

PECE Algorithms for the Solution of Stiff Systems of Ordinary Differential Equations

By R. W. Klopfenstein and C. B. Davis

Abstract. This paper presents a study of a class of PECE algorithms consisting of an application of a predictor followed by application of one iteration of a pseudo Newton-Raphson method to a corrector. Such algorithms require precisely two evaluations of the derivative function for each forward step. Theorems 1 and 4 show that the stability properties of such algorithms compare favorably with those obtained with application of the Newton-Raphson method to the corrector iterated to convergence.

A subclass of these algorithms have local truncation error of second order and some have local truncation error of third order. Theorems 2 and 3 exhibit members of this subclass wherein an estimate of the local truncation error is explicit in the algorithm at each step.

Initially (in Theorem 1) these algorithms are characterized in terms of their stability properties in the limit as the interval of integration becomes indefinitely large. In Section 5, their properties for other intervals of integration are discussed through the study of some enclosure properties.

I. Introduction. The phenomenon of "stiffness" in the numerical solution of ordinary differential equations occurs in problems from a great variety of sources. Among these are the simulation of parabolic partial differential equations where the space variable has been discretized so that a system of ordinary differential equations in the time variable results. The efficacy of implicit formulas in such problems has been recognized for some time and, indeed, Richtmeyer [8] has proposed local linearization (the Newton-Raphson method) for solving the nonlinear systems of equations which may arise therein. Another important area often more limited by the "stiffness" phenomenon is the numerical determination of the transient response of electrical circuits [1], [11]. In this context, the difficulty has been dubbed "the small time constant problem" and is an infamous source of frustration to users of large computer programs for the analysis of such circuits. Recently, considerable progress has been made in the alleviation of difficulties arising in such problems.

Dahlquist [3] defined the property of A -stability for linear multistep methods which requires that the solution of the equation $y' = \lambda y$ tend to zero as the number of steps approaches infinity for arbitrary interval size, h , whenever λ is in the left-half complex plane. He showed that this property was obtainable only with implicit formulas of order equal to or less than two, and that among formulas of second order the trapezoidal rule has the smallest error constant. Widlund [12] showed that formulas of orders three and four could be obtained if the domain of stability was restricted to a wedge-shaped region within the left-half plane rather than the entire left-half plane. Liniger and Willoughby [7] have exhibited implicit formulas through order four

Received August 4, 1969, revised June 11, 1970.

AMS 1969 subject classifications. Primary 6561, 6568.

Key words and phrases. Ordinary differential equations, numerical solution, numerical stability, predictor-corrector method, Newton-Raphson approximation.

Copyright © 1971, American Mathematical Society

having the property of A -stability provided the class of linear multistep methods is broadened to include terms involving the second derivative.

More recently, a number of authors [4], [7], [10] have been concerned with the implementation of such implicit formulas in general purpose algorithms for the numerical solution of stiff systems of ordinary differential equations. In the instances cited, such algorithms have been based on a particular implicit formula of the type discussed above with the solution to this implicit formula being obtained by an iterative application of the Newton-Raphson method. Typically, the previous point (y_n) is taken as first approximation to the new point (y_{n+1}). Sandberg and Shichman [10] discuss some results corresponding to a single iteration of the Newton-Raphson method as well as iterative application when the open Euler formula is used as an implicit formula.

One motivation for use of iterative applications of the Newton-Raphson method is quite clear. To the extent that iteration to convergence is carried out, the stability properties of the algorithm are completely described by the stability properties of the implicit formula on which the algorithm is based. This is generally an easier question to study than the stability properties of an algorithm based on a fixed number of iterations. There are some drawbacks to such algorithms, however. First of all, the number of iterations (and, hence, derivative evaluations) per step forward is open-ended and possibly quite large. Secondly, if (y_n) is taken as first approximation, there is no direct estimation of the truncation error incurred at each step forward as an explicit part of the algorithm (though indirect methods may be used [10]). Finally, there are limitations on the convergence of the iterative process itself (see [7] and our Theorem 4).

We should also mention some implicit algorithms patterned after Runge-Kutta processes, first proposed by Rosenbrock [9]. These algorithms require only a fixed number of derivative evaluations and may be of third or higher order. Callahan [2] describes an attractive third-order algorithm of this type and also mentions the importance of nonsensitivity to the value of the Jacobian matrix. Haines [5] discusses third-order algorithms of this type from a somewhat more general point of view and develops algorithms which include a mechanism for estimation of the truncation error at each step.

In this paper, we are concerned with the properties of one-step and two-step algorithms of the form

$$\begin{aligned}
 \mathbf{p}_{n+1} &= \mathbf{y}_n + h[\alpha \mathbf{f}_n + \beta \mathbf{f}_{n-1}], \\
 \mathbf{c}_{n+1} &= \mathbf{y}_n + h[v \mathbf{f}(x_{n+1}, \mathbf{p}_{n+1}) + u \mathbf{f}_n], \\
 \mathbf{y}_{n+1} &= \mathbf{p}_{n+1} + [aI - v h \tilde{\mathbf{J}}]^{-1}(\mathbf{c}_{n+1} - \mathbf{p}_{n+1}),
 \end{aligned}
 \tag{1.1}$$

where, as usual, we have the differential equations

$$\mathbf{y}' = \frac{d\mathbf{y}}{dx} = \mathbf{f}(x, \mathbf{y}), \quad \mathbf{y}_0 = \mathbf{y}(0).
 \tag{1.2}$$

\mathbf{y}_n is the result of the n th application of the algorithm and \mathbf{f}_n denotes $\mathbf{f}(x_n, \mathbf{y}_n)$. The α , β , u , v , and a are real constants while I is the identity matrix of order N (N being the number of equations in the system (1.2)) and $\tilde{\mathbf{J}}$ is a matrix of order N which approximates the Jacobian matrix of (1.2), i.e., $\partial \mathbf{f} / \partial \mathbf{y}$.

Such algorithms are of Adams-type (involve only the most recent value of y_n) and require precisely two derivative evaluations per step forward. When “ a ” is equal to unity, (1.1) corresponds to one application of a predictor followed by one application of the Newton-Raphson method to a corrector, and, hence, we refer to these algorithms as PECE algorithms. We shall see later that values of “ a ” somewhat different from unity furnish a most useful parameter in the design of algorithms of this class.

Algorithms of the form (1.1) lend themselves well to organization of intermediate results, and when the predictor and corrector are independently of the same (first or second) order, a useful error estimator is explicitly present (see Theorems 2 and 3). Algorithms with $\beta = 0$ in (1.1) enjoy the further advantage of requiring no special starting procedures.

We shall study the stability properties of these algorithms in terms of the difference between the approximate Jacobian matrix, \tilde{J} , in the algorithm and the actual Jacobian matrix, J , of the system (1.2).

II. Asymptotic Absolute Stability. We consider the system of ordinary differential equations (1.2) and a $(k + 1)$ -step algorithm

$$(2.1) \quad y_{n+1} = \mathbf{F}(h, \tilde{J}, y_n, \dots, y_{n-k}, f_n, \dots, f_{n-k}),$$

which generates a sequence of approximate solution values to (1.2) with h a real positive scalar. We refer further to the linear system of ordinary differential equations with constant coefficients

$$(2.2) \quad \mathbf{y}' = \frac{d\mathbf{y}}{dx} = J\mathbf{y} + \mathbf{c},$$

with \mathbf{c} a constant vector, and J a constant matrix of order N with eigenvalues properly in the left-half complex plane.

Definition 1. The algorithm (2.1) is said to be asymptotically ($h \rightarrow \infty$) absolutely stable in solving the system (2.2) if the sequence $\{y_n\}_{n=0}^{\infty}$ is uniformly bounded in norm

1. for all positive h for some \tilde{J} which may depend on h , and
2. for all sufficiently large $h \geq h_0 > 0$ when the eigenvalues of

$$h\Delta = h(J - \tilde{J})$$

lie within a region R containing a circle of radius $c_1 > 0$ centered on the origin where c_1 is independent of h .

Definition 2. The region R of Definition 1 is called the “asymptotic region of absolute stability” for the algorithm (2.1).

Throughout this paper, the norms of vectors and matrices will refer to the 2-norm thereof [13, p. 80].

From these definitions, it can be seen that we are focusing on the behavior of an algorithm (2.1) when applied to a (possibly nonlinear) system (1.2) by studying its behavior when applied to a related linear system (2.2), i.e., the “locally linearized” equivalent (1.2). Our justification for doing this is that any results so obtained will be strictly applicable to such linear systems while at the same time providing insight into the behavior of algorithms when applied to other systems (nonlinear or linear

with variable coefficients). Differences between J and \tilde{J} may occur for a number of reasons. It may not be possible to obtain accurate values for the elements of J , for example, if they are obtained by numerical differentiation. Also due to errors in the process of matrix inversion, the inverse matrix appearing in (1.1) will correspond to a somewhat different Jacobian matrix, i.e., \tilde{J} . Finally, when (1.1) is being applied to a nonlinear problem, the matrix \tilde{J} may be out-of-date, i.e., it no longer furnishes an accurate value for the current J . It is perhaps worth noting that except for changes in the matrix, \tilde{J} , the inverse matrix in (1.1) need only be formed when the interval of integration changes. It is also worth observing that when the matrix whose inverse is indicated in (1.1) is sparse (a common occurrence in practice), efficiencies may be effected by solving the sparse system of linear equations implied by the third equation of (1.1) at each step of the integration rather than retaining a preformed matrix inverse [7].

It is convenient in what follows to express the third equation of (1.1) as

$$(2.3) \quad \mathbf{y}_{n+1} = \mathbf{p}_{n+1} + (I - \nu h(J - \Delta_1))^{-1}(\mathbf{c}_{n+1} - \mathbf{p}_{n+1}),$$

where $\Delta_1 = J - \tilde{J} - \frac{1-a}{\nu h} I$,

so that the parameter "a" no longer appears explicitly in the algorithm. It is sufficient to apply the algorithm (1.1) to (2.2) with the constant term set equal to zero. It may be verified by direct substitution that the sequence \mathbf{y}_n , so obtained, leads readily through $\mathbf{y}_n(\mathbf{c}) = \mathbf{y}_n - J^{-1}\mathbf{c}$ to that obtained for nonzero \mathbf{c} , and $\mathbf{y}_n(\mathbf{c})$ is bounded in norm if and only if \mathbf{y}_n is. When this is done, the following two-step difference equation is obtained for the \mathbf{y}_n

$$(2.4) \quad (I - \nu h(J - \Delta_1))\mathbf{y}_{n+1} = [I + u h J + \nu h \Delta_1(I + \alpha h J)]\mathbf{y}_n + \beta \nu h^2 \Delta_1 J \mathbf{y}_{n-1}.$$

We restrict our attention to those matrices J which can be fully diagonalized by a similarity transformation and define

$$(2.5) \quad Q = h P J P^{-1}, \quad \mathbf{z}_n = P \mathbf{y}_n, \quad D = h P \Delta_1 P^{-1},$$

where Q is a diagonal matrix whose elements lie entirely in the left-half complex plane. When these variables are introduced into (2.4) along with premultiplication of both sides of (2.4) by the matrix P followed finally by the change of variable

$$(2.6) \quad \mathbf{t}_n = Q \mathbf{z}_n,$$

we obtain

$$(2.7) \quad [(I + \nu D)Q^{-1} - \nu I]\mathbf{t}_{n+1} = [(I + \nu D)Q^{-1} + (uI + \alpha \nu D)]\mathbf{t}_n + \beta \nu D \mathbf{t}_{n-1}.$$

We shall now look for conditions on D and on the parameters of the algorithm such that the conditions of Definition 1 are satisfied.

We note first that for $D = 0$, and, hence,

$$(2.8) \quad \tilde{J} = J - \frac{1-a}{\nu h} I,$$

that

$$(2.9) \quad \mathbf{t}_{n+1} = (I - \nu Q)^{-1}(I + uQ)\mathbf{t}_n.$$

Since the matrix on the right-hand side of (2.9) is diagonal, we can see that when the elements of Q are in the left-half complex plane, the solutions of (2.9) tend to zero for all $h > 0$ and $n \rightarrow \infty$ provided that

$$(2.10) \quad v > 0, \quad |u| \leq v.$$

Thus, under conditions (2.10) we obtain the \tilde{J} required in the first part of Definition 1.

In studying other values for D , we shall first place (2.7) in the form of two simultaneous first-order recursions. We obtain

$$R_1 \mathbf{w}_{n+1} = S_1 \mathbf{w}_n,$$

with

$$(2.11) \quad \mathbf{w}_n = \begin{bmatrix} \mathbf{t}_n \\ \mathbf{t}_{n-1} \end{bmatrix}, \quad R_1 = \begin{bmatrix} R & 0 \\ 0 & I \end{bmatrix}, \quad S_1 = \begin{bmatrix} S & T \\ I & 0 \end{bmatrix},$$

where

$$\begin{aligned} R &= vI + (I + vD)Q^{-1}, \\ S &= uI + \alpha vD + (I + vD)Q^{-1}, \\ T &= \beta vD. \end{aligned}$$

We can see that the last terms in the expressions for R and S will be small in some sense for large intervals of integration. We proceed at this point by setting these terms equal to zero and in Appendix 1 make precise how this relates to the limiting behavior demanded in Definition 1. We may then write (2.11) in the form

$$(2.12) \quad \mathbf{w}_{n+1} = C \mathbf{w}_n, \quad \text{with } C = \begin{bmatrix} -\frac{u}{v} I - \alpha D & -\beta D \\ I & 0 \end{bmatrix}.$$

It is easily seen that the solutions of (2.12) will tend to zero when the eigenvalues of C are strictly less than one in absolute value. We seek these eigenvalues through

$$(2.13) \quad Cz = \lambda z, \quad \text{with } z = \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix},$$

and these lead to the simultaneous equations

$$(2.14) \quad \left(-\frac{u}{v} I - \alpha D \right) \mathbf{u} - \beta D \mathbf{v} = \lambda \mathbf{u}, \quad \mathbf{u} = \lambda \mathbf{v}.$$

Eliminating \mathbf{u} , we obtain

$$(2.15) \quad D \mathbf{v} = -\frac{\lambda}{v} \frac{v\lambda + u}{\alpha\lambda + \beta} \mathbf{v},$$

and can now assert that the eigenvalues of C will be less than unity when the eigenvalues of D lie within the complement of the map of the exterior of the unit circle in the λ -plane under the transformation

$$(2.16) \quad d(\lambda) = -\frac{\lambda}{v} \frac{v\lambda + u}{\alpha\lambda + \beta}.$$

The region R so defined is symmetrical with respect to the real axis since the parameters of (2.16) are real. When it is not empty, it can be translated through selection of an appropriate value of the real parameter "a" to provide the circular region required in Definition 1.

The mapping (2.16) is precisely two-to-one and may be written in the form

$$(2.17) \quad \lambda^2 + (\gamma + \alpha d)\lambda + \beta d = 0, \quad \text{with } \gamma = u/v, |\gamma| \leq 1.$$

Thus, to establish the existence of a nonempty region R , it is sufficient to exhibit a real point, d , both of whose images are within the unit circle. $d = 0$ is a suitable point when $|\gamma| < 1$, since its images are

$$(2.18) \quad \lambda_1 = 0, \quad \lambda_2 = -\gamma.$$

To examine the two limiting cases, we expand (2.17) through first order in d , obtaining

$$(2.19) \quad \lambda_1 = -\frac{\beta}{\gamma}d + O(d^2), \quad \lambda_2 = -\gamma + \frac{\beta - \alpha\gamma}{\gamma}d + O(d^2),$$

where γ is either plus one or minus one. From (2.19), we are assured of a value d such that both of its images are interior to the unit circle when $|\gamma| = 1$, except possibly when $\beta = \alpha$ (for $\gamma = 1$) and $\beta = -\alpha$ (for $\gamma = -1$). But in these cases, the roots of (2.17) can be expressed exactly as

$$(2.20) \quad \lambda_{1,2} = -\alpha d, -1; \quad \lambda_{1,2} = -\alpha d, 1;$$

respectively. In these last cases, the y_n do not tend to zero but the conditions of Definition 1 are still met since they remain bounded.

We have therefore proved the following:

THEOREM 1. *The PECE algorithm (1.1) with $v > 0$, $|u| \leq v$, $\alpha^2 + \beta^2 \neq 0$, is asymptotically ($h \rightarrow \infty$) absolutely stable in solving the system (2.2) when the eigenvalues of the matrix*

$$(2.21) \quad h\Delta = h(J - \tilde{J}),$$

lie within a closed subset of the open region R defined by

$$(2.22) \quad d(\lambda) = \frac{1}{v} \left\{ 1 - a - \lambda \frac{v\lambda + u}{\alpha\lambda + \beta} \right\}, \quad \lambda = e^{i\varphi}, \quad -\pi \leq \varphi \leq \pi,$$

where "a" may be selected as different from unity, if necessary, so that R properly contains the origin.

Hence, we see that algorithms of the form (1.1) are asymptotically stable under very broad conditions on the parameters. The parameter "a" is available to "center" the region R on the origin. The remaining parameters are available to achieve other desirable properties in the algorithm.

III. First- and Second-Order Algorithms. In practice, one would like the algorithm (1.1) to have truncation errors of at least second order in h and perhaps third-order truncation errors would be desirable in many cases. We have retained the arbitrary parameters up to this point mainly to exhibit under what broad conditions the property of asymptotic absolute stability may be obtained. From this point on,

the parameters α and v will be restricted as given by

$$(3.1) \quad \alpha = 1 - \beta, \quad v = 1 - u.$$

This results in the predictor and corrector each being of at least first order. Furthermore, for the values

$$(3.2) \quad \beta = -\frac{1}{2}, \quad u = \frac{1}{2},$$

the predictor and corrector, respectively, will be of second order. While it is possible for the predictor to have order one less than the corrector without reducing the order of the PECE algorithm, we will generally refrain from this in order to have a direct estimate of the truncation error incurred at each step.

Let the exact solution of (1.2) at any given point be denoted by $y(x_n)$ and suppose, in the usual way, that $y_n = y(x_n)$. Then

$$(3.3) \quad \begin{aligned} y(x_{n+1}) &= y_n + hy'(x_n) + \frac{1}{2}h^2y''(x_n) + \frac{1}{6}h^3y'''(x_n) + O(h^4), \\ f_{n-1} &= y'(x_n) - hy''(x_n) + \frac{1}{2}h^2y'''(x_n) + O(h^3), \\ f_n &= y'(x_n), \\ f_{n+1} &= y'(x_n) + hy''(x_n) + \frac{1}{2}h^2y'''(x_n) + O(h^3), \end{aligned}$$

so that

$$(3.4) \quad \begin{aligned} e_p &= y(x_{n+1}) - p_{n+1} = (\beta + \frac{1}{2})h^2y''(x_n) - \frac{1}{2}(\beta - \frac{1}{3})h^3y'''(x_n) + O(h^4), \\ e_c &= y(x_{n+1}) - c_{n+1} = (u - \frac{1}{2})h^2y''(x_n) + \frac{1}{2}(u - \frac{2}{3})h^3y'''(x_n) + O(h^4), \\ c_{n+1} - p_{n+1} &= (\beta - u + 1)h^2y''(x_n) - \frac{1}{2}(\beta + u - 1)h^3y'''(x_n) + O(h^4). \end{aligned}$$

It should be noted that the use of f_{n+1} in place of $f(x_{n+1}, p_{n+1})$ does not affect the second-order terms of (3.4) (for a first-order predictor) or the third-order terms (for a second-order predictor). The Eqs. (3.4) furnish the basis for assertions in the following two theorems.

THEOREM 2. *The algorithm (1.1) with $\alpha = 1 - \beta, v = 1 - u, u < \frac{1}{2}$,*

1. *is asymptotically ($h \rightarrow \infty$) absolutely stable in solving the system (2.2) when the eigenvalues of the matrix $h\Delta = h(J - \tilde{J})$ lie within a closed subset of the open region R given by (2.22),*

2. *is of first order with local truncation error given by*

$$(3.5) \quad \begin{aligned} e_v &= y(x_{n+1}) - y_{n+1} \\ &= [(\beta + \frac{1}{2})I - (\beta - u + 1)(aI - (1 - u)h\tilde{J})^{-1}]h^2y''(x_n) + O(h^3), \end{aligned}$$

3. *has a first-order estimate for the local truncation error incurred at each step given by*

$$(3.6) \quad e_{n+1} = p_{n+1} - y_{n+1} + \frac{\beta + \frac{1}{2}}{\beta - u + 1} (c_{n+1} - p_{n+1}).$$

THEOREM 3. *The algorithm (1.1) with $\alpha = \frac{3}{2}, \beta = -\frac{1}{2}, v = \frac{1}{2}, u = \frac{1}{2}$,*

1. *is asymptotically ($h \rightarrow \infty$) absolutely stable in solving the system (2.2) when the*

eigenvalues of the matrix $h\Delta = h(J - \tilde{J})$ lie within a closed subset of the open region R given by (2.22),

2. is of second order with local truncation error given by

$$(3.7) \quad \begin{aligned} e_v &= y(x_{n+1}) - y_{n+1} \\ &= \left[\frac{5}{12} I - \frac{1}{2} \left(aI - \frac{h}{2} \tilde{J} \right)^{-1} \right] h^3 y'''(x_n) + O(h^4), \end{aligned}$$

3. has a first-order estimate for the local truncation error incurred at each step given by

$$(3.8) \quad e_{n+1} = p_{n+1} - y_{n+1} + \frac{5}{6}(c_{n+1} - p_{n+1}).$$

It should be noted that the parameter "a" has not been specified in the algorithms of Theorems 2 and 3. It may be selected in any given case to "center" the region R on the origin. One possibility is to so select "a" such that a circle centered on the origin of maximum radius may be inscribed in R . However, other criteria may be relevant in some special cases.

For the first-order algorithms described in Theorem 2, we will restrict our attention to the case $\beta = 0$ since that case has the important advantage of requiring no special starting procedures. The predictor is now the Euler formula. In this case, (2.22) can be written

$$(3.9) \quad d(\lambda) = - \left\{ \lambda + \frac{a - (1 - u)}{1 - u} \right\},$$

and it is clear that if we set $a = 1 - u$, the map of the unit circle will be a circle of unit radius centered on the origin. We may obtain larger regions up to circles with radius approaching two by using other values for β but this would be at the expense of the self-starting feature.

For the second-order algorithm of Theorem 3, the region R specified by (2.22) is exhibited in Fig. 1 where "a" has been set equal to one. The bounding curve of this

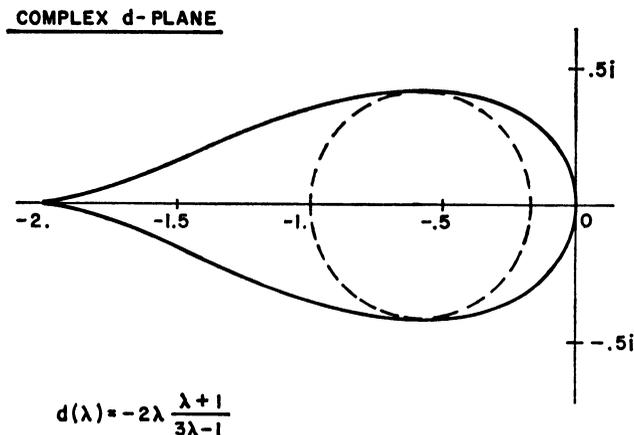


FIGURE 1. Stable Region for Algorithm of Theorem 3

region has a cusp at $d = -2$ since this is a branch point of the mapping. The largest circle that can be inscribed in this region has center at about $d = -.58$ and radius of about 0.42. The appropriate value of "a" that will shift the region R so that the center of this inscribed circle is at the origin is $a = 0.71$.

IV. Comparison with Iterated Corrector Algorithm. Under the restrictions of Theorem 1, the second equation of (1.1), considered as an independent implicit formula, is A -stable, that is, stable unconditionally for systems of differential equations whose eigenvalues lie entirely in the left-half complex plane. It is natural to seek solutions of this implicit equation through an iterative application of the Newton-Raphson method and, indeed, several authors have suggested this [4], [7], [10]. We would like to examine the properties of such an algorithm within the same frame of reference which has been applied here to PECE algorithms in order to judge their relative merits.

The general first-order corrector with iterative applications of the Newton-Raphson method can be expressed as

$$(4.1) \quad \begin{aligned} \mathbf{c}_{n+1}^{(m+1)} &= \mathbf{y}_n + h[v\mathbf{f}(x_{n+1}, \mathbf{y}_{n+1}^{(m)}) + u\mathbf{f}_n], \\ \mathbf{y}_{n+1}^{(m+1)} &= \mathbf{y}_{n+1}^{(m)} + (I - vh\tilde{J})^{-1}(\mathbf{c}_{n+1}^{(m+1)} - \mathbf{y}_{n+1}^{(m)}), \end{aligned}$$

with

$$m = 0, 1, 2, 3, \dots; \quad \Delta = J - \tilde{J}; \quad v = 1 - u.$$

We apply (4.1) to the system (2.2) and obtain

$$(4.2) \quad (I - vh\tilde{J})\mathbf{y}_{n+1}^{(m+1)} = vh\Delta\mathbf{y}_{n+1}^{(m)} + (I + uhJ)\mathbf{y}_n + hc.$$

The convergent value of (4.1), however, satisfies

$$(4.3) \quad (I - vh\tilde{J})\mathbf{y}_{n+1} = vh\Delta\mathbf{y}_{n+1} + (I + uhJ)\mathbf{y}_n + hc.$$

We therefore define

$$(4.4) \quad \boldsymbol{\varepsilon}_m = \mathbf{y}_{n+1}^{(m)} - \mathbf{y}_{n+1},$$

and subtract (4.3) from (4.2) to obtain

$$(4.5) \quad (I - vh\tilde{J})\boldsymbol{\varepsilon}_{m+1} = vh\Delta\boldsymbol{\varepsilon}_m.$$

Again, we restrict our attention to those matrices \tilde{J} which can be fully diagonalized by a similarity transformation and define

$$(4.6) \quad Q = hP\tilde{J}P^{-1}, \quad D = hP\Delta P^{-1}, \quad \mathbf{t}_m = P\boldsymbol{\varepsilon}_m,$$

obtaining upon substitution into (4.5)

$$(4.7) \quad (I - vQ)\mathbf{t}_{m+1} = vD\mathbf{t}_m.$$

P is selected so that Q is a diagonal matrix whose elements generally lie in the left-half complex plane, though there may be exceptions to this in border line cases since it is the approximate \tilde{J} that has been diagonalized rather than J . (4.7) may therefore be examined by its components and we have established the following:

THEOREM 4. *The iterated corrector algorithm (4.1) is convergent when applied to the system (2.2), provided that the eigenvalues of $h\Delta = h(J - \tilde{J})$ lie entirely within a closed subset of an open circular region centered on the origin with radius*

$$r = \frac{d_{\min}}{1 - u},$$

where

$$(4.8) \quad d_{\min} = \text{Min}_i |1 - (1 - u)h\tilde{\lambda}_i|,$$

the $\tilde{\lambda}_i$ being the eigenvalues of \tilde{J} . If the $\tilde{\lambda}_i$ lie entirely in the left-half complex plane $d_{\min} \geq 1$.

Theorem 4 provides a basis for comparison of the iterated corrector algorithm (4.1) with the PECE algorithms of Theorems 2 and 3. Asymptotically ($h \rightarrow \infty$), the iterated corrector algorithm (4.1) has a much larger region of convergence than the stable region of the PECE algorithms of Theorems 2 and 3 since the radius of that region approaches a constant ($\tilde{\lambda}_{\min}$) rather than being of order $1/h$. However, note that it is the smallest eigenvalue of \tilde{J} which limits the convergence of (4.1) so that for much of the solution we would have d_{\min} only slightly larger than one. Only as the solution of a system of equations was closely approaching its steady-state value, would values of d_{\min} substantially greater than one be admissible while retaining an appropriately small error in the solution function.

Thus, we see that in spite of the asymptotic superiority of the iterated corrector algorithm (4.1), for many purposes, the explicit PECE algorithms of Theorems 2 and 3 may be comparable insofar as sensitivity to the value of J is concerned and have the added advantage that only an explicit number (two) of derivative evaluations are required at each step forward.

V. Enclosure Properties. We desire to study the properties of PECE algorithms for values of Q other than the infinite limit. As a first step in this direction, we consider the case where Δ can be simultaneously diagonalized with J . The results from such an assumption are strictly applicable for decoupled systems of equations and by implication should be relevant for lightly coupled systems. We proceed and note that from (2.7) the characteristic equation can be written in this case in the form

$$(5.1) \quad d = -\frac{\lambda}{v} \frac{v\lambda + u - (\lambda - 1)w}{\alpha\lambda + \beta - \lambda(\lambda - 1)w},$$

where $w = 1/q$, d and q are used in place of any corresponding components of D and Q , and λ is a corresponding eigenvalue of (2.13). Note that w is in the left-half complex plane when q is.

Definition 3. An algorithm (1.1) is said to have the property of enclosure for all q in a set S_q if for every q in S_q the map of the exterior of the unit circle in the λ -plane under (5.1) is a subset of that map with $q \rightarrow \infty$ ($w = 0$).

If an algorithm has this property its stability properties for q in S_q are at least as favorable as they are at infinity. Note that the point $\lambda = 1$ is common to the mapping (5.1) for all q .

We consider first the first-order algorithms of Theorem 2 with $\beta = 0$. In this case, (5.1) simplifies to

$$(5.2) \quad \delta = d + \frac{u}{v} = -\frac{\lambda - (\lambda - 1)w}{1 - (\lambda - 1)w}.$$

This mapping is bilinear in λ for all w and, hence, it maps the unit circle in the λ -plane onto circles and lines in the δ -plane. In fact, all of these mappings have in common the point $\delta = -1$ with a vertical tangent (corresponding to $\lambda = 1$). This situation is depicted in Fig. 2.

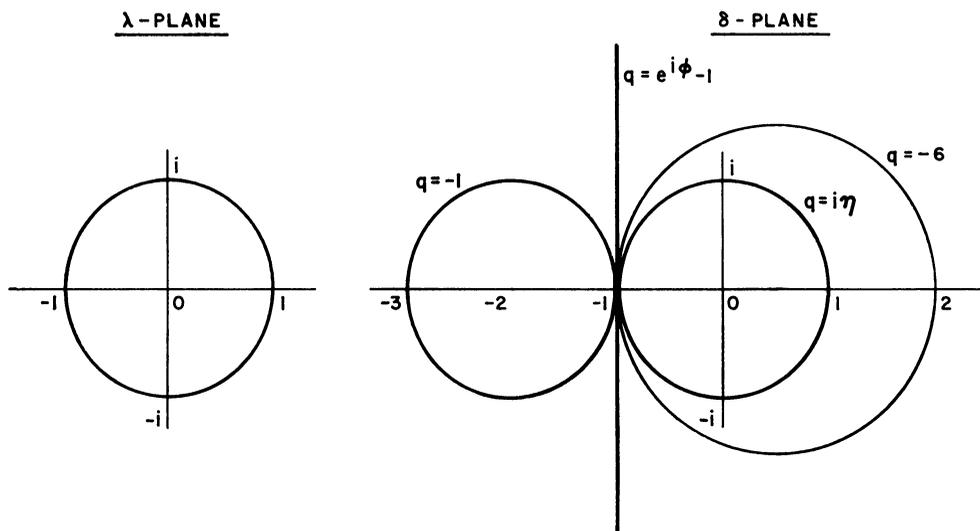


FIGURE 2. Stable Region for Algorithm of Theorem 2 with $\beta = 0, a = 1 - u$

The exterior of the unit circle in the λ -plane maps in the δ -plane onto

- (1) the exterior of the unit circle for q on the imaginary axis and not equal to zero,
- (2) the exterior of circles enclosing the unit circle for q in the left-half plane exterior to the circle $e^{i\phi} - 1$,
- (3) the half-plane to the left of the line $\delta = -1$ for q on the circle $e^{i\phi} - 1$ and not equal to zero, and
- (4) the interior of circles tangent to and to the left of the line $\delta = -1$ for q interior to the circle $e^{i\phi} - 1$,

since the infinite point in the λ -plane maps into $\delta = -1 + q$.

We have, therefore, established the following:

THEOREM 5. *The first-order algorithms of Theorem 2 with $\beta = 0$ have the property of enclosure for all q in the left-half complex plane.*

For the second-order algorithm of Theorem 3, (5.1) becomes

$$(5.3) \quad d = -2\lambda \frac{\lambda + 1 - 2(\lambda - 1)w}{3\lambda - 1 - 2\lambda(\lambda - 1)w}.$$

This mapping is considerably more difficult to study than that of (5.2) since it is no longer bilinear in λ .

The mapping for $w = 0$ is exhibited in Fig. 1. A complex number which is an exterior normal to the bounding curve at a point (w, λ) is given by

$$(5.4) \quad n = \lambda d_\lambda, \quad \text{where } d_\lambda = \partial d / \partial \lambda.$$

On the other hand, the incremental change in d due to a change in w is given by

$$(5.5) \quad \Delta d = d_w \Delta w + O((\Delta w)^2), \quad \text{where } d_w = \partial d / \partial w.$$

The requirement that such an increment in w not carry a point on the boundary into the interior of the region can be expressed as

$$(5.6) \quad -\frac{\pi}{2} \leq \text{Arg} \left(\frac{\lambda d_\lambda}{d_w} \right) - \text{Arg} (\Delta w) \leq \frac{\pi}{2}.$$

Applying this to (5.3), we find that

$$(5.7) \quad f = \frac{\lambda d_\lambda}{d_w} = \frac{(3 - 2w)\lambda + 1 + 2w}{2(\lambda - 1)^2},$$

and for $\lambda = e^{i\varphi}$,

$$(5.8) \quad f = -\frac{1}{4}(1 - \cos \varphi)^{-1}[3 - 2w + (1 + 2w)e^{-i\varphi}].$$

The argument (phase angle) of f will limit the permissible arguments for Δw in (5.6) for a given (w, λ) . We now have the following result.

THEOREM 6. *The second-order algorithm of Theorem 3 has the property of enclosure in a region S_q which*

1. *contains the entire negative real half-line, and*
2. *is bounded asymptotically as $|q| \rightarrow \infty$ by the lines*

$$(5.9) \quad -\frac{\pi}{2} + \text{Arcsin} \left(\frac{1}{3} \right) \leq \text{Arg} (-q) \leq \frac{\pi}{2} - \text{Arcsin} \left(\frac{1}{3} \right).$$

The boundary of the region S_q for this algorithm is exhibited in Fig. 3. The details of the calculation leading to the determination of this boundary are discussed in Appendix 2.

It should be mentioned that the property of enclosure is not necessary for useful stability properties for an algorithm. Neither is it sufficient in view of the limitation on the diagonalizability of D introduced at the beginning of this section. However, it is one meaningful attribute which is useful in characterizing an algorithm. Additional work is needed to obtain results similar to those of this section for more general systems of ordinary differential equations.

VI. Computational Experiences. An adaptive subroutine based on the algorithm of Theorem 3 with $a = .71$ was programmed in FORTRAN IV for the RCA Spectra 70/45 computer. This subroutine has proved effective in the numerical solution of a variety of stiff systems of ordinary differential equations arising in the transient response of electronic circuits and in some one-dimensional heat flow problems.

It is not our intent to report here detailed experimental verification of the results of the theory presented in the preceding sections of this paper. However, experimental

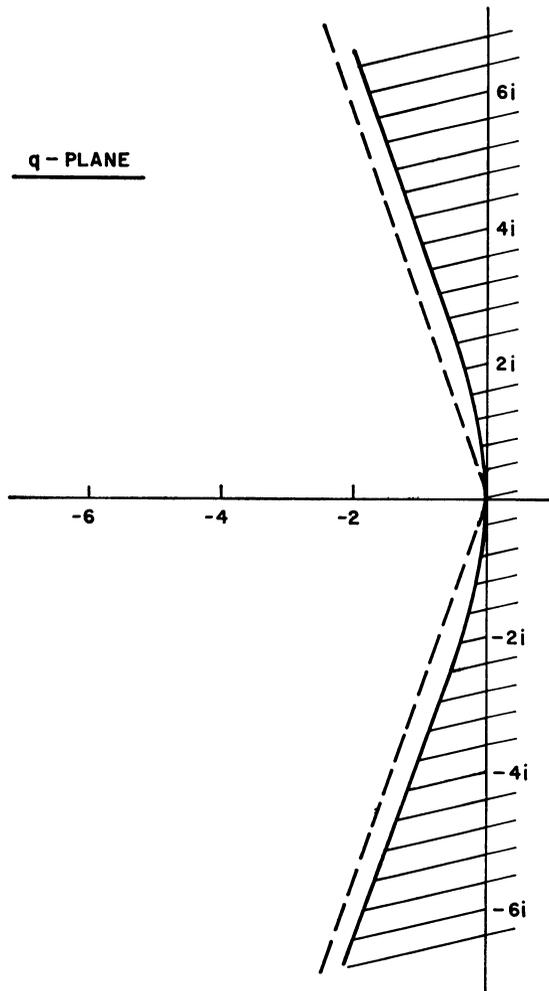


FIGURE 3. *Region S_q of Enclosure for Algorithm of Theorem 3*

calculations have been carried out in connection with a linear system of differential equations with the Jacobian matrix perturbed by a family of matrices with known eigenvalues. These experiments have been sufficient to verify the main features of the asymptotically stable region exhibited in Fig. 1.

VII. Conclusions. In this paper we have studied the properties of a class of PECE algorithms consisting of an application of a predictor followed by application of one iteration of a pseudo Newton-Raphson method to a corrector. Such algorithms require precisely two evaluations of the derivative function for each forward step. We have shown in Theorems 1 and 4 that the stability properties of such algorithms compare favorably with those obtained with application of the Newton-Raphson method to the corrector iterated to convergence.

An important subclass of these algorithms have local truncation errors of second-order and some have local truncation errors of third-order. In Theorems 2

and 3, we exhibit members of this subclass wherein an estimate of the local truncation error is explicit in the algorithm at each step.

Whereas initially (in Theorem 1) these algorithms were characterized in terms of their stability properties in the limit as the interval of integration becomes indefinitely large, we discuss their properties for other intervals of integration in Section 5 through the study of some enclosure properties.

VIII. Acknowledgment. We would like to acknowledge with pleasure the contribution of Dr. Werner Liniger, Thomas J. Watson Research Center, Yorktown Heights, New York, to the material presented in this paper through several stimulating discussions. We would also like to thank the referee for suggestions leading to the more precise statement of Theorems 1, 2, 3, and 4.

Appendix 1. Precise Limiting Process for Definition 1. In determining the limiting region of absolute stability for Theorem 1, we simply set the terms involving Q^{-1} in (2.11) to zero. In this appendix, we examine with more precision the effect on the eigenvalues as those terms approach zero.

If we divide the members of (2.7) by $-v$, it becomes

$$(A1.1) \quad [I - A]t_{n+1} = -\left[\frac{u}{v}I + \alpha D + A\right]t_n - \beta D t_{n-1},$$

$$A = \frac{1}{v}(I + vD)Q^{-1},$$

which in turn can be written as a single first-order recursion

$$R_1 w_{n+1} = S_1 w_n,$$

with

$$(A1.2) \quad w_n = \begin{bmatrix} t_n \\ t_{n-1} \end{bmatrix}, \quad R_1 = \begin{bmatrix} R & 0 \\ 0 & I \end{bmatrix}, \quad S_1 = \begin{bmatrix} S & T \\ I & 0 \end{bmatrix},$$

where

$$R = I - A, \quad S = -\frac{u}{v}I - \alpha D - A, \quad T = -\beta D.$$

We now study the eigenvalue problem

$$S_1 z = \lambda R_1 z,$$

with

$$(A1.3) \quad R_1 = I + E, \quad S_1 = S_0 + E, \quad E = -\begin{bmatrix} A & 0 \\ 0 & 0 \end{bmatrix}.$$

A very closely related problem is discussed by Wilkinson [13, p. 135] which differs from the present one only in that E appears only in S_1 and not in R_1 . By following procedures exactly similar to those followed there, we find

$$(A1.4) \quad \min_i |\lambda - \lambda_i| \leq |1 - \lambda| \|P_1\| \|E\| \|P_1^{-1}\|,$$

where λ is an eigenvalue of (A1.3), the λ_i are the eigenvalues of the unperturbed system (2.13), and P_1 is the matrix which diagonalizes S_0 . (A1.4) holds for an arbitrary matrix E though we are interested primarily in the case where $\|E\|$ is small. As noted before, the norms of vectors and matrices used here are the 2-norms [13, p. 80].

We will restrict our attention to the case $|\lambda| < 1$ for which (A1.1) has solutions approaching zero. In that case, $|1 - \lambda| < 2$, and further

$$\begin{aligned} (A1.5) \quad \|E\| = \|A\| &\leq \frac{1}{v} \|Q^{-1}\| \|I + vD\| \\ &\leq \frac{1 + v\|D\|}{vq_{\min}}, \end{aligned}$$

where q_{\min} is h times the magnitude of the smallest eigenvalue of J (see (2.5)). Since the eigenvalues of Δ and, hence, of D are restricted by (2.22), we have that

$$(A1.6) \quad \text{Min}_i |\lambda - \lambda_i| \leq K/h,$$

where K is positive and bounded. This assures us that the eigenvalues of (A1.3) lie within a union of discs of radius K/h , with the eigenvalues of (2.13) as centers. By taking h sufficiently large, the radius of these discs can be made as small as desired.

By requiring the λ_i to be within a circle of radius $1 - K/h$, we can be assured that the eigenvalues of (A1.3) will be less than unity. The complement of the map of the exterior of this circle will be internal to the region of stability R referred to in Theorem 1. However, by selecting h large enough this region, say R_h , will also enclose the origin since R properly encloses the origin.

It is important to note that we are not asserting that absolute stability can be obtained for a given fixed Δ for arbitrarily large h , but only for Δ which continue to meet the requirement of (2.22) as h increases.

Appendix 2. Enclosure Property for Second-Order Algorithm. In order to determine the region in the q -plane having the property of enclosure, we set

$$(A2.1) \quad d(w, \lambda) = d(0, \eta),$$

where $d(w, \lambda)$ is given by (5.3). Values of w , λ , and η , satisfying (A2.1), represent an intersection of the regions of stability R and R_h . If this equation is solved for w , we obtain explicitly

$$(A2.2) \quad q = \frac{1}{w} = \frac{-2\lambda(\lambda - 1)(\eta - 1)^2}{\lambda(\lambda + 1)(3\eta - 1) - \eta(\eta + 1)(3\lambda - 1)}.$$

The envelope of curves traced out by (A2.2) in the q -plane as λ and η range over

$$(A2.3) \quad \lambda = e^{i\varphi}, \quad \eta = e^{i\alpha}, \quad -\pi \leq \varphi, \alpha \leq \pi,$$

furnish the boundary of S_q .

This can be thought of as a one-parameter family of curves in the real (x, y) plane with

$$\begin{aligned}
 q &= x + iy, \\
 (A2.4) \quad q(\lambda, \eta) &= F(\varphi, \alpha) + iG(\varphi, \alpha), \quad \text{so that} \\
 x &= F(\varphi, \alpha), \\
 y &= G(\varphi, \alpha).
 \end{aligned}$$

A condition that must be satisfied on the envelope of this one-parameter family of curves is

$$(A2.5) \quad J\left(\frac{F, G}{\varphi, \alpha}\right) = F_\varphi G_\alpha - F_\alpha G_\varphi = 0.$$

The last two equations together determine the envelope.

Now

$$(A2.6) \quad \frac{\partial q}{\partial \varphi} = i\lambda \frac{\partial q}{\partial \lambda} = F_\varphi + iG_\varphi, \quad \frac{\partial q}{\partial \alpha} = i\eta \frac{\partial q}{\partial \eta} = F_\alpha + iG_\alpha,$$

so that

$$(A2.7) \quad \text{Imag} \left[\lambda \bar{\eta} \frac{\partial q}{\partial \lambda} \frac{\partial \bar{q}}{\partial \eta} \right] = 0,$$

where $\text{Imag} [\]$ denotes the imaginary part of and the upper bar denotes complex conjugate. Upon performing the operations indicated in (A2.7), we obtain

$\text{Imag} [f_1 \bar{f}_2] = 0$, where

$$\begin{aligned}
 (A2.8) \quad f_1 &= \lambda \left[\frac{1}{\lambda} + \frac{1}{\lambda - 1} - \frac{1}{\lambda - \eta} - \frac{3\eta - 1}{3\lambda\eta - \lambda - \eta - 1} \right], \quad \text{and} \\
 f_2 &= \eta \left[\frac{2}{\eta - 1} + \frac{1}{\lambda - \eta} - \frac{3\lambda - 1}{3\lambda\eta - \lambda - \eta - 1} \right].
 \end{aligned}$$

Computationally, one proceeds by selecting a value for φ and finding a corresponding value for α satisfying (A2.8). This (φ, α) pair is used in (A2.2) to obtain a point q on the boundary of S_α . The result of this computation is exhibited in Fig. 3. To within the accuracy of the computation, these points lie on the left branch of the hyperbola

$$(A2.9) \quad 8(x - \frac{1}{2})^2 - y^2 - 2 = 0,$$

where $q = x + iy$. However, we know of no way of directly inferring this from (A2.2) and (A2.7).

RCA Laboratories
Princeton, New Jersey 08540

1. F. H. BRANIN, JR., "Computer methods of network analysis," *Proc. IEEE*, v. 55, 1967, pp. 1787-1801.

2. D. A. CALLAHAN, "A stable, accurate method of numerical integration for nonlinear systems," *Proc. IEEE*, v. 56, 1968, p. 744.

3. G. G. DAHLQUIST, "A special stability problem for linear multistep methods," *BIT*, v. 3, 1963, pp. 27-43. MR 30 #715.

4. C. W. GEAR, "The automatic integration of stiff ordinary differential equations," *Proc. IFIPS Congress*, 1968, pp. A81-A85.
5. C. F. HAINES, "Implicit integration processes with error estimate for the numerical solution of differential equations," *Comput. J.*, v. 12, 1969/70, pp. 183-187. MR 39 #7817.
6. W. LINIGER, "A criterion for A -stability of linear multistep integration formulae," *Computing (Arch. Elektron. Rechnen)*, v. 3, 1968, pp. 280-285. MR 39 #1120.
7. W. LINIGER & R. A. WILLOUGHBY, "Efficient numerical integration of stiff systems of ordinary differential equations," *SIAM J. Numer. Anal.*, v. 7, 1970, pp. 47-66.
8. R. D. RICHTMYER, *Difference Methods for Initial-Value Problems*, Interscience Tracts in Pure and Appl. Math., Tract 4, Interscience, New York, 1957. MR 20 #438.
9. H. H. ROSENBROCK, "Some general implicit processes for the numerical solution of differential equations," *Comput. J.*, v. 5, 1962/63, pp. 329-330. MR 27 #5368.
10. I. W. SANDBERG & H. SHICHMAN, "Numerical integration of stiff nonlinear differential equations," *Bell System Tech. J.*, v. 47, 1968, pp. 511-527. MR 38 #1836.
11. S. R. SEDORE, *SCEPTRE: A Second Generation Transient Analysis Program*, Proc. Computer-Aided Circuit Design Seminar, (NASA), Cambridge, Mass., 1967, pp. 55-61.
12. O. B. WIDLUND, "A note on unconditionally stable linear multistep methods," *BIT*, v. 7, 1967, pp. 65-70. MR 35 #6373.
13. J. H. WILKINSON, *Rounding Errors in Algebraic Processes*, Prentice-Hall, Englewood Cliffs, N. J., 1963. MR 28 #4661.