stichting
mathematisch
centrum

MC

AFDELING NUMERIEKE WISKUNDE NW 15/75 MARCH

J.G. VERWER

S-STABILITY AND STIFF-ACCURACY FOR TWO CLASSES OF GENERALIZED INTEGRATION METHODS

2e boerhaavestraat 49 amsterdam

Printed at the Mathematical Centre, 49, 2e Boerhaavestraat, Amsterdam.

The Mathematical Centre, founded the 11-th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O), by the Municipality of Amsterdam, by the University of Amsterdam, by the Free University at Amsterdam, and by industries.

AMS (MOS) subject classification scheme (1970): 65L05

S-stability and stiff-accuracy for two classes of generalized integration methods

by

J.G. Verwer

ABSTRACT

The S-stability and stiff-accuracy (concepts proposed by A. PROTHERO and A. ROBINSON in Math. Comp., Vol. 28, No 125, pp.145-162, 1974) are studied for a class of generalized, linear multistep methods and generalized Runge-Kutta methods. These integration methods are characterized by the fact that the coefficients in the integration formulas are matrices, which depend on the Jacobian matrix of the differential equation under consideration. An S-stable, stiffly accurate two-point Runge-Kutta method is developed.

KEY WORDS & PHRASES: *Numerical analysis, ordinary differential equations, initial value problems, stiff equations, S-stability, stiff-accuracy.*

CONTENTS

1. Introduction 1

2. Generalized multistep methods with zero-parasitic roots 3

3. Generalized Runge-Kutta methods 7

4. S-stability 11

5. Stiff-accuracy 17

6. S-stability and stiff-accuracy for the class of generalized multistep methods 19

7. S-stability and stiff-accuracy for generalized Runge-Kutta methods 21

8. An S-stable generalized Runge-Kutta formula 28

9. Discussion 33

References 39

1. INTRODUCTION

Let

(1.1) $y' = f(x,y)$

represent a set of *stiff* differential equations subject to the initial condition

$$y = y_0 \text{ at } x = x_0.$$

A basic difficulty in the numerical solution of stiff systems is the requirement of stability. The stability analysis of integration methods for stiff systems has centred largely on the stability of numerical solutions to the linear test equation

(1.2) $y' = \delta y$, δ complex, $\mathrm{Re}(\delta) < 0$,

where δ stands for an eigenvalue of the Jacobian matrix of system (1.1). Almost all concepts of stability, developed for equation (1.2), are modifications of the concept of *A-stability* proposed by DAHLQUIST [3].

In using A-stable, implicit one-step methods to solve stiff, non-linear systems with eigenvalues located in widely separated clusters, PROTHERO & ROBINSON [9] have found that

(a) some A-stable methods give highly unstable solutions, and

(b) the accuracy of the solutions obtained often appears to be unrelated to the order of the method used.

This has caused them to re-examine the form of stability required when stiff systems are solved and to question the relevance of the concept of order of accuracy for stiff problems.

In their paper, PROTHERO & ROBINSON [9] propose the test-equation

(1.3) $y' = g'(x) + \delta(y-g(x))$, δ complex, $\mathrm{Re}(\delta) < 0$,

where g' is any given bounded function. They analyse the stability of numerical approximations to the solution $y \equiv g$ of equation (1.3) and derive necessary and sufficient conditions for such stability, which is termed

S-stability. The new stability concept generalizes the concept of A-stability.

They also propose a new consistency concept for stiff equations, namely the concept of *stiff-accuracy*. Roughly spoken, the accuracy of numerical approximations to the solution $y \equiv g$ of equation (1.3) is discussed by considering the *asymptotic form of the local truncation error* in the limit

$$h\ \mathrm{Re}(\delta) \to -\infty \text{ and } h \to 0,$$

where h denotes an integration stepsize.

The new concepts of S-stability and stiff-accuracy provide a greater insight into the numerical difficulties that are encountered with stiff systems. PROTHERO & ROBINSON [9] investigated implicit one-step methods based on quadrature formulas. We applied the new theory to two classes of generalized integration methods, namely *generalized, linear multistep methods* with *zero-parasitic roots* (see HOUWEN, VAN, DER & VERWER [8] and VERWER [13]) and *generalized Runge-Kutta methods* (see e.g. HOUWEN, VAN, DER [6]). These generalized methods are characterized by the fact that the coefficients in the integration formulas are matrices which depend on the Jacobian matrix of the system to be integrated.

The results of our investigation are described in this report. In section 2 and 3 we give short outlines of the integration methods to be analyzed. Section 4 and 5 are devoted to the new concepts of S-stability and stiff-accuracy. In section 6 and section 7 we discuss results of the generalized multistep- and one-step methods, respectively, concerning the new concepts of stability and accuracy. In section 8 we propose a two-point generalized Runge-Kutta scheme which is developed according to the new criteria. In the last section, this new scheme, as well as some others, is applied to a test-equation of the form (1.3) and to a non-linear scalar equation. More extensive numerical results of the new Runge-Kutta scheme will appear in a forthcoming paper.

2. GENERALIZED MULTISTEP METHODS WITH ZERO-PARASITIC ROOTS

The multistep formulas we are going to analyse belong to the class of *generalized, linear k-step methods* with *zero parasitic roots* and *adaptive stability function* (principal root). These integration methods are developed in HOUWEN, VAN, DER & VERWER [8] and further discussed in VERWER [13]. In this section we only give an outline of the method; for details the reader is referred to the references mentioned above.

Let the points x_j, $j = n+1,n,\dots,n+1-k$ denote the reference points of the k-step formula. Let the parameter h_n denote the steplength, $h_n = x_{n+1}-x_n$. Further, let y_n denote the numerical approximation to the solution $y(x_n)$ of system (1.1) at $x = x_n$, and let $f_n = f(x_n,y_n)$. Define

(2.1) $\qquad R$ and B_ℓ, $\ell = 1,\dots,k$,

to be *rational functions* with real coefficients. Then, the generalized, linear k-step method with zero-parasitic roots and adaptive principal root is defined by

(2.2) $$y_{n+1} = R(h_nJ_n)y_n + h_n \sum_{\ell=1}^{k} B_\ell(h_nJ_n)[f_{n+1-\ell} - J_ny_{n+1-\ell}],$$

where J_n denotes the *Jacobian matrix* of system (1.1) at the point (x_n,y_n). *The principal root is identified with the prescribed stability function R.*

Define the numbers

(2.3) $$q_\ell = (x_{n-\ell}-x_n)/h_n, \quad \ell = -1,0,\dots,k-1,$$

and let $y^{(j)}(x_n)$ denote

$$y^{(j)}(x_n) = \frac{d^j}{dx^j}\, y(x)\Big|_{x=x_n}, \quad j = 0,1,\dots .$$

By substituting a solution y of the differential equation into the right-hand side of (2.2), and by expanding

$$y'(x_{n+1-\ell}) \text{ and } y(x_{n+1-\ell})$$

about $x = x_n$, we obtain the formal expansion

$$(2.4) \qquad y_{n+1} = \sum_{j=0} C_j(h_n J_n) h_n^j y^{(j)}(x_n),$$

where the functions C_j are defined by

$$(2.5) \qquad C_0(z) = R(z) - \sum_{\ell=1}^{k} z\, B_\ell(z),$$

$$(2.6) \qquad C_j(z) = \frac{1}{j!} \sum_{\ell=1}^{k} (j - zq_{\ell-1}) q_{\ell-1}^{j-1}\, B_\ell(z),\ j = 1,2,\ldots\ .$$

Now it is clear that a possible choice for *consistency conditions* for *p-th order accuracy* is

$$(2.7) \qquad \begin{aligned} &C_j(z) = 1/j!,\ j = 0,\ldots,p-1, \\ &C_p(z) = 1/p! + O(z) \text{ as } z \to 0. \end{aligned}$$

By assuming that R, the stability function, is consistent of order k, i.e.

$$(2.8) \qquad \left.\frac{d^i}{dz^i} R(z)\right|_{z=0} = 1,\ i = 0,\ldots,k,$$

it may be proved that the maximal attainable order, in case of conditions (2.7), is $p = k$, if $k \geq 2$. For $k = 1$ the order $p = 2$ if R is consistent of order greater than one. In this report we shall assume that $p = k$. Then, the consistency conditions (2.7) may be transformed to relations which determine uniquely the functions B_ℓ, $\ell = 1,\ldots,k$, i.e.

$$(2.9) \qquad \sum_{\ell=1}^{k} q_{\ell-1}^{j-1}\, B_\ell(z) = D_j(z),\ j = 1,\ldots,k,$$

$$(2.10) \qquad D_1(z) = \frac{R(z)-1}{z},$$

$$(2.11) \qquad D_{j+1}(z) = \frac{jD_j(z)-1}{z},\ j = 1,\ldots,k-1.$$

In fact, the k-step scheme of order k with consistency conditions (2.7)-(2.8) is completely determined by the adaptive stability function R.

For our analysis of S-stability and stiff-accuracy we need results on the *asymptotic behaviour of the rational functions* B_ℓ, $\ell = 1,\dots,k$, and C_j, $j = k,k+1,\dots$, as $z \to \infty$. Denote the k-th order stability function R with

$$(2.12) \qquad R(z) = (\sum_{i=0}^{m} \alpha_i z^i)/(\sum_{i=0}^{m} \beta_i z^i), \; \alpha_0 = \beta_0 = 1.$$

LEMMA 2.1. *The rational functions* R *and* B_ℓ, $\ell = 1,\dots,k$, *as defined by* (2.9), (2.10) *and* (2.11) *have the same denominator.*

PROOF. The recurrence relation (2.11) yields

$$(2.13) \qquad D_j(z) = (j-1)! \, \frac{R(z)-P_{j-1}(z)}{z^j}, \; j = 1,\dots,k,$$

where

$$(2.14) \qquad P_{j-1}(z) = 1 + z + \dots + \frac{z^{j-1}}{(j-1)!} \, .$$

The function R is consistent of order k, i.e.

$$(2.15) \qquad R(z) = P_k(z) + O(z^{k+1}) \text{ as } z \to 0.$$

Relation (2.15) implies that

$$(2.16) \qquad \begin{aligned} &R(z) - 1 = O(z), \text{ and} \\ &jD_j(z) - 1 = O(z), \; j = 1,\dots,k-1, \end{aligned}$$

as $z \to 0$. This means that

$$R(z) - 1,$$

and

$$jD_j(z) - 1,$$

have a zero at $z = 0$. Thus, according to (2.10) and (2.11), the rational functions D_j, $j = 1,\dots,k$, have the denominator of R. As the functions B_ℓ, $\ell = 1,\dots,k$, are linear combinations of the D_j, this is also valid for the B_ℓ. □

LEMMA 2.2. *If* $\beta_m \neq 0$, *we have*

(a) *All the* B_ℓ *have a zero at infinity while, for* $k \geq 2$, *at least one* B_ℓ *has a single zero at infinity.*

(b) *The functions* C_j, $j = k,k+1,\dots$, *do not have a pole at infinity.*

PROOF. If β_m is not equal to zero, we see, from the preceding proof, that

$$D_1(z) = \frac{\dots+(\alpha_m-\beta_m)z^{m-1}}{1+\beta_1 z+\dots+\beta_m z^m}, \text{ and} \tag{2.17}$$

$$D_j(z) = \frac{\dots-\beta_m z^{m-1}}{1+\beta_1 z+\dots+\beta_m z^m}, \quad j = 2,\dots,k. \tag{2.18}$$

By means of relation (2.9) we find

$$B_\ell(z) = \frac{b_\ell^{(0)}+\dots+b_\ell^{(m-1)}z^{m-1}}{1+\beta_1 z+\dots+\beta_m z^m}, \quad \ell = 1,\dots,k. \tag{2.19}$$

Because the Vandermonde Matrix

$$(q_{\ell-1}^{j-1}), \quad j,\ell = 1,\dots,k,$$

provides a non-trivial solution, at least one $b_\ell^{(m-1)}$ is not equal to zero establishing assertion (a). Assertion (b) is now easily proved by means of relations (2.6) and (2.19). □

EXAMPLE 2.3. For future reference, we give the three-step scheme generated by the third order, L-acceptable (see REMARK (2,4)) Padé approximation

$$R(z) = \frac{1+\frac{1}{3}z}{1-\frac{2}{3}z+\frac{1}{6}z^2}. \tag{2.20}$$

For constant stepsizes, i.e. $q_\ell = -\ell$, the functions B_ℓ are defined by

$$(2.21)\qquad B_1(z) = \frac{\frac{23}{12} - \frac{1}{2}z}{1 - \frac{2}{3}z + \frac{1}{6}z^2},$$

$$(2.22)\qquad B_2(z) = \frac{-\frac{4}{3} + \frac{1}{2}z}{1 - \frac{2}{3}z + \frac{1}{6}z^2},$$

$$(2.23)\qquad B_3(z) = \frac{\frac{5}{12} - \frac{1}{6}z}{1 - \frac{2}{3}z + \frac{1}{6}z^2}.$$

REMARK 2.4. In this paper we pay no special attention to rational approximations to the exponential. For all properties and definitions which are used the reader is referred to LAMBERT [10]. We will, however, give one definition: let R be A-acceptable, then R is said to be strongly A-acceptable, if

$$\lim_{Re(z)\to-\infty} |R(z)| < 1.$$

Further, we let the variable z as well as the rational functions, which we shall meet, range over the extended complex plane. If F is a rational function of which the behaviour for $Re(z) \to -\infty$ must be studied, we consider $F(1/z)$. We rewrite $F(1/z)$ as a rational function $\bar{F}(z)$ and set $F(\infty) = \bar{F}(0)$, while the order of a zero or pole at infinity is defined as the order of a zero or pole of $\bar{F}$ at the origin (see AHLFORS [1]). Moreover, we shall always assume that all the formula manipulation is legal.

3. GENERALIZED RUNGE-KUTTA METHODS

Let h_n denote the steplength $h_n = x_{n+1} - x_n$, and let y_n denote the numerical approximation to the analytical solution $y(x_n)$ of system (1.1) at $x = x_n$. Let J_n denote the *Jacobian matrix* of system (1.1) at the point

(x_n, y_n). Further, define

(3.1) $$\Lambda_{j,\ell}, \; j = 0,\dots,m; \; \ell = 0,\dots,j-1,$$

to be *rational functions* with real coefficients. Then, the *generalized m-point Runge-Kutta method* is defined by

(3.2) $$y_{n+1} = y_n + \sum_{j=0}^{m-1} \Lambda_{m,j}(h_n J_n) k_n^{(j)},$$

(3.3) $$k_n^{(j)} = h_n f(x_n + \mu_j h_n, \; y_n + \sum_{\ell=0}^{j-1} \Lambda_{j,\ell}(h_n J_n) k_n^{(\ell)}),$$

where the parameters μ_j are given by

(3.4) $$\mu_j = \sum_{\ell=0}^{j-1} \Lambda_{j,\ell}(0).$$

These one-step methods are extensively discussed in HOUWEN, VAN, DER [6].

Let us define the $m \times m$ matrix

(3.5) $$\Lambda = \begin{bmatrix} 0 & 0 & \cdots & \cdots & 0 \\ \Lambda_{1,0} & 0 & \cdots & \cdots & 0 \\ \Lambda_{2,0} & \Lambda_{2,1} & \cdot & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ \Lambda_{m-1,0} & \Lambda_{m-1,1} & \cdots & \Lambda_{m-1,m-2} & 0 \end{bmatrix},$$

and the m-vector

(3.6) $$\Lambda_m = [\Lambda_{m,0}, \Lambda_{m,1}, \dots, \Lambda_{m,m-1}]^T.$$

Then scheme (3.2) may be formally characterized by the $(m+1) \times m$ matrix

(3.7) $$\left[\frac{\Lambda}{\Lambda_m^T}\right].$$

The generalized Runge-Kutta method (3.2) is said to be *accurate of order p*, when the expansion of y_{n+1} in powers of h_n agrees with p+1 terms with the Taylor expansion of the solution y(x) about the point $x = x_n$. For establishing accuracy of order p, we need the derivatives of the functions $\Lambda_{j,\ell}$ at z = 0. Therefore we set

$$(3.8) \qquad \lambda_{j,\ell}^{(i)} = \frac{d^i}{dz^i} \Lambda_{j,\ell}(z)\Big|_{z=0}.$$

For a two-point scheme, the *consistency conditions* for p = 1 up to p = 4 are given in table 3.1 (cf. HOUWEN, VAN, DER [6]).

TABLE 3.1 Consistency conditions for scheme (3.2), m = 2.

p	
$p \geq 1$	$\lambda_{2,0}^{(0)} + \lambda_{2,1}^{(0)} = 1;$
$p \geq 2$	$\lambda_{2,0}^{(1)} + \lambda_{2,1}^{(1)} + \lambda_{2,1}^{(0)}\lambda_{1,0}^{(0)} = \frac{1}{2};$
$p \geq 3$	$\lambda_{2,0}^{(2)} + \lambda_{2,1}^{(2)} + 2(\lambda_{2,1}^{(0)}\lambda_{1,0}^{(1)} + \lambda_{2,1}^{(1)}\lambda_{1,0}^{(0)}) = \frac{1}{3},$
	$\lambda_{2,1}^{(0)}\lambda_{1,0}^{(0)2} = \frac{1}{3};$
$p \geq 4$	$\lambda_{2,0}^{(3)} + \lambda_{2,1}^{(3)} + 3(\lambda_{2,1}^{(0)}\lambda_{1,0}^{(2)} + 2\lambda_{2,1}^{(1)}\lambda_{1,0}^{(1)} + \lambda_{2,1}^{(2)}\lambda_{1,0}^{(0)}) = \frac{1}{4},$
	$\lambda_{2,1}^{(0)}\lambda_{1,0}^{(0)}\lambda_{1,0}^{(1)} = \frac{1}{8},$
	$\lambda_{2,1}^{(1)}\lambda_{1,0}^{(0)2} = \frac{1}{12},$
	$\lambda_{2,1}^{(0)}\lambda_{1,0}^{(0)3} = \frac{1}{4}.$

When applied to the scalar *test equation*

$$y' = \delta y,$$

formula (3.2) yields a relation of the type

$$y_{n+1} = R(z)y_n,$$

where $z = h_n\delta$. The function R is called the *stability function* of the integration method and is composed of the coefficient functions $\Lambda_{j,\ell}$.

In HOUWEN, VAN, DER [7], first to fourth order formulas of the class (3.2) are proposed, which possess an *adaptive stability function*. For p = 1, 2 and 3, the generating matrices of these formulas are presented below. The function $R^{(p)}$ denotes a stability function of order p.

First order methods:

(3.9)
$$\begin{bmatrix} 0 \\ \\ \dfrac{R^{(1)}(z) - 1}{z} \end{bmatrix}$$

Second order methods:

(3.10)
$$\begin{bmatrix} 0 \\ \\ \dfrac{R^{(2)}(z) - 1}{z} \end{bmatrix}$$

Third order methods:

(3.11)
$$\begin{bmatrix} 0 & 0 \\ \\ \frac{4}{3}\dfrac{R^{(3)}(z) - 1 - z}{z^2} & 0 \\ \\ \frac{1}{4} & \frac{3}{4} \end{bmatrix}$$

Two other formulas which are known in the literature, are a second order formula of ROSENBROCK [11], given by

$$(3.12)\qquad \begin{bmatrix} 0 & 0 \\ \frac{\sqrt{2}-1}{2-(2-\sqrt{2})z} & 0 \\ 0 & \frac{2}{2-(2-\sqrt{2})z} \end{bmatrix},$$

and a third order formula of CALAHAN [2], given by

$$(3.13)\qquad \begin{bmatrix} 0 & 0 \\ \frac{-8\sqrt{3}}{12-(6+2\sqrt{3})z} & 0 \\ \frac{9}{12-(6+2\sqrt{3})z} & \frac{3}{12-(6+2\sqrt{3})z} \end{bmatrix}.$$

Observe, however, that the formulas of Rosenbrock and Calahan are not adaptive.

4. S-STABILITY

Let g denote the solution vector of the *stiff* initial value problem (1.1). In the neighbourhood of $y \equiv g$, we approximate the differential system by

$$(4.1)\qquad \begin{aligned} y' &= f(x,g(x)) + J(g(x))(y-g(x)) \\ &= g'(x) + J(g(x))(y-g(x)), \end{aligned}$$

where $J(g(x))$ denotes the Jacobian matrix along the solution g. In a sufficiently small neighbourhood of $y \equiv g$, the stiffness properties of system (1.1) may be characterized by means of the linear approximation (4.1).

Let h be a relevant stepsize, i.e. for a stiff eigenvalue $\delta(x)$ of $J(g(x))$ we have $h\,\mathrm{Re}(\delta(x)) << -1$. Now we shall make two assumptions. *The first assumption* is that, to a first approximation, we may neglect the dependence of $J(g(x))$ with x on each interval $[x_n, x_n+h]$. This means that, per integration step, we may characterize the stiffness properties of system (4.1) with the linear system

$$(4.2) \qquad y' = g'(x) + J(y-g(x)),$$

where J is a constant matrix. *The second assumption* is that the matrix J may be diagonalized. Then, the numerical difficulties arising from the stiffness of system (1.1) may be analysed in terms of the *stability* and *accuracy* of numerical solutions to a single equation of the form

$$(4.3) \qquad y' = g'(x) + \delta(y-g(x)), \ \delta \text{ complex}, \ \mathrm{Re}(\delta) < 0,$$

where g' is *any given bounded function*. Equation (4.3) is considered as *a new test-equation*. For a more extensive derivation of this test-equation the reader is referred to PROTHERO & ROBINSON [9]. In particular, they assume that the eigenvalues of the Jacobian matrices under consideration are located in widely separated clusters.

The analytical solution of equation (4.3), with the initial condition $y(x_0) = y_0$, is given by

$$(4.4) \qquad y(x) = e^{\delta(x-x_0)}[y(x_0)-g(x_0)] + g(x).$$

Regardless of the initial condition at $x = x_0$, the solution (4.4) tends to $g(x)$ for $x > x_0$. Therefore, we may examine the stability of numerical approximations to the solution $y \equiv g$. A stable behaviour of the numerical solution y_{n+1} with respect to $g(x_{n+1})$ includes a stable behaviour of y_{n+1} to $y(x_{n+1})$ where $y(x)$ is given by (4.4). PROTHERO & ROBINSON [9] give the following definition of stablity.

DEFINITION 4.1. An integration method is said to be *S-stable* if, for a differential equation of the form (4.3) and for any real negative constant δ_0, a real positive constant h_0 exists, such that

$$(4.5) \qquad \left|\frac{y_{n+1} - g(x_{n+1})}{y_n - g(x_n)}\right| < 1,$$

provided $y_n \neq g(x_n)$, for all stepsizes $0 < h < h_0$ and all complex δ with $\mathrm{Re}(\delta) \leq \delta_0$.

If the conditions of this definition only hold for $|\arg(h\delta)| < \alpha$, the integration method is said to be $S(\alpha)$-stable. In fact, *the concept of S-stability generalizes the concept of A-stability*. Analogously, the following definition generalizes the concept of L-stability.

DEFINITION 4.2. An S-stable integration method is said to be *L-S-stable* if in addition

$$\frac{y_{n+1} - g(x_{n+1})}{y_n - g(x_n)} \to 0$$

as $\mathrm{Re}(\delta) \to -\infty$, for all stepsizes $h > 0$.

REMARK 4.3. Suppose $g \equiv c$, c constant. When applied to equation (4.3), the integration methods discussed in this paper result in a relation of the form

$$y_{n+1} = R(h_n\delta)y_n + (1-R(h_n\delta))c,$$

where R is the stability function of the method. Now inequality (4.5) reads

$$\left|\frac{y_{n+1} - c}{y_n - c}\right| = |R(h_n\delta)| < 1.$$

This means that *if g is a constant function, S-stability is equivalent to A-stability*. Of course, this is also valid for other types of stability.

When applied to the *S-stability equation* (4.3), each formula discussed in the preceding sections will result into a relation of the form

(4.6) $$y_{n+1} = R(h_n\delta)y_n + h_nT_n(h_n\delta,h_n;g),$$

where R is the stability function of the method. In general, the expression T_n contains values of g and g' for several arguments x. Observe, however, that all methods discussed in this paper are not implicit. As a consequence, the values $g(x_n+h_n)$ and $g'(x_n+h_n)$ do not occur in T_n.

We define (cf. (4.5))

(4.7) $$\varepsilon_n = y_n - g(x_n),$$

and derive

(4.8) $$\varepsilon_{n+1} = R(h_n\delta)\varepsilon_n + h_n[T_n(h_n\delta,h_n;g) + (R(h_n\delta)g(x_n) - g(x_{n+1}))/h_n].$$

The S-stability of the method is governed by the difference equation (4.8). To establish conditions for S-stability we use a lemma which is a slight modification of a lemma proposed by PROTHERO & ROBINSON [9].

<u>LEMMA 4.4</u>. *Let* $E = (0,\bar{h}] \subset \mathbb{R}$ *with* $\bar{h} > 0$. *Let* S *be some region in* $\mathbb{C}$. *For all pairs* $(h,z) \in E \times S$ *and all* $\varepsilon_0 \in \mathbb{C}$ *we define*

(4.9) $$\varepsilon = \alpha(z)\varepsilon_0 + h\beta(h,z).$$

Then, a real positive $h_0 = h_0(\varepsilon_0;S) \le \bar{h}$ *exists, such that*

(4.10) $$|\varepsilon| < |\varepsilon_0|$$

for all pairs $(h,z) \in (0,h_0] \times S$, *if and only if*

(a) $|\alpha(z)| < 1$ *on* S,

(b) *a real positive* $h_1 \le \bar{h}$ *exists such that* $\beta(h,z)/(1-|\alpha(z)|)$ *is uniformly bounded on* $(0,h_1] \times S$.

PROOF. Suppose conditions (a) and (b) are satisfied. Then, a positive number K_0 exists such that

$$\frac{|\beta(h,z)|}{1 - |\alpha(z)|} < K_0$$

for all $(h,z) \in (0,h_1] \times S$. Further, we have

$$|\varepsilon| \le |\alpha(z)||\varepsilon_0| + h|\beta(h,z)| =$$
$$= |\varepsilon\ | - \{1 - |\alpha(z)|\}\{|\varepsilon_0| - h|\beta(h,z)|/(1-|\alpha(z)|)\}.$$

Because $|\alpha(z)| < 1$ on S, the choice

$$h_0 = \min\{h_1, |\varepsilon_0|/K_0\},$$

yields

$$|\varepsilon| < |\varepsilon_0|$$

for all $z \in S$, all $\varepsilon_0 \neq 0$ and all $h \in (0,h_0]$.

Next we suppose condition (a) is not satisfied. Then

$$|\alpha(z^*)| \ge 1$$

for some $z^* \in S$. As a consequence, for any $h \in E$, there exists an $\varepsilon_0 \neq 0$ such that

$$|\varepsilon| \ge |\varepsilon_0|$$

at (h,z^*). This establishes a contradiction.

Finally we suppose condition (b) is not satisfied. Then, for any real positive K_1, a pair $(h^*,z^*) \in E \times S$ exists, with h^* arbitrarily small, such that

$$\frac{|\beta(h^*,z^*)|}{1 - \alpha(z^*)} > K_1.$$

By the choice

$$\varepsilon_0 = h^*(1-|\alpha(z^*)|) \neq 0,$$

we have

$$h^*|\beta(h^*,z^*)|/|\varepsilon_0| > K_1.$$

Since

$$|\varepsilon| = |\varepsilon_0|\,|\alpha(z^*) + h^*\beta(h^*,z^*)/\varepsilon_0|,$$

there exists K_1, such that for (h^*,z^*) holds

$$|\varepsilon| \geq |\varepsilon_0|.$$

This establishes a second contradiction and completes the proof of the lemma. □

For a given function g this lemma establishes conditions for the stability of solutions to equation (4.8).

THEOREM 4.5. *The integration method related to the error equation* (4.8) *is* S-*stable, if and only if*
(a) *The stability function* R *is strongly* A-*acceptable,*
(b) *A constant* $\bar{h} > 0$ *exists, such that*

$$(4.11) \qquad T_n(z,h_n;g) + (R(z)g(x_n) - g(x_n+h_n))/h_n$$

is uniformly bounded on $\{(h_n,z) \mid \mathrm{Re}(z) < 0,\ h_n \in (0,\bar{h}]\}$.

PROOF. After observing that the strong A-acceptability of R implies

$$\lim_{\mathrm{Re}(z)\to-\infty} |R(z)| < 1,$$

the theorem is easily proved by applying definition 4.1 and lemma 4.4 to the error equation 4.8. □

THEOREM 4.6. *The integration method related to the error equation* (4.8) *can not be* L-S-*stable.*

PROOF. As already observed, such a method is not implicit. This implies that the value $g(x_n+h_n)$ does not occur in T_n. A consequence is that always a function g exists, with g' bounded, such that for any $h_n > 0$,

$$\lim_{Re(z)\to-\infty} T_n(z,h_n;g) + (R(z)g(x_n) - g(x_n+h_n))/h_n \neq 0. \tag{4.12}$$

However, from equation (4.8) and definition 4.2 we see that we have L-S-stability, if and only if, for any $h_n > 0$, both R(z) and expression (4.11) tend to zero as $Re(z) \to -\infty$. From (4.12) it thus follows that the method can not be L-S-stable. □

REMARK 4.7. In the terminology of PROTHERO & ROBINSON [9], L-S-stability is called strong S-stability. They prove that certain implicit one-step formulas, which are based on Radau or Lobatto quadrature, are strongly S-stable.

REMARK 4.8. The effect of S-stability is, that for given δ, the upper bound to the range of stability h_0', which depends on δ, ε_n and g, does not tend to zero if $Re(\delta) \to -\infty$. So, in contrast with A-stability, quantitative aspects of S-stability do not depend exclusively on the integration method, but also on the differential equation. As is to be expected, S-stability is meant to compare qualitatively the stability properties of integration formulas which share certain theoretical properties. Especially, for explicit integration methods, as discussed in this paper, we have to restrict ourselves to a comparative analysis.

5. STIFF-ACCURACY

It is well-known that, when integrating stiff differential equations, the local truncation error yields no reliable information about the size of the global error (see e.g. HENRICI [5], p.4 and VELDHUIZEN, VAN, [12]). Therefore, when analyzing integration methods for stiff systems, not only the stability of the method has to be considered, but also the *accuracy*. One has to relate the stiffness of the differential equation with the accuracy of the integration formula. For our methods we shall perform the accuracy analysis, as proposed by PROTHERO & ROBINSON [9].

As $Re(\delta) \to -\infty$, the true solution to equation (4.3) tends to $g(x)$ for any fixed $x > x_0$, regardless of the initial condition at $x = x_0$. Thus, for a stiff equation, a qualitative measure of the accuracy of the integration method is provided by the difference

$$(5.1) \qquad d_n = \bar{y}_{n+1} - g(x_{n+1}),$$

where $\bar{y}_{n+1}$ denotes the numerical solution at $x = x_{n+1}$ to the initial value problem

$$(5.2) \qquad y' = g'(x) + \delta(y-g(x)),\ y(x_n) = g(x_n).$$

The error d_n is dependent on the stepsize h_n, the eigenvalue δ and the function g. We shall derive an *asymptotic relation* for d_n in the limit

$$(5.3) \qquad Re(h_n\delta) \to -\infty \text{ and } h_n \to 0,$$

which is of the form

$$(5.4) \qquad d_n \sim C_0 h_n^{p_h+1} g^{(p_h+1)}(x_n) + \sum_{j=1}^{s} C_j z^{t_j} h_n^{p_j+1} g^{(p_j+1)}(x_n),$$

where $z = h_n\delta$.

In relation (5.4) C_j, $j = 0,\ldots,s$ are real constants; p_h, p_j, $j = 1,\ldots,s$, are integers greater or equal to zero; s is an integer greater or equal to 1; t_j, $j = 1,\ldots,s$, are integers. This asymptotic relation provides a good view on the accuracy of a method when applied to stiff problems. In particular, it will appear from the results that the order of a method may be rather *misleading* for stiff problems.

To obtain accurate results it is necessary that the integers t_j are all *negative*. A method with this property shall be called *stiffly accurate*.

By putting $\varepsilon_n = 0$, i.e. $y_n = g(x_n)$, in equation (4.8), we find

(5.5) $$d_n = h_n[T_n(z,h_n;g) + (R(z)g(x_n) - g(x_n+h_n))/h_n].$$

Thus, the equation for the error ε_n may be written as

(5.6) $$\varepsilon_{n+1} = R(z)\varepsilon_n + d_n.$$

This implies that stiff-accuracy is necessary for S-stability. In theorem 4.5 we have proved that the strong A-acceptability of R is also necessary for S-stability. From equation (5.6) it is now clear, that we may expect a better S-stability behaviour if, in addition, R is L-acceptable. Because, if R is L-acceptable, the local truncation errors d_n are very quickly damped out for $Re(z) << -1$. This implies that if R is L-acceptable and $Re(z) << -1$, the global errors ε_{n+1} approximately equal the local truncation errors d_n.

6. S-STABILITY AND STIFF-ACCURACY FOR THE CLASS OF GENERALIZED MULTISTEP METHODS

When applied to the S-stability test equation (4.3), the adaptive multistep scheme (2.2) yields (compare (4.6))

(6.1) $$y_{n+1} = R(z)y_n + h_nT_n(z,h_n;g),$$

where $z = h_n\delta$ and

(6.2) $$T_n(z,h_n;g) = \sum_{\ell=1}^{k} B_\ell(z)[g'(x_{n+1-\ell}) - \delta g(x_{n+1-\ell})].$$

Assuming g to be sufficiently differentiable, we may state

THEOREM 6.1. *The adaptive multistep scheme (2.2) is* S-*stable if and only if the stability function* R *is strongly* A-*acceptable.*

PROOF. The necessity of the strong A-acceptability of R follows immediately from assertion (a) of theorem 4.5.

Suppose R is strongly A-acceptable. According to theorem 4.5, we just have to prove that a constant $\bar{h} > 0$ exists such that

$$\frac{d_n}{h_n} = \sum_{\ell=1}^{k} B_\ell(z)[g'(x_{n+1-\ell}) - \delta g(x_{n+1-\ell})] + \frac{R(z)g(x_n) - g(x_n+h_n)}{h_n},$$

is uniformly bounded on $(0,\bar{h}] \times \{z \mid Re(z) < 0\}$. From expansion (2.4) and conditions (2.7) it follows that for any z, $Re(z) < 0$, d_n may be formally expanded as

$$d_n = \sum_{j=k} [C_j(z) - 1/j!]\, h_n^j g^{(j)}(x_n), \quad h_n \to 0. \tag{6.3}$$

Because R is strongly A-acceptable the results of lemma 2.2 are valid. As a result, the functions B_ℓ, $\ell = 1,\ldots,k$, and C_j, $j = k,k+1,\ldots$, are uniformly bounded on $\{z \mid Re(z) < 0\}$. As in expression (6.3) the integer $k \geq 1$, the uniform boundedness of the B_ℓ and C_j implies that for any fixed $\bar{h} > 0$, d_n/h_n is uniformly bounded on $(0,\bar{h}] \times \{z \mid Re(z) < 0\}$. □

By means of expansion (6.3) we may readily prove

THEOREM 6.2. *If the stability function* R *is A-acceptable, the* k*-th order adaptive multistep scheme* (2.2) *is stiffly accurate, with* $p_h \geq k - 1$.

PROOF. If R is A-acceptable we may apply lemma 2.2. As the functions C_j, $j = k,k+1,\ldots$, do not have a pole at infinity, we may expand

$$C_j(z) = \bar{c}_j^{(0)} + \bar{c}_j^{(1)} z^{-1} + \ldots, \quad \text{as } Re(z) \to -\infty.$$

Substitution into expansion (6.3) yields

$$d_n = \sum_{j=k} [-1/j! + \sum_{i=0} \bar{c}_j^{(i)} z^{-i}]\, h_n^j g^{(j)}(x_n), \quad h_n \to 0 \text{ and } Re(z) \to -\infty. \tag{6.4}$$

If $\bar{c}_k^{(0)} \neq 1/k!$ and $\bar{c}_k^{(1)} \neq 0$, we have

(6.5) $$d_n \sim (\bar{c}_k^{(0)} - 1/k!)h_n^k g^{(k)}(x_n) + \bar{c}_k^{(1)} z^{-1} h_n^k g^{(k)}(x_n),$$

as $h_n \to 0$ and $\mathrm{Re}(z) \to -\infty$, establishing the proof of the theorem. □

REMARK 6.3. Observe that the adaptive multistep scheme is stiffly accurate, if it is S-stable. The converse need not be true. This depends on the stability function R.

7. S-STABILITY AND STIFF-ACCURACY FOR GENERALIZED RUNGE-KUTTA METHODS

We rewrite the S-stability test equation (4.3) as

(7.1) $$y' = \delta y + r,$$

where the function r denotes

(7.2) $$r = g' - \delta g.$$

We introduce the abbreviations

(7.3) $$r_{n,j} = r(x_n + \mu_j h_n),\ j = 0,\dots,m-1,$$

and the m-vector

(7.4) $$r_n = [r_{n,0}, r_{n,1}, \dots, r_{n,m-1}]^T.$$

Further, we define the m-vectors,

(7.5) $$e = [1,1,\dots,1]^T,$$

and, for scalar equations,

(7.6) $$k_n = [k_n^{(0)}, k_n^{(1)}, \dots, k_n^{(m-1)}]^T.$$

When applied to equation (7.1), the Runge-Kutta method (3.2) supplies the increment functions

$$(7.7)\qquad k_n^{(j)} = h_n r_{n,j} + z(y_n + \sum_{\ell=0}^{j-1} \Lambda_{j,\ell}(z) k_n^{(\ell)}),$$

where $z = h_n\delta$. Let I denote the identity matrix of order m. Then, the increment vector k_n may be expressed as (compare PROTHERO & ROBINSON [9])

$$(7.8)\qquad k_n = (I - z\Lambda(z))^{-1}[zy_n e + h_n r_n].$$

Now it is easily seen that y_{n+1} can be expressed as

$$(7.9)\qquad y_{n+1} = [1 + z\Lambda_m(z)(I - z\Lambda(z))^{-1}e]y_n + \Lambda_m(z)(I - z\Lambda(z))^{-1}h_n r_n.$$

The matrix $\Lambda(z)$ and the vector $\Lambda_m(z)$ are given by (3.5) and (3.6), respectively.

If $g \equiv 0$, equations (7.1) and (7.9) reduce to

$$y' = \delta y,$$

and

$$y_{n+1} = [1 + z\Lambda_m(z)(I - z\Lambda(z))^{-1}e]y_n,$$

respectively. As a result, the stability function of method (3.1) is given by

$$(7.10)\qquad R(z) = 1 + z\Lambda_m(z)(I - z\Lambda(z))^{-1}e.$$

We set

$$(7.11)\qquad T_n(z,h_n;g) = \Lambda_m(z)(I - z\Lambda(z))^{-1}r_n,$$

so that equation (7.9) can be rewritten as (cf. (4.6))

$$(7.12)\qquad y_{n+1} = R(z)y_n + h_n T_n(z,h_n;g).$$

Next, we transform R and T_n in such a way that the coefficient functions $\Lambda_{j,\ell}$ are easier to recognize. To that end we introduce the functions

(7.13) $$\sigma_{j,k,\ell} = \sum \Lambda_{i_{k+1},i_k} \Lambda_{i_k,i_{k-1}} \cdots \Lambda_{i_2,i_1} \Lambda_{i_1,i_0},$$

where the summation runs over all k+2-tuples $(i_{k+1}, i_k, \ldots, i_0)$ which satisfy: $j = i_{k+1} > i_k > \ldots > i_1 > i_0 = \ell$.

LEMMA 7.1. *The inverse of the matrix* $I - z\Lambda(z)$ *is given by*

(7.14) $$\eta(z) = \begin{bmatrix} 1 & 0 & 0 & & 0 \\ \eta_{1,0}(z) & 1 & 0 & & \\ \eta_{2,0}(z) & \eta_{2,1}(z) & 1 & & \\ \vdots & \vdots & & \ddots & 0 \\ \eta_{m-1,0}(z) & \eta_{m-1,1}(z) & \cdots & \eta_{m-1,m-2}(z) & 1 \end{bmatrix},$$

where

(7.15) $$\eta_{j,\ell}(z) = \sum_{k=0}^{j-1-\ell} z^{k+1} \sigma_{j,k,\ell}(z),\ j = 1,\ldots,m-1;\ \ell = 0,\ldots,j-1.$$

PROOF. The matrix $I - z\Lambda(z)$ is lower triangular with all diagonal elements equal to 1. After this observation the proof may be obtained by elementary matrix algebra. □

The inner-product

$$\Lambda_m(z)(I - z\Lambda(z))^{-1}e$$

may thus be expressed as

$$\Lambda_m(z)(I - z\Lambda(z))^{-1}e = \sum_{j=0}^{m-1} \Lambda_{m,j}(z)\ (1 + \sum_{\ell=0}^{j-1} \eta_{j,\ell}(z)),$$

so that R(z) equals,

(7.16) $$R(z) = 1 + \sum_{j=0}^{m-1} \Lambda_{m,j}(z)z + \sum_{j=1}^{m-1} \sum_{\ell=0}^{j-1} \sum_{k=0}^{j-1-\ell} \Lambda_{m,j}(z)\sigma_{j,k,\ell}(z)z^{k+2}.$$

To transform the second inner-product, i.e.

$$T_n(z,h_n;g) = \Lambda_m(z)(I - z\Lambda(z))^{-1}r_n,$$

we denote the ℓ-th column of $\eta(z)$ with $\eta_\ell(z)$, $\ell = 0,\ldots,m-1$. Then,

$$T_n(z,h_n;g) = \sum_{\ell=0}^{m-1} \Lambda_m(z)\ \eta_\ell(z)\ r_{n,\ell} =$$

$$\sum_{\ell=0}^{m-1} [\sum_{j=\ell+1}^{m-1} \Lambda_{m,j}(z)\ \eta_{j,\ell}(z) + \Lambda_{m,\ell}(z)]\ r_{n,\ell} =$$

$$\sum_{\ell=0}^{m-1} [\Lambda_{m,\ell}(z) + \sum_{j=\ell+1}^{m-1} \sum_{k=0}^{j-1-\ell} \Lambda_{m,j}(z)\ \sigma_{j,k,\ell}(z)z^{k+1}]\ r_{n,\ell}.$$

For future reference, we define

$$T_{m,\ell}(z) = \Lambda_{m,\ell}(z) + \sum_{j=\ell+1}^{m-1} \sum_{k=0}^{j-1-\ell} \Lambda_{m,j}(z)\ \sigma_{j,k,\ell}(z)z^{k+1},\ \ell = 0,\ldots,m-1, \tag{7.17}$$

so that

$$T_n(z,h_n;g) = \sum_{\ell=0}^{m-1} T_{m,\ell}(z)\ r_{n,\ell}. \tag{7.18}$$

The functions $T_{m,\ell}$ shall be useful for establishing S-stability and stiff-accuracy. Before proceeding with the theorem which establishes S-stability for our one-step methods, we give an example.

EXAMPLE 7.2. The two-point generalized Runge-Kutta scheme reads

$$\begin{aligned} y_{n+1} = y_n + h_n\Lambda_{2,0}(h_nJ_n)\ f(x_n,y_n) + \\ h_n\Lambda_{2,1}(h_nJ_n)\ f(x_n + \lambda_{1,0}^{(0)}h_n, y_n + h_n\Lambda_{1,0}(h_nJ_n)f(x_n,y_n)). \end{aligned} \tag{7.19}$$

The generating matrix (3.7), for scheme (7.19), equals

$$\left[\begin{array}{cc} 0 & 0 \\ \Lambda_{1,0} & 0 \\ \hline \Lambda_{2,0} & \Lambda_{2,1} \end{array}\right]. \tag{7.20}$$

The vector r_n is given by $r_n = [r_{n,0}, r_{n,1}]^T$, where

$$r_{n,0} = g'(x_n) - \delta g(x_n),$$

$$r_{n,1} = g'(x_n+\lambda^{(0)}_{1,0}h_n) - \delta g(x_n+\lambda^{(0)}_{1,0}h_n).$$

Expression (7.16), for the stability function R, yields

$$R(z) = 1 + (\Lambda_{2,0}(z) + \Lambda_{2,1}(z))z + \Lambda_{2,1}(z)\Lambda_{1,0}(z)z^2. \tag{7.21}$$

Finally, the expressions (7.17), for the functions $T_{m,\ell}$, yield

$$\begin{aligned} T_{2,0}(z) &= \Lambda_{2,0}(z) + z\Lambda_{2,1}(z)\Lambda_{1,0}(z), \\ T_{2,1}(z) &= \Lambda_{2,1}(z). \end{aligned} \tag{7.22}$$

THEOREM 7.3. *The generalized Runge-Kutta method* (3.2) *is* S-*stable, if and only if*

(a) *The stability function* R *is strongly* A-*acceptable,*

(b) *The rational functions* $T_{m,\ell}$, $\ell = 0,\ldots,m-1$, *have a zero at infinity.*

PROOF. The necessity and sufficiency of assertion (a) follow immediately from assertion (a) of theorem 4.5.

Suppose assertion (b) holds. Expression (4.11) now reads

$$\frac{d_n}{h_n} = \sum_{\ell=0}^{m-1} T_{m,\ell}(z)[g'(x_n+\mu_\ell h_n)-\delta g(x_n+\mu_\ell h_n)] + \frac{R(z)g(x_n)-g(x_n+h_n)}{h_n}.$$

By assuming g to be sufficiently differentiable, and by expanding $g(x_n+\mu_\ell h_n)$ and $g'(x_n+\mu_\ell h_n)$ into Taylor series, d_n may be formally expanded as

$$d_n = [R(z) - \sum_{\ell=0}^{m-1} z\, T_{m,\ell}(z) - 1]\, g(x_n) + \sum_{j=1} [\sum_{\ell=0}^{m-1} T_{m,\ell}(z)(j\mu_\ell^{j-1}-z\mu_\ell^j) - 1]\,\frac{h_n^j}{j!}\, g^{(j)}(x_n), \tag{7.23}$$

for $h_n \to 0$. From relations (7.10), (7.11), (7.18) and the choice $g \equiv 1$, it follows that

$$(7.24) \qquad R(z) - \sum_{\ell=0}^{m-1} z\, T_{m,\ell}(z) - 1 = 0.$$

As a result, (7.23) is reduced to

$$(7.25) \qquad d_n = \sum_{j=1} \Big[\sum_{\ell=0}^{m-1} T_{m,\ell}(z)(j\mu_\ell^{j-1} - z\mu_\ell^j) - 1\Big] \frac{h_n^j}{j!} g^{(j)}(x_n),$$

which implies that d_n/h_n is bounded for any fixed z, $Re(z) < 0$, as $h_n \to 0$. Moreover, the functions $z\, T_{m,\ell}(z)$, $\ell = 0,\dots,m-1$, are uniformly bounded on $\{z \mid Re(z) < 0\}$. This implies that for every fixed $\bar{h} > 0$, d_n/h_n is uniformly bounded on $(0,\bar{h}] \times \{z \mid Re(z) < 0\}$, establishing the sufficiency of assertion (b).

The necessity of assertion (b) follows trivially from the expression for d_n/h_n. □

Next we discuss the stiff-accuracy properties of the generalized Runge-Kutta method.

<u>THEOREM 7.4</u>. *The generalized Runge-Kutta method* (3.2) *is stiffly accurate, if and only if*

(a) *The rational function* $T_{m,0}$ *has no pole at infinity,*

(b) *The rational functions* $T_{m,\ell}$, $\ell = 1,\dots,m-1$, *have a zero at infinity.*

<u>PROOF</u>. Because $\mu_0 = 0$, the expansion (7.25), for the error d_n, reads

$$d_n = \Big[\sum_{\ell=0}^{m-1} T_{m,\ell}(z) - \sum_{\ell=1}^{m-1} z\, T_{m,\ell}(z)\mu_\ell - 1\Big] h_n g^{(1)}(x_n) +$$

$$\sum_{j=2} \Big[\sum_{\ell=1}^{m-1} T_{m,\ell}(z)(j\mu_\ell^{j-1} - z\mu_\ell^j) - 1\Big] \frac{h_n^j}{j!} g^{(j)}(x_n).$$

If at least one $T_{m,\ell}$, $\ell = 1,\dots,m-1$, has no zero at infinity, there exists an integer j_0, such that

$$\sum_{\ell=1}^{m-1} T_{m,\ell}(z)(j_0\mu_\ell^{j_0-1} - z\mu_\ell^{j_0})$$

has a pole at infinity. In such a situation the method can not be stiffly accurate. This establishes the necessity of hypothesis (a) and (b).

If hypothesis (a) and (b) are satisfied, we have the expansions

$$(7.26) \qquad T_{m,\ell}(z) = t_{m,\ell}^{(0)} + t_{m,\ell}^{(1)}z^{-1} + t_{m,\ell}^{(2)}z^{-2} + \ldots , \text{ as } z \to \infty,$$

where

$$t_{m,\ell}^{(0)} = 0, \ \ell = 1,\ldots,m-1.$$

Substitution into the expansion for d_n yields

$$(7.27) \qquad d_n = [-1 + \sum_{i=0} (\sum_{\ell=0}^{m-1} t_{m,\ell}^{(i)} - \sum_{\ell=1}^{m-1} \mu_\ell t_{m,\ell}^{(i+1)})z^{-i}] \, h_n g^{(1)}(x_n) +$$

$$\sum_{j=2} [-1 + \sum_{i=0} (\sum_{\ell=1}^{m-1} (j\mu_\ell^{j-1} t_{m,\ell}^{(i)} - \mu_\ell^j t_{m,\ell}^{(i+1)}))z^{-i}] \frac{h_n^j}{j!} g^{(j)}(x_n),$$

as $h_n \to 0$ and $\mathrm{Re}(z) \to -\infty$. This expansion establishes the stiff-accuracy. □

<u>REMARK 7.5</u>. Observe that the generalized one-step scheme is stiffly accurate if it is S-stable. The converse need not be true. This depends on $T_{m,0}$ and R.

Let us define the constants K_1 and K_2 by

$$(7.28) \qquad K_1 = -1 + t_{m,0}^{(0)} - \sum_{\ell=1}^{m-1} \mu_\ell \, t_{m,\ell}^{(1)},$$

and

$$(7.29) \qquad K_2 = \sum_{\ell=0}^{m-1} t_{m,\ell}^{(1)} - \sum_{\ell=1}^{m-1} \mu_\ell \, t_{m,\ell}^{(2)}.$$

If $K_1 \neq 0$ and $K_2 \neq 0$, it is easily seen that

$$d_n \sim K_1 h_n g^{(1)}(x_n) + K_2 z^{-1} h_n g^{(1)}(x_n) \text{ as } h_n \to 0 \text{ and } \mathrm{Re}(z) \to -\infty.$$

As far as we know, there does not yet exist generalized Runge-Kutta methods with the property

$$K_1 = K_2 = 0.$$

This implies that the existing methods share a rather disappointing accuracy behaviour for stiff problems. For all existing methods we have $p_h = 0$. Therefore, it seems worthwhile to develop new formulas with a better stiff-accuracy behaviour. In the next section we propose such a formula.

8. AN S-STABLE GENERALIZED RUNGE-KUTTA FORMULA

In this section we shall consider the two-point scheme (7.19), i.e.

$$(8.1)\qquad y_{n+1} = y_n + h_n\Lambda_{2,0}(h_nJ_n)\, f(x_n,y_n) + h_n\Lambda_{2,1}(h_nJ_n)\, f(x_n+\lambda_{1,0}^{(0)}h_n, y_n+h_n\Lambda_{1,0}(h_nJ_n)f(x_n,y_n)).$$

An additional property which we shall demand from the method is that it will not become useless when the *Jacobian matrix is inaccurately evaluated.* For scheme (8.1) this means that the order has to remain two in case of inaccurately evaluated Jacobians, while the order equals three in case of accurate Jacobians. To obtain this property we have to impose the consistency conditions (see HOUWEN, VAN, DER [6])

$$(8.2)\qquad \lambda_{2,0}^{(0)} = \frac{1}{4},\ \lambda_{2,1}^{(0)} = \frac{3}{4},\ \lambda_{1,0}^{(0)} = \frac{2}{3},$$

$$(8.3)\qquad \lambda_{2,0}^{(1)} = -\frac{1}{4} + \frac{3}{4}(\lambda_{2,0}^{(2)} + \lambda_{2,1}^{(2)}) + \frac{9}{8}\lambda_{1,0}^{(1)}.$$

$$(8.4)\qquad \lambda_{2,1}^{(1)} = -\lambda_{2,0}^{(1)}.$$

For practical reasons, we shall further demand that the rational functions $\Lambda_{j,\ell}$ share the same denominator of degree two. Thus, we denote

(8.5) $$\Lambda_{2,0}(z) = \frac{\frac{1}{4} + a_0 z}{1+b_1 z+b_2 z^2},$$

(8.6) $$\Lambda_{2,1}(z) = \frac{\frac{3}{4} + a_1 z}{1+b_1 z+b_2 z^2},$$

(8.7) $$\Lambda_{1,0}(z) = \frac{\frac{2}{3} + a_2 z}{1+b_1 z+b_2 z^2},$$

where $b_2 \neq 0$. By means of expressions (7.22), we see that assertion (b) of the S-stability theorem 7.3 is satisfied. Further, from theorem (7.4) it follows that the method, generated by (8.5), (8.6) and (8.7) is stiffly accurate.

Next, we shall demand that the numbers K_1 and K_2, defined by (7.28) and (7.29), respectively, equal zero, i.e.

(8.8) $$-1 + t^{(0)}_{2,0} - \mu_1 t^{(1)}_{2,1} = 0,$$

(8.9) $$t^{(1)}_{2,0} + t^{(1)}_{2,1} - \mu_1 t^{(2)}_{2,1} = 0.$$

From (7.26) we see that $t^{(0)}_{2,0} = 0$, and $\mu_1 = \Lambda_{1,0}(0) = \frac{2}{3}$, so that

(8.10) $$t^{(1)}_{2,1} = -\frac{3}{2}.$$

From expressions (7.22), (7.26) and (8.6) it follows that

(8.11) $$t^{(1)}_{2,1} = \frac{a_1}{b_2}, \quad t^{(2)}_{2,1} = \frac{1}{b_2}\left(\frac{3}{4} - \frac{b_1}{b_2} a_1\right),$$

hence

(8.12) $$a_1 = -\frac{3}{2} b_2,$$

and

(8.13) $$t_{2,1}^{(2)} = \frac{\frac{3}{2} b_1 + \frac{3}{4}}{b_2}, \quad t_{2,0}^{(1)} = \frac{3}{2} + \frac{2b_1+1}{2b_2}.$$

From expressions (7.22), (7.26) and (8.4)-(8.6) it follows that

(8.14) $$t_{2,0}^{(1)} = \frac{a_0}{b_2} + \frac{a_1 a_2}{b_2^2}.$$

Then, from condition (8.13), we find

(8.15) $$a_0 - \frac{3}{2} a_2 = \frac{3}{2} b_2 + b_1 + \frac{1}{2}.$$

As a result, if the parameters a_0, a_1, and a_2 satisfy relations (8.12) and (8.15), we have $K_1 = K_2 = 0$.

Our next step is to satisfy the consistency conditions (8.3) and (8.4). To that end we calculate

$$\lambda_{2,0}^{(1)} = a_0 - \frac{b_1}{4}, \quad \lambda_{2,0}^{(2)} = \frac{1}{2} b_1^2 - \frac{1}{2} b_2 - 2b_1 a_0,$$

(8.16) $$\lambda_{2,1}^{(1)} = a_1 - \frac{3}{4} b_1, \quad \lambda_{2,1}^{(2)} = \frac{3}{2} b_1^2 - \frac{3}{2} b_2 - 2b_1 a_1,$$

$$\lambda_{1,0}^{(1)} = a_2 - \frac{2}{3} b_1.$$

From condition (8.4) it follows that

(8.17) $$a_0 = b_1 - a_1 = b_1 + \frac{3}{2} b_2,$$

so that relation (8.15) yields

(8.18) $$a_2 = -\frac{1}{3}.$$

Further, from condition (8.3) we find

(8.19) $$b_2 = -\frac{5}{24} - \frac{b_1}{2}.$$

As a result, there is left one free parameter b_1, while the others are given by

$$(8.20)\qquad \begin{aligned} a_0 &= -\frac{5}{16} + \frac{1}{4} b_1, \quad a_1 = \frac{5}{16} + \frac{3}{4} b_1, \\ a_2 &= -\frac{1}{3}, \quad b_2 = -\frac{5}{24} - \frac{1}{2} b_1. \end{aligned}$$

For this set of parameters the stability function R is given by (cf. (7.21))

$$(8.21)\qquad R(z) = \frac{1+(1+2b_1)z+(\frac{1}{12}+b_1+b_1^2)z^2-(\frac{1}{4}+\frac{5}{12}b_1)z^3-(\frac{35}{576}+\frac{1}{4}b_1+\frac{1}{4}b_1^2)z^4}{(1+b_1z-(\frac{5}{24}+\frac{1}{2}b_1)z^2)^2},$$

and is consistent of order three. In this paper we shall use the parameter b_1 to satisfy the relation $R(\infty) = 0$. Then an elementary calculation yields $b_1 = -7/12$, and R is given by

$$(8.22)\qquad R(z) = \frac{144 - 24z - 23z^2 - z^3}{(z-3)^2(z-4)^2}.$$

According to assertion (a) of theorem 7.3, R has to be strongly A-acceptable to obtain S-stability. This property may be established by means of the maximum modulus theorem in complex analysis (see AHLFORS [1]).

LEMMA 8.1. *The rational function* (8.22) *is* L-*acceptable.*

PROOF. The rational function (8.22) is L-acceptable, if
(a) R is analytic for all z with $\mathrm{Re}(z) \le 0$,
(b) $\lim R(z) = 0$ as $\mathrm{Re}(z) \to -\infty$,
(c) $|R(z)| \le 1$ for all z with $\mathrm{Re}(z) = 0$.
Conditions (a) and (b) are trivially satisfied, while condition (c) may be verified by an elementary calculation. □

As L-acceptability implies strongly A-acceptability, assertion (a) of theorem 7.3 is satisfied.

The last step is to determine the asymptotic relation for the error d_n. Expansion (7.27) and relations (8.8), (8.9) yield

$$(8.23) \qquad d_n \sim -\frac{1}{2}(1+\mu_1^2 t_{2,1}^{(1)})\, h_n^2 g^{(2)}(x_n) + (t_{2,0}^{(2)} + t_{2,1}^{(2)} - \mu_1 t_{2,1}^{(3)}) z^{-2} h_n g^{(1)}(x_n) + \frac{1}{2}(2\mu_1 t_{2,1}^{(1)} - \mu_1^2 t_{2,1}^{(2)}) z^{-1} h_n^2 g^{(2)}(x_n),$$

as $\mathrm{Re}(z) \to -\infty$ and $h_n \to 0$, where $\mu_1 = \lambda_{1,0}^{(0)}$. To determine the right hand side of (8.23) we still need

$$(8.24) \qquad t_{2,1}^{(3)} = \frac{-a_1 - b_1(\frac{3}{4}+\frac{3}{2}b_1)}{b_2^2},$$

and

$$(8.25) \qquad t_{2,0}^{(2)} = \frac{\frac{2}{3}a_1 + \frac{3}{4}a_2 + 3b_1 a_2}{b_2^2}.$$

Substitution of the parameter values (8.20), with $b_1 = -7/12$, yields

$$(8.26) \qquad d_n \sim -\frac{1}{6} h_n^2 g^{(2)}(x_n) + \frac{118}{144} z^{-2} h_n g^{(1)}(x_n) - \frac{2}{3} z^{-1} h_n^2 g^{(2)}(x_n),$$

as $\mathrm{Re}(z) \to -\infty$ and $h_n \to 0$.

Summarized, the matrix

$$(8.27) \qquad \begin{bmatrix} 0 & 0 \\ \dfrac{\frac{2}{3}-\frac{1}{3}z}{1-\frac{7}{12}z+\frac{1}{12}z^2} & 0 \\ \dfrac{\frac{1}{4}-\frac{11}{24}z}{1-\frac{7}{12}z+\frac{1}{12}z^2} & \dfrac{\frac{3}{4}-\frac{1}{8}z}{1-\frac{7}{12}z+\frac{1}{12}z^2} \end{bmatrix}$$

generates an *S-stable, third order* scheme of the class (8.1), which is *stiffly accurate* with the asymptotic error relation (8.26). The stability function of the scheme is *L-acceptable*. The scheme remains consistent of order two, in case of *inaccurately evaluated Jacobian matrices*.

In a forthcoming paper the author intends to present numerical results of this scheme when applied to a number of stiff non-linear systems.

REMARK 8.2. As scheme (8.1) is not implicit, i.e. $\mu_1 < 1$, we see from relations (8.8) and (8.23), that for such a scheme always holds $p_h \leq 1$.

REMARK 8.3. After this construction it is clear that it is not possible to develop generalized methods of the class (8.1), which are adaptive, such as the formulas of van der Houwen, and which

(a) satisfy consistency conditions (8.2)-(8.4),
(b) are S-stable,
(c) provide an asymptotic relation for d_n of the form

$$d_n \sim K_1 h_n^2 g^{(2)}(x_n) + K_2 z^{-2} h_n g^{(1)}(x_n) + K_3 z^{-1} h_n^2 g^{(2)}(x_n),$$

as $h_n \to 0$ and $\mathrm{Re}(z) \to -\infty$.

9. DISCUSSION

In this last section we shall investigate some of the schemes, earlier mentioned, with respect to their S-stability and stiff-accuracy. Further, we shall apply these methods to a test problem of the form (4.3) and to a non-linear scalar equation. Here we emphasize once more that, especially for explicit integration methods, S-stability is meant to compare qualitatively the stability properties of methods which share certain theoretical properties.

I. *The multistep scheme generated by* (2.20)-(2.23).

According to theorem 6.1 the scheme is S-stable and thus stiffly accurate. By means of relation (6.6), the asymptotic relation for the error d_n is found to be

$$d_n \sim h_n^3 g^{(3)}(x_n) - \frac{4}{3} z^{-1} h_n^3 g^{(3)}(x_n).$$

II. *The one-step scheme generated by* (3.11).

For this scheme the functions $T_{2,0}$ and $T_{2,1}$ are given by

$$T_{2,0}(z) = R^{(3)}(z) - z - \frac{3}{4},$$

$$T_{2,1}(z) = \frac{4}{3}\,\frac{R^{(3)}(z)-1-z}{z^2}.$$

Regardless of the choice for $R^{(3)}$, the function $T_{2,0}$ has a pole at infinity. According to theorem (7.3) and (7.4), the scheme is neither S-stable nor stiffly accurate. If $R^{(3)}$ is given by (2.20), we obtain a third order, L-stable scheme. Presently we shall apply this scheme to the test problem (9.1) and to the non-linear scalar equation (9.7).

III. *The one-step scheme generated by* (3.13).

The functions $T_{2,0}$ and $T_{2,1}$ are given by

$$T_{2,0}(z) = \frac{108 - (27+18\sqrt{3})z}{(12-(6+2\sqrt{3})z)^2},$$

$$T_{2,1}(z) = \frac{-8\sqrt{3}}{12 - (6+2\sqrt{3})z},$$

and the stability function equals

$$R(z) = \frac{1 - 0.578z - 0.456z^2}{1 - 1.578z + 0.622z^2}.$$

From theorem (7.3) and (7.4) we see that the method is S-stable and stiffly accurate. The numbers K_1 and K_2, defined by (7.28) and (7.29), respectively, are given by

$$K_1 = \frac{10 - 2\sqrt{3}}{6 + 2\sqrt{3}},$$

$$K_2 = \frac{1383 + 748\sqrt{3}}{320 + 184\sqrt{3}}.$$

The asymptotic relation for d_n may thus be written as

$$d_n \sim K_1 h_n g^{(1)}(x_n) + K_2 z^{-1} h_n g^{(1)}(x_n).$$

IV. *The one-step scheme generated by* (8.27).

According to the results of section 8, this scheme is S-stable and stiffly accurate. The asymptotic relation for d_n is given by (8.26).

From these methods it is clear, that for stiff problems, the effective order of a method is generally lower than for non-stiff problems. This change in the effective order will result in a lower level of accuracy for stiff problems. Especially the generalized Runge-Kutta methods share this disadvantage. Further, observe that the multistep scheme I and the one-step scheme II possess both the stability function (2.20). In illustration of our results, we apply the methods I-IV to two scalar equations.

EXAMPLE A. The first equation is the test equation (cf. (4.3))

$$(9.1)\qquad \begin{aligned} &y' = g' + \delta(y-g),\ y(0) = g(0),\ \delta \text{ real},\\ &g(x) = 10 - (10+x)e^{-x}, \end{aligned}$$

with the solution y g. The integration has been performed over the interval [0,1], with the stepsize $h = 0.1$, for four values of δ: $\delta = -10000$, -1000, -10, -1. In the tables of results, we give the number of significant digits, i.e.

$$(9.2)\qquad -{}^{10}\log\left|1 - \frac{y_n}{g(x_n)}\right|.$$

The additional starting values for the three-step scheme are chosen equal to the values g(x) at x = -0.2, -0.1, 0.

Table 9.1, $\delta = -10000$

x	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
I	2.1	2.4	2.6	2.7	2.8	2.9	3.0	3.1	3.2	3.3
II	-2.7	-2.4	-2.2	-2.0	-1.9	-1.8	-1.7	-1.7	-1.6	-1.5
III	-0.1	0.9	0.5	1.0	0.8	1.1	1.0	1.2	1.2	1.4
IV	1.8	2.2	2.4	2.5	2.6	2.7	2.8	2.9	3.0	3.0

Table 9.2, $\delta = -1000$

x	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
I	2.1	2.4	2.6	2.7	2.8	2.9	3.0	3.1	3.2	3.3
II	-1.7	-1.4	-1.2	-1.0	-0.9	-0.8	-0.7	-0.6	-0.6	-0.5
III	-0.1	0.8	0.5	1.0	0.8	1.1	1.1	1.2	1.2	1.3
IV	1.9	2.2	2.4	2.5	2.6	2.7	2.8	2.9	3.0	3.0

Table 9.3, $\delta = -10$

x	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
I	2.7	3.0	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.8
II	0.9	1.1	1.2	1.3	1.5	1.6	1.6	1.7	1.8	1.9
III	0.3	0.5	0.7	0.8	0.9	1.0	1.1	1.2	1.2	1.3
IV	1.9	2.1	2.2	2.4	2.5	2.5	2.6	2.7	2.8	2.9

Table 9.4, $\delta = -1$

x	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
I	4.5	4.8	5.0	5.1	5.3	5.4	5.5	5.6	5.6	5.7
II	3.7	3.7	3.8	3.8	3.8	3.8	3.9	3.9	3.9	3.9
III	2.1	2.1	2.1	2.1	2.2	2.2	2.2	2.2	2.3	2.3
IV	4.5	4.5	4.5	4.6	4.6	4.6	4.7	4.7	4.7	4.7

From the tables of results we may conclude that

(a) the lower accuracy for stiff eigenvalues is clearly exhibited by all methods, even for $\delta = -10$;

(b) the lack of stiff-accuracy for method II causes unacceptable results;

(c) for method III, the change of the effective order is too large ($p_h=0$), and, compared with the two other one-step schemes, the method is also inaccurate for non-stiff eigenvalues;

(d) the new one-step method IV is, for non-stiff eigenvalues, also more accurate than the two other one-step methods.

EXAMPLE B. The second example originates from the stiff system (GEAR [4])

$$
(9.3)\qquad
\begin{aligned}
y_1' &= -0.013\, y_1 - 1000\, y_1 y_3,\\
y_2' &= -2500\, y_2 y_3,\\
y_3' &= -0.013\, y_1 - 1000\, y_1 y_3 - 2500\, y_2 y_3,
\end{aligned}
$$

with initial values $y_1(0) = y_2(0) = 1$ and $y_3(0) = 0$. For system (9.3) holds

$$y_3' = y_1' + y_2',$$

so that, by virtue of the initial conditions,

$$y_1 + y_2 - y_3 = 2.$$

By eliminating y_3 we obtain

$$
(9.4)\qquad
\begin{aligned}
y_1' &= -.013\, y_1 - 1000\, y_1(y_1+y_2-2),\\
y_2' &= -2500\, y_2(y_1+y_2-2).
\end{aligned}
$$

A first analysis of system (9.4) reveals that

$$(9.5)\qquad 2.5 \frac{d}{dx} \ln(y_1) - \frac{d}{dx} \ln(y_2) = -0.0325.$$

Then, with the initial conditions $y_1(0) = y_2(0) = 1$, we find

(9.6) $$y_2 = y_1^{2.5} e^{0.0325}.$$

Substitution of the right hand side of (9.6) into the first equation of system (9.4) yields the single equation

(9.7) $$y_1' = -0.013\, y_1 - 2500\, y_1(y_1 + y_1^{2.5} e^{0.0325x} - 2),\quad y_1(0) = 1.$$

We have integrated equation (9.7) over the interval [0,1]. On this integration interval, the problem has a very smooth solution; however, it is also very stiff. The eigenvalue varies from -8750 at $x = 0$ to -8782 at $x = 1$. The integration has been performed for four values of the stepsize h: $h = 0.005, 0.01, 0.05, 0.1$. The additional starting values for the three-step scheme are computed by a second order starting mechanism (see VERWER [13]). In the table of results we give the number of significant digits (see (9.2)), while a dash stands for an unstable result. The reference solution used at $x = 1$ is given by $y_1(1) = 0.9906310343$.

Table 9.5

h	0.005	0.01	0.05	0.1
I	10.0	10.0	9.9	9.0
II	3.0	-	-	-
III	4.5	4.1	3.4	3.1
IV	7.0	7.3	6.6	5.7

Again we may conclude, from table 9.5, that the lack of S-stability and stiff-accuracy for method II causes unacceptable results, while the new method IV is more accurate than the two other one-step methods. The very accurate results of the multistep method I are due to the particularly smooth behaviour of the solution $y_1(x)$.

REFERENCES

[1] AHLFORS, L.V., *Complex Analysis*, 2nd edition, McGraw-Hill, New York, 1966.

[2] CALAHAN, D.A., *A stable, accurate method of numerical integration for non-linear systems*, Proc. IEEE 56, 744-747, 1968.

[3] DAHLQUIST, G.G., *A special stability problem for linear multistep methods*, BIT, 3, 1963, pp.27-43.

[4] GEAR, C.W., *Numerical integration of stiff ordinary differential equations*, Rep. 221, Dept. of Computer Science, Univ. of Illinois, Urbana, 1967.

[5] HENRICI, P., *Error propagation for difference methods*, The SIAM Series in Applied Mathematics, John Wiley and Sons, Inc., New York, London, 1963.

[6] HOUWEN, P.J. VAN DER, *Construction of integration formulas for initial value problems*, North-Holland Publishing Company, Amsterdam, (to be published).

[7] HOUWEN, P.J. VAN DER, *One-step methods with adaptive stability functions for the integration of differential equations*, Lecture Notes in Mathematics 333, pp. 164-174, Springer-Verlag, Berlin, 1972.

[8] HOUWEN, P.J. VAN DER & VERWER J.G., *Generalized linear multistep methods I, Development of algorithms with zero-parasitic roots*, Report NW 10/74, Mathematisch Centrum, Amsterdam, 1974.

[9] PROTHERO, A. & ROBINSON, A., *On the stability and accuracy of one-step methods for solving stiff systems of ordinary differential equations*, Math. Comp., 28, No 125, pp.145-162, 1974.

[10] LAMBERT, J.D., *Computational methods in ordinary differential equations*, John Wiley and Sons, London, 1973.

[11] ROSENBROCK, H.H., *Some general implicit processes for the numerical solution of differential equations*, Comput. J., 5, pp.329-330, 1963.

[12] VELDHUIZEN, VAN, M., *Convergence of one-step discretizations for stiff differential equations* (Thesis), Mathematical Institute, Univ. of Utrecht, Budapestlaan 6, Utrecht, Netherlands.

[13] VERWER, J.G., *Generalized linear multistep methods II: Numerical applications*, Report NW 12/74, Mathematisch Centrum, Amsterdam, 1974.