda = re^{i\theta}$, we have

$$\frac{\partial \psi}{\partial \theta} = \frac{2mr \sin \theta}{|\hat{\alpha} - re^{i\theta}|^4} (\hat{\alpha}^2 - r^2 - \hat{\alpha}m), \tag{31}$$

and since $m \geq \hat{\alpha}$, $\partial\psi/\partial\theta \leq 0$ for $\pi/2 \leq \theta \leq \pi$, which is the desired result.

With $\lambda = iy$, we have

$$\frac{\partial \psi}{\partial y} = \frac{2my}{(\hat{\alpha}^2 + y^2)^2} (2\hat{\alpha} - m), \tag{32}$$

and thus ψ is an increasing function of y if $m < 2\hat{\alpha}$, and a decreasing function of y if $m > 2\hat{\alpha}$.

This derivation and a desire for computational efficiency suggest the following procedure: Estimate $|\lambda_{\max}|$ from J using a power method, and $|\lambda_{\min}|$ from G_f using an inverse power method. The total cost is about seven matrix-vector multiplies when a new Jacobian is formed. Since the largest $|\mu|$ for a given $|\lambda|$ occurs on the imaginary axis, we assume λ is purely imaginary and let

$$m_1 = \hat{\alpha} + \frac{|\lambda_{\max}|^2}{\alpha} \geq m_2 = \hat{\alpha} + \frac{|\lambda_{\min}|^2}{\alpha}. \tag{33}$$

If $m_1 < 2\hat{\alpha}$, eq. (32) indicates ψ is an increasing function for each $|\lambda|$, $\lambda \in \sigma(J)$, and we can minimize $|\mu|$ by using the largest value of $|\lambda|$, set $m = m_1$. If $m_2 > 2\hat{\alpha}$, a similar argument sets $m = m_2$. If neither is true, m is obtained by setting the value of $|\mu|$ to be equal at the extreme points, $|\lambda_{\max}|$, $|\lambda_{\min}|$. From eq. (32), this happens for $m = 2\hat{\alpha}$.

Given m, c is given by eq. (28). The corresponding estimates for $|\mu|$ are

$$|\mu| = \begin{cases} \dfrac{|\alpha - \hat{\alpha}|\,|\lambda_{\max}|}{[(|\lambda_{\max}|^2 + \hat{\alpha}^2)(|\lambda_{\max}|^2 + \alpha^2)]^{1/2}} & \text{if } m = m_1 \\ \dfrac{|\alpha - \hat{\alpha}|\,|\lambda_{\min}|}{[(|\lambda_{\min}|^2 + \hat{\alpha}^2)(|\lambda_{\min}|^2 + \alpha^2)]^{1/2}} & \text{if } m = m_2 \\ \dfrac{|\alpha - \hat{\alpha}|}{(\alpha + \hat{\alpha})} & \text{if } m = 2\hat{\alpha}. \end{cases} \tag{34}$$

Another approach to getting a low cost reduction in the spectral radius of $\hat{G}^{-1}(G - \hat{G})$ which the first author believes has real promise is described in [5]. Lest someone believe we are missing the obvious, we would like to note that because of computation noise, we have found ratios such as

$$\frac{\|\delta \mathbf{y}_{n+1}^{(j+1)}\|}{\|\delta \mathbf{y}_{n+1}^{(j)}\|}$$

or

$$\frac{\|\mathbf{r}_{n+1}^{(j+1)}\|}{\|\mathbf{r}_{n+1}^{(j)}\|}$$

hazardous for use in making computational decisions. Also, in part because we don't need such ratios, we very frequently use only a single iteration.

5. CONCLUDING REMARKS

We have a code under development which is based on the ideas presented here. Even though the preceding analysis does not apply to variable order-variable step methods or to nonlinear problems, the analysis for the simplified case has proved very useful. A few highlights on how these ideas are used in this code follow. We compute a new iteration matrix when we estimate that the largest eigenvalue of $\hat{G}^{-1}(G - \hat{G})$ exceeds the value given in Table I for $m = 2$, and $k =$ the current order of the predictor. This estimate is the sum of $|\mu|_{\max}$ and an estimate of the magnitude of the largest eigenvalue of $\hat{G}^{-1}(J - \hat{J})$. This estimate requires the equivalent of a matrix-vector multiply. A new J is computed when forming a new G_f only if it appears that without a new J stability would require more than one function evaluation per step. On most problems we tend to evaluate J much less frequently than the iteration matrix is factored.

ACKNOWLEDGMENT

The comments of the referees have helped in clarifying the ideas presented in this paper.

REFERENCES

1. JACKSON, K.R. AND SACKS-DAVIS, R. An alternative implementation of variable step-size multistep formulas for stiff ODEs. *ACM Trans. Math. Softw. 6*, 3 (Sept. 1980), 295–318.
2. KLOPFENSTEIN, R W Numerical differentiation formulas for stiff systems of ordinary differential equations *RCA Review 32* (1971), 447–462.

3. KROGH, F.T. Changing stepsize in the integration of differential equations using modified divided differences. In *Proc Conference on the Numerical Solution of Ordinary Differential Equations*, (Austin, Texas, Oct 19–20, 1972), Lecture Notes in Math., vol. 362, Springer-Verlag, New York, 1974, pp. 22–71.
4. KROGH, F.T Notes on partitioning in the solution of stiff equations. Section 366, Internal Memorandum No. 388, Jet Propulsion Laboratory, Pasadena, Calif., March 1982.
5. STEWART, K. AND KROGH, F.T. Preliminary comparison test results for STRUT, LSODE and LSODA Section 366, Internal Memorandum No. 499, Jet Propulsion Laboratory, Pasadena, Calif., July 1983.

Received Oct. 1981; revised July 1983; accepted July 1983