Proceedings of the International Symposium
On Numerical Analysis
Middle East Technical University
Ankara, Turkey, September 1-4, 1987

Convergence of Runge-Kutta methods on classes of stiff initial value problems

Willem H. Hundsdorfer
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

For certain stiff initial value problems the order of convergence of implicit Runge-Kutta methods can be much lower than for nonstiff problems. In this paper we consider for some classes of stiff initial value problems convergence results which are independent of the stiffness, such as the *B*-convergence results for nonlinear dissipative problems.

Mathematics Subject Classification: Primary 65L05, Secondary 65M20.
Keywords and Phrases: Numerical analysis, stiff initial value problems, implicit Runge-Kutta methods, *B*-convergence.

Note: Paper for the International Symposium on Numerical Analysis, Ankara, Turkey, September 1-4, 1987. The text is largely based on joint work with K. Burrage and J.G. Verwer [3], [4].

1. Introduction

1.1. Convergence on Classes of Problems

Consider an initial value problem

$$u'(t) = f(t,u(t)) \quad (0 \leqslant t \leqslant 1)\,, \quad u(0) = u_0 \tag{1.1}$$

where $u_0 \in \mathbb{C}^m$ and $f:[0,1]\times\mathbb{C}^m \to \mathbb{C}^m$ are given. Let $h>0$ be a stepsize, and $t_n = nh (n=0,1,2,...)$. Using an implicit Runge-Kutta method, approximations u_n to $u(t_n)$ are computed recursively by

$$u_{n+1} = u_n + h\sum_{i=1}^{s} b_i f(t_n + c_i h, u_{in})\,, \tag{1.2a}$$

$$u_{in} = u_n + h\sum_{j=1}^{s} a_{ij} f(t_n + c_j h, u_{jn}) \quad (i=1,2,...,s). \tag{1.2b}$$

Here $s \in \mathbb{N}$ and the a_{ij}, b_i, c_i are real parameters. For convenience it will be assumed that all c_i satisfy $0 \leqslant c_i \leqslant 1$. The internal vectors u_{in}, defined by (1.2b), can be regarded as approximations to $u(t_n + c_i h)$.

In the following, the Euclidean inner product on $\mathbb{C}^m$ will be denoted by (v,w), and $|v| = (v,v)^{1/2}$ stands for the corresponding norm. The induced spectral norm for $m \times m$ matrices H is denoted by $\|H\|$. Further we shall use

$$\|u\|^{(r)} = \max\{|u^{(j)}(t)| : 0 \leqslant t \leqslant 1,\ j = 0,1,...,r\}$$

as a measure for the smoothness of the solution u. We shall be concerned with error bounds of the form

$$|u(t_n) - u_n| \leqslant Ch^p \|u\|^{(r)} \quad (\text{for } 0 \leqslant t_n \leqslant 1,\ 0 < h \leqslant \bar{h}) \tag{1.3}$$

where $C, \bar{h} > 0$ and $r \in \mathbb{N}$ are not affected by stiffness or the dimension of the initial value problem. Let $\mathcal{P}$ stand for a class of initial value problems of the form (1.1). The Runge-Kutta method is said to be convergent of order p on $\mathcal{P}$ if there exist $C, \bar{h} > 0, r \in \mathbb{N}$ such that (1.3) holds whenever

$u \in C^{(r)}[0,1]$ is a solution of a problem in $\mathcal{P}$ and the u_n are computed from (1.2). Here it is essential that (1.3) should hold uniformly on $\mathcal{P}$, not only for each individual problem in $\mathcal{P}$. By the order of the Runge-Kutta method on $\mathcal{P}$ we shall always mean the largest p for which (1.3) holds uniformly on $\mathcal{P}$. Usually a method is said to have order p if a bound (1.3) is valid individually for each problem where f is smooth and satisfies a Lipschitz condition. We will refer to this as the classical order.

In the following we shall consider some classes $\mathcal{P}$ which contain problems with arbitrarily large Lipschitz constants. The error bounds we thus obtain are not affected by stiffness. These results are also relevant for the numerical solution of time dependent partial differential equations by the method of lines. Suppose, for example, that (1.1) stands for a semi-discrete (space-discretized) PDE, in which case f will contain negative powers $(\Delta x)^{-k}$ of the meshwidth in space. If $\mathcal{P}$ contains all the problems with Δx ranging from 0 to some $(\Delta x)_{\max}$, then convergence on $\mathcal{P}$ with order p guarantees that these negative powers are not present in the error bound (1.3), and we can then prove convergence to the PDE solution without restrictions on the ratio $h/(\Delta x)^k$ (cf. [14], [17], [18]).

Let α,β,γ be real parameters with $\alpha,\gamma \geqslant 0$. By $\mathcal{W}(\beta)$ we shall denote the well known class of non-linear test problems (1.1) (see [9] for example) where $m \in \mathbb{N}, u_0 \in \mathbb{C}^m$ are arbitrary and $f:[0,1]\times\mathbb{C}^m \to \mathbb{C}^m$ satisfies

$$\mathrm{Re}\langle f(t,\tilde{v})-f(t,v),\tilde{v}-v\rangle \leqslant \beta|\tilde{v}-v|^2 \tag{1.4}$$

for all $t \in [0,1]$ and $\tilde{v},v \in \mathbb{C}^m$. Further we shall consider the class $\mathcal{S}(\alpha,\beta,\gamma)$ of semi-linear problems where f can be written as

$$f(t,v) = H(t)v + g(t,v) \quad \text{for } t \in [0,1],\ v \in \mathbb{C}^m, \tag{1.5}$$

with $H(t) \in L(\mathbb{C}^m)$ and $g:[0,1]\times\mathbb{C}^m \to \mathbb{C}^m$ such that

$$|g(t,\tilde{v})-g(t,v)| \leqslant \alpha|\tilde{v}-v|, \tag{1.6a}$$

$$\mathrm{Re}\langle H(t)v,v\rangle \leqslant \beta|v|^2, \tag{1.6b}$$

$$\|(I-\tau H(t))^{-1}(H(t+\tau)-H(t))\| \leqslant \gamma \quad (\text{for } 0 \leqslant \tau < \bar{\tau}, t+\tau \leqslant 1) \tag{1.6c}$$

for all $\tilde{v},v \in \mathbb{C}^m$ and $t \in [0,1]$, with $\bar{\tau}$ such that $1-\bar{\tau}\beta \geqslant 0$. We note that $\mathcal{S}(\alpha,\beta,\gamma) \subset \mathcal{W}(\alpha+\beta)$.

Convergence on $\mathcal{W}(\beta)$ for certain Runge-Kutta methods was proved by FRANK, SCHNEID and UEBERHUBER [11], who used the term B-convergence. In this paper some results on the order on $\mathcal{W}(\beta)$ and $\mathcal{S}(\alpha,\beta,\gamma)$ of [3], [4] and [8] will be reviewed and slightly generalized. The results, which are refinements on the theory of FRANK et al., can be found in section 2, together with some remarks on their practical relevance. We note that all results remain valid if we consider real initial value problems (1.1), where $u_0 \in \mathbb{R}^m$ and $f:[0,1]\times\mathbb{R}^m \to \mathbb{R}^m$.

1.2. The Stage Order

If we insert an exact solution of (1.1) into the Runge-Kutta scheme (1.2), we get

$$u(t_{n+1}) = u(t_n) + h\sum_{i=1}^{s} b_i u'(t_n+c_i h) + r_{on}, \tag{1.7a}$$

$$u(t_n+c_i h) = u(t_n) + h\sum_{j=1}^{s} a_{ij} u'(t_n+c_j h) + r_{in} \quad (i=1,2,\ldots,s) \tag{1.7b}$$

with residual errors r_{in} $(i=0,1,\ldots,s)$. The stage order is defined by

$$q = \min\{q_0,q_1,\ldots,q_s\} \tag{1.8}$$

where the q_i are the largest numbers such that $r_{in} = O(h^{q_i+1})$ $(h \downarrow 0)$ for any smooth solution u. This stage order q, introduced in [11], plays a central role in the convergence results for stiff problems.

Let $A=(a_{ij})$ be the $s\times s$ matrix containing the coefficients of the Runge-Kutta method, $b=(b_1,b_2,\ldots,b_s)^T$ and $c^j=(c_1^j,c_2^j,\ldots,c_s^j)^T$ for $j \in \mathbb{N}$. By a Taylor series development it is easily seen

that q is the largest integer for which the following two simplifying order conditions hold,

$$B(q): \quad b^T c^{j-1} = 1/j \quad (j=1,2,...,q),$$

$$C(q): \quad Ac^{j-1} = c^j/j \quad (j=1,2,...,q),$$

and that

$$r_{in} = d_i h^{q+1} u^{(q+1)}(t_n) + d_i' h^{q+2} u^{(q+2)}(t_n) + \cdots \quad (i=0,1,...,s) \tag{1.9}$$

with error constants

$$d_0 = \frac{1}{q!}\left(\frac{1}{q+1} - b^T c^q\right), \tag{1.10a}$$

$$d = (d_1,d_2,...,d_s)^T = \frac{1}{q!}\left(\frac{1}{q+1} c^{q+1} - Ac^q\right). \tag{1.10b}$$

The stage order is often considerably lower than the classical order. For example, the methods based on Gauss quadrature have classical order $2s$ and stage order s, and the diagonally implicit Runge-Kutta methods with $a_{11} \neq 0$ can have only stage order 1. More details can be found in [9; section 7.3].

2. The convergence Results

Convergence on $\mathcal{S}(\alpha,\beta,\gamma)$ can be derived from (1.2), (1.7) in a similar way as done in [11] for the class $\mathcal{N}(\beta)$. Under suitable internal stability assumptions on the method, which means that one step of the Runge-Kutta process (1.2) is not unduely disturbed by small perturbations on the internal stages (1.2b), it can be shown that the discretization error which is introduced in one step is of $O(h^{q+1})$ for smooth solutions, uniformly on the problem class. By a usual stability argument the bound (1.3) can then be obtained with order $p = q$. Numerical experiments in [9] and [17] show however that in many cases (1.3) holds with $p = q+1$. This discrepancy, which is due to cancellation and damping effects, will be the main matter of interest here.

In order to formulate our results we define for $Z = diag(\zeta_1,\zeta_2,...,\zeta_s)$, $\zeta_j \in \mathbb{C}$,

$$K(Z) = 1 + b^T Z(I-AZ)^{-1} e, \tag{2.1}$$

$$L(Z) = d_0 + b^T Z(I-AZ)^{-1} d \tag{2.2}$$

where I is the $s \times s$ identity matrix, $e = (1,1,...,1)^T \in \mathbb{R}^s$ and $d_0 \in \mathbb{R}$, $d \in \mathbb{R}^s$ are defined by (1.10).

The Class $\mathcal{S}(\alpha,\beta,\gamma)$. Let $\mathcal{C} = \{\zeta I : \zeta \in \mathbb{C}, \mathrm{Re}\,\zeta \leq 0\}$. We note that $K(\zeta I)$ is the stability function of the method (1.2), so that A-stability is equivalent with

$$|K(Z)| \leq 1 \quad \text{for all} \quad Z \in \mathcal{C}. \tag{2.3}$$

Besides this it will be assumed that $I - AZ$ is regular for $Z \in \mathcal{C}$, and all elements of

$$(I-AZ)^{-1} \text{ and } b^T Z(I-AZ)^{-1} \text{ are uniformly bounded for } Z \in \mathcal{C}. \tag{2.4}$$

These are internal stability assumptions, and for most methods which are A-stable condition (2.4) is also satisfied, for example for any method where all eigenvalues of A have positive real part (see [4], [6] for more details).

Theorem 2.1. *Let* $\alpha,\gamma \geq 0$ *and* $\beta \in \mathbb{R}$. *Suppose the Runge-Kutta method is A-stable and satisfies (2.4). Suppose also* $L(Z) \not\equiv 0$ *for some* $Z \in \mathcal{C}$. *Then the method is convergent on* $\mathcal{S}(\alpha,\beta,\gamma)$ *with order*

$$p = q+1 \quad \textit{if } C(\mathcal{C}) = \sup\{|(1-K(Z))^{-1} L(Z)| : Z \in \mathcal{C}\} < \infty,$$

$$p = q \quad \textit{otherwise.}$$

In case $L(Z)\equiv 0$ on $\mathcal{C}$ oders larger than $q+1$ might be possible, but it is not clear whether there are methods of practical significance with this property. The above theorem slightly generalizes results of [8] on the necessity of $C(\mathcal{C})<\infty$ for having order $q+1$, and of [4] where the sufficiency of the conditions for convergence on the class of semi-linear problems with constant linear part $\mathcal{S}(\alpha,\beta,0)$ was proved. The proof of this theorem will be given in section 3.

For most A-stable methods we have $C(\mathcal{C})<\infty$, for example for all those methods which satisfy (2.4) and $K(\zeta I)\neq 1$ for $\zeta=i\eta$, $\eta\in\mathbb{R}\cup\{\infty\}, \eta\neq 0$ (see [4]). An important exception to this are the Gauss-methods with $s\geqslant 2$, for which $C(\mathcal{C})=\infty$, and since they have stage order s, we have $p=s$ for these methods in theorem 2.1. This result is due to [8]. We note that the Radau II-A methods also have stage order s, but $C(\mathcal{C})<\infty$ if $s\geqslant 2$, and thus these methods are convergent with order $p=s+1$ on $\mathcal{S}(\alpha,\beta,\gamma)$ (cf. [4]).

REMARK 2.2. For the subclass $\mathcal{S}_0(\alpha,\beta,\gamma)$, consisting of the problems in $\mathcal{S}(\alpha,\beta,\gamma)$ with Hermitian matrix $H(t)$ $(0\leqslant t\leqslant 1)$, the results of theorem 2.1 are still valid if we consider instead of $C(\mathcal{C})$ the supremum of $|(1-K(Z))^{-1}L(Z)|$ over $\mathcal{C}_0=\{\xi I : \xi\in\mathbb{R}, \xi\leqslant 0\}$. On this subclass all Gauss methods with an odd number of stages are convergent with order $s+1$, whereas the order is still s if s is even (this is due to the fact that $K(\infty I)=1$ for s even). For convergence on $\mathcal{S}_0(\alpha,\beta,\gamma)$ it is sufficient that (2.3), (2.4) hold with $\mathcal{C}$ replaced by $\mathcal{C}_0$. These results can be easily proved along the lines of the proof for $\mathcal{S}(\alpha,\beta,\gamma)$ (see section 3).

The Class $\mathcal{N}(\beta)$. Let $\mathcal{B}=\{diag(\zeta_1,\zeta_2,...,\zeta_s) : \zeta_j\in\mathbb{C}, \mathrm{Re}\,\zeta_j\leqslant 0 \text{ for } j=1,2,...,s\}$. Convergence on $\mathcal{N}(\beta)$ will be discussed here for Runge-Kutta methods which satisfy nonlinear analogues of (2.3), (2.4). We now assume that the method is B-stable, which can be expressed by the condition $|K(Z)|\leqslant 1$ (for all $Z\in\mathcal{B}$), or more conveniently by the algebraic (stability) condition

$$B>0\,,\ BA+A^TB+bb^T\geqslant 0 \tag{2.5}$$

(cf. [2], [5], [9]). Here $B=diag(b_1,b_2,....,b_s)$ and $>0(\geqslant 0)$ refers to positive (semi-)definiteness. In order to ensure internal stability on $\mathcal{N}(\beta)$ (called BSI-and BS-stability in [10]), it will be assumed that there exists a positive definite diagonal matrix D such that

$$DA+A^TD\ >0. \tag{2.6}$$

This assumption implies that $I-AZ$ is regular for $Z\in\mathcal{B}$, and that $(I-AZ)^{-1}$ and $b^TZ(I-AZ)^{-1}$ are uniformly bounded for $Z\in\mathcal{B}$ (see for instance [12; lemma 2.4.3]), and hence it can be considered as the counterpart of (2.4). Condition (2.6) holds for most well known B-stable methods, though not for the Lobatto III-C methods with $s\geqslant 3$ (see [7], [9]; convergence results for these Lobatto methods can be found in [15], [16]).

THEOREM 2.3. *Let $\beta\in\mathbb{R}$. Suppose the Runge-Kutta method satisfies (2.5) and (2.6). Then the method is convergent on $\mathcal{N}(\beta)$ with order*

$$p=q+1 \quad \text{if } C(\mathcal{B})=\sup\{|(1-K(Z))^{-1}L(Z)| : Z\in\mathcal{B}\}<\infty,$$

$$p\in[q,q+1) \quad \text{if } C(\mathcal{B})=\infty \text{ and } c_i-c_j\notin\mathbb{Z} \text{ for } i\neq j.$$

We note that the assumption $L(Z)\neq 0$ for some $Z\in\mathcal{B}$ (cf. theorem 2.1) can be omitted here, because this is already implied by our assumption (2.5) (see for example [3], lemma 2.3 and its proof). The most important fact in the above theorem, convergence with order $p\geqslant q$, was proved in [11]. A proof of convergence with order 2 for the implicit midpoint rule, which has stage order 1, can be found in [13]. The remaining results were proved in [3] for $\beta=0$, and this proof can be easily extended for arbitrary $\beta\in\mathbb{R}$. In [3] it was implicitly assumed that the order p is always integer valued, in which case $p\in[q,q+1)$ implies $p=q$. This is not correct. It was shown by K. DEKKER (private communications) that fractional orders may occur here as well.

Although the theorems 2.1 and 2.3 look very similar, the outcome is quite different. Whereas we

have order $q+1$ on $\mathcal{S}(\alpha,\beta,\gamma)$ for most A-stable methods, the order $q+1$ result on $\mathcal{W}(\beta)$ only holds for few B-stable methods. We have, see [3],

$$C(\mathcal{B})<\infty \quad \textit{iff} \quad d_0=0 \quad \text{and} \quad d_1=d_2=\cdots=d_s.$$

This holds for the implicit midpoint rule (the Gauss method with $s=1$), but for methods with $s>1$ the condition $d_1=d_2=\cdots=d_s$ is very restrictive. It is satisfied by some special low order methods only, see [3].

Generally speaking, we have order $q+1$ on $\mathcal{S}(\alpha,\beta,\gamma)$ and order $p<q+1$ on $\mathcal{W}(\beta)$, under suitable stability assumptions. (The Gauss-methods are an exception to this rule.) At first sight the class $\mathcal{W}(\beta)$ looks more relevant than $\mathcal{S}(\alpha,\beta,\gamma)$ for practical problems, most of which are nonlinear after all. However, in the numercal experiments in [9], [17] orders less than $q+1$ are not observed (except for the 2-stage Gauss method), which indicates that the results of theorem 2.3 are often too pessimistic. Probably this is caused by the fact that $\mathcal{W}(\beta)$ contains some extreme problems, such as $u'(t)=\lambda(t)u(t)+g(t)$ with $\lambda(t)$ very wildly oscillating in the range $\{\zeta:\zeta\in\mathbf{C},\mathrm{Re}\,\zeta\leqslant\beta\}$. It seems that for stiff problems (1.1) with $\partial f(t,u)/\partial u$ rather slowly varying along the solution $u(t)$, the results of theorem 2.1 are more relevant than those of theorem 2.3.

REMARK 2.4. In the classes $\mathcal{S}(\alpha,\beta,\gamma)$ and $\mathcal{W}(\beta)$ all derivatives of $f(t,v)$ are allowed to be arbitrarily large. For stiff problems $\partial f(t,v)/\partial v$ is always large, since the norm of this derivative is only bounded by the Lipschitz constant of f, but for certain stiff problems some of the other derivatives may be of moderate size. If this is taken into account orders larger than q or $q+1$ can be obtained. Convergence results with order $p=q+2$ were derived by CROUZEIX, cf. [1], [6], for Runge-Kutta methods applied to linear problems with constant coefficients in $\mathcal{S}(0,0,0)$, where

$$f(t,v) = Hv + g(t)$$

and derivatives of $g(t)$ are bounded (by a moderate constant, which then enters into the estimate for the global error). We note that the results in [6] were formulated for arbitrary Banach spaces. Equations of this type arise for instance by discretization in space of initial-boundary value problems for linear partial differential equations with constant boundary values; for time dependent boundary values some derivatives of g will be very large, see e.g. [14]. In [11] it was shown that, for certain Runge-Kutta methods of Gauss, Radau or Lobatto type, the discretization error introduced in one step (1.2) can be bounded by Ch^{q+2} for nonlinear problems in $\mathcal{W}(\beta)$, where C then depends on bounds for certain derivatives of f, but not on the Lipschitz constant. This leads to convergence with order $p\in[q+1,q+2]$ ($q+2$ if there is sufficient damping or cancellation).

3. DERIVATION OF THE CONVERGENCE RESULTS FOR $\mathcal{S}(\alpha,\beta,\gamma)$

3.1. Preliminaries

We shall consider here only scalar equations in $\mathcal{S}(\alpha,\beta,\gamma)$. This case contains already all essential difficulties. By using the material presented in [4], where convergence on $\mathcal{S}(\alpha,\beta,0)$ was proved, the extension to systems can be easily obtained.

In this section some technical results will be given. We use the following notations, $z_{0n}=hH(t_n)$, $Z_{0n}=z_{0n}I$ with I the $s\times s$ identity matrix, and $Z_n=diag(z_{1n},z_{2n},\ldots,z_{sn})$ with

$$z_{in} = hH(t_n+c_ih) + h\int_0^1 \frac{\partial}{\partial u} g(t_n + c_ih, u_{in} + \theta(u(t_n+c_ih)-u_{in}))d\theta.$$

The set $\{\zeta\in\mathbf{C}:\mathrm{Re}\,\zeta\leqslant 0\}$ will be be denoted by $\mathbf{C}^-$. Further $C,C_0,C_1,\ldots$ and $\bar{h}$ will stand for positive constants which depend exclusively on α,β,γ and the coefficients of the Runge-Kutta method.

LEMMA 3.1. *Assume (2.3), (2.4). There are C_0,C_1,C_2 and $\bar{h}>0$ such that for $0<h\leqslant\bar{h}$ the matrix*

$I-AZ_n$ is regular, and

$$\|(I-AZ_n)^{-1}\| \leqslant C_0, \tag{3.1}$$

$$|b^T Z_n (I-AZ_n)^{-1} r| \leqslant C_1 |r| \quad (\textit{for all } r \in \mathbb{C}^s), \tag{3.2}$$

$$|1 + b^T Z_n (I-AZ_n)^{-1} e| \leqslant 1 + C_2 h. \tag{3.3}$$

PROOF. We have

$$I-AZ_n = (I-AZ_{0n})\,(I-(I-AZ_{0n})^{-1} A(Z_n - Z_{0n})). \tag{3.4}$$

We note that (2.4) implies that there exists an $\omega>0$ such that $I-A\zeta$ is regular for all $\zeta \in \mathbb{C}$ with $\mathrm{Re}\,\zeta \leqslant \omega$. Therefore $I-AZ_{0n} = I-Az_{0n}$ is regular provided that $h\beta \leqslant \omega$. Further we have

$$(I-AZ_{0n})^{-1} A(Z_n - Z_{on}) = \{(I-AZ_{0n})^{-1} A(I-Z_{0n})\}\,\{(I-Z_{0n})^{-1}(Z_n - Z_{0n})\}.$$

The matrix $(I-AZ_{0n})^{-1}A(I- Z_{0n}) = (I-Az_{0n})^{-1}(A-I)+I$ is bounded uniformly for $\mathrm{Re}\, z_{0n} \leqslant \omega$, and the assumptions (1.6) imply that, for $h\beta \leqslant 1$,

$$\|(I-Z_{0n})^{-1}(Z_n - Z_{0n})\| \leqslant Ch \tag{3.5}$$

for some $C>0$. Regularity of $I-AZ_n$ and (3.1) now follow from (3.4).

In order to prove (3.2) and (3.3) we note that

$$Z_n(I-AZ_n)^{-1} - Z_{0n}(I-AZ_{0n})^{-1} = Z_n(I-AZ_n)^{-1} - (I-Z_{0n}A)^{-1} Z_{0n} =$$
$$= (I-Z_{0n}A)^{-1}\{(I-Z_{0n}A)Z_n - Z_{0n}(I-AZ_n)\}(I-AZ_n)^{-1} =$$
$$= (I-AZ_{0n})^{-1}(Z_n - Z_{0n})(I-AZ_n)^{-1}.$$

Hence

$$b^T Z_n (I-AZ_n)^{-1} = b^T Z_{0n}(I-AZ_{0n})^{-1} + \tag{3.6}$$
$$+ \{b^T (I-AZ_{0n})^{-1}(I-Z_{0n})\}\{(I-Z_{0n})^{-1}(Z_n - Z_{0n})\}(I-AZ_n)^{-1}.$$

The second term on the right hand side can be bounded by Ch for some $C>0$, by using (2.4), (3.1) and (3.5). The inequalities (3.2) and (3.3) now easily follow from (2.3) and (2.4). □

In the following lemma we consider some rational function whose coefficients are determined by those of the Runge-Kutta method.

LEMMA 3.2. *Let ψ be a rational function which is bounded on $\mathbb{C}^-$. Then there are $C_3, \bar{h} > 0$ such that, for $0 < h \leqslant \bar{h}$, we have $|\psi(z_{0n})| \leqslant C_3$ and*

$$|\psi(z_{0n+1}) - \psi(z_{0n})\ | \leqslant C_3 h.$$

PROOF. Let $\omega>0$ be such that ψ is bounded on $\{\zeta \in \mathbb{C}: \mathrm{Re}\,\zeta \leqslant \omega\}$. We assume $h\beta \leqslant \omega$, so that $\psi(z_{0n})$ is bounded. Write ψ in irreducible form as

$$\psi(\zeta) = \nu \prod_{j=1}^{\sigma} \psi_j(\zeta)\,,\ \psi_j(\zeta) = (1+\mu_j\zeta)^{-1}\,(1+\lambda_j\zeta).$$

We shall prove, by induction to the degree σ, that there is a $C>0$ such that

$$|\psi(\tilde{\zeta}) - \psi(\zeta)| \leqslant C|(1-\zeta)^{-1}(\tilde{\zeta} - \zeta)|$$

for all $\tilde{\zeta}, \zeta \in \mathbb{C}$ with real part $\leqslant \omega$. The proof of the lemma then follows from (1.6c).

First assume $\sigma = 1$. Then

$$\psi(\tilde{\zeta}) - \psi(\zeta) = \nu\,(1+\mu_1\zeta)^{-1}\,\{(1+\mu_1\zeta)(1+\lambda_1\tilde{\zeta}) - (1+\lambda_1\zeta)\,(1+\mu_1\tilde{\zeta})\}(1+\mu_1\tilde{\zeta})^{-1} =$$

$$= \nu(\lambda_1-\mu_1)\{(1+\mu_1\zeta)^{-1}(1-\zeta)\}\{(1-\zeta)^{-1}(\tilde{\zeta}-\zeta)\}(1+\mu_1\tilde{\zeta})^{-1}.$$

Both $(1+\mu_1\zeta)^{-1}(1-\zeta)$ and $(1+\mu_1\zeta)^{-1}$ are uniformly bounded for $\mathrm{Re}\zeta\leqslant\omega$, so that the inequality for $\sigma=1$ follows. For $\sigma>1$ we have

$$\psi(\tilde{\zeta})-\psi(\zeta) = \nu\{\psi_1(\tilde{\zeta})-\psi_1(\zeta)\}\prod_{j=2}^{\sigma}\psi_j(\tilde{\zeta}) + \nu\psi_1(\zeta)\{\prod_{j=2}^{\sigma}\psi_j(\tilde{\zeta})-\prod_{j=2}^{\sigma}\psi_j(\zeta)\},$$

and the proof follows by induction. □

3.2. Order results for global errors

By subtraction of (1.2) from (1.7) one arrives after some manipulation (see [4]) at the following recursion for the global errors $\epsilon_n=u(t_n)-u_n$ $(n=0,1,2,...)$,

$$\epsilon_{n+1} = (1+b^TZ_n(I-AZ_n)^{-1}e)\,\epsilon_n + b^TZ_n(I-AZ_n)^{-1}r_n + r_{0n}. \tag{3.7}$$

Here

$$r_{0n} = d_0h^{q+1}u^{(q+1)}(t_n) + h^{q+2}\rho_{0n}, \tag{3.8a}$$

$$r_n = dh^{q+1}u^{(q+1)}(t_n) + h^{q+2}\rho_n \tag{3.8b}$$

with $d_0\in\mathbb{R}, d\in\mathbb{R}^s$ given by (1.10), and with remainder terms ρ_{0n} and $\rho_n=(\rho_{1n},\rho_{2n},...,\rho_{sn})^T$ such that $|\rho_{in}|\leqslant C_4\|u\|^{(q+2)}$ $(i=0,1,...,s)$ for some $C_4>0$. Defining $\sigma_n=b^TZ_n(I-AZ_n)^{-1}\rho_n+\rho_{0n}$, we obtain by using the functions $K(Z)$ and $L(Z)$, introduced in section 2,

$$\epsilon_{n+1} = K(Z_n)\epsilon_n + \delta_n \tag{3.9a}$$

where

$$\delta_n = L(Z_n)h^{q+1}u^{(q+1)}(t_n) + h^{q+2}\sigma_n, \tag{3.9b}$$

and, in view of (3.2),

$$|\sigma_n|\leqslant C_4\,(1+C_1)\|u\|^{(q+2)}. \tag{3.9c}$$

Here and in the following it is assumed that $0<h\leqslant\bar{h}$ with $\bar{h}>0$ such that the estimations of section 3.1 can be applied.

The above recursion will be used to derive convergence results for methods satisfying (2.3) and (2.4). From lemma 3.1 we directly obtain

$$|\epsilon_{n+1}|\leqslant(1+C_2h)|\epsilon_n| + C_5h^{q+1}\|u\|^{(q+2)} \tag{3.10}$$

for some $C_5>0$. This leads, in a standard way, to convergence on $\mathcal{S}(\alpha,\beta,\gamma)$ with order $p\geqslant q$. By a more careful analysis of (3.9) we can prove an order $q+1$ result, provided that the stability factor $K(Z_n)$ and local error δ_n are related in the following way:

there are ξ_n,η_n (for $n=0,1,2,...$) and $C_6>0$ such that

$$\delta_n = (1-K(Z_n))\xi_n + \eta_n \tag{3.11}$$

with $|\xi_n|\leqslant C_6h^{q+1}\|u\|^{(q+2)}, |\eta_n|\leqslant C_6h^{q+2}\|u\|^{(q+2)}$ and $|\xi_{n+1}-\xi_n|\leqslant C_6h^{q+2}\|u\|^{(q+2)}$.

LEMMA 3.3. *Consider (3.9a), and assume that (3.3) and (3.11) hold. Then there are $C_7,\bar{h}>0$ such that $|\epsilon_n|\leqslant C_7h^{q+1}\|u\|^{(q+2)}$ (for all $n\geqslant 0$, $0\leqslant t_n\leqslant 1$, $0<h\leqslant\bar{h}$).*

PROOF. Define $\hat{\epsilon}_n=\epsilon_n-\xi_n$ for $n=0,1,2,...$. Then we have for all n

$$\hat{\epsilon}_{n+1} = K(Z_n)\hat{\epsilon}_n + \eta_n + \xi_n-\xi_{n+1}.$$

Hence

$$|\hat{\epsilon}_{n+1}| \leq (1+C_2 h)|\hat{\epsilon}_n| + 2C_6 h^{q+2}\|u\|^{(q+2)},$$

which leads to the global estimate

$$|\hat{\epsilon}_n| \leq e^{C_2 t_n}|\hat{\epsilon}_0| + 2C_6 C_2^{-1}(e^{C_2 t_n}-1)h^{q+1}\|u\|^{(q+2)} \quad (n=0,1,2,\ldots).$$

Since $|\hat{\epsilon}_n - \epsilon_n| \leq C_6 h^{q+1}\|u\|^{(q+1)}$ (for all n), by assumption, the proof follows. □

This result and its proof were inspired by the analysis of KRAAIJEVANGER [13] for the implicit midpoint rule. As we shall see later on, the assumption (3.11) is also necessary for having order $q+1$ for problems (3.12), where $K(Z_n)$ and δ_n are constant. First it will be shown that boundedness of $(1-K(Z))^{-1}L(Z)$ on $\mathcal{C}$ is sufficient for (3.11) to hold.

In view of (3.9b) we can write δ_n in the form $(1-K(Z_n))\xi_n + \eta_n$ with

$$\xi_n = (1-K(Z_{0n}))^{-1}L(Z_{0n})h^{q+1}u^{(q+1)}(t_n),$$

$$\eta_n = \{L(Z_n) - (1-K(Z_n))\ (1-K(Z_{0n}))^{-1}L(Z_{0n})\}h^{q+1}u^{(q+1)}(t_n) + h^{q+2}\sigma_n.$$

The bounds in (3.11) for $|\xi_n|$ and $|\xi_{n+1}-\xi_n|$ follow by applying lemma 3.2 with $\psi(\zeta)=(1-K(\zeta I))^{-1}L(\zeta I)$. Further we have

$$\eta_n = \{L(Z_n) - L(Z_{0n}) + (K(Z_n)-K(Z_{0n}))(1-K(Z_{0n}))^{-1}L(Z_{0n})\}h^{q+1}u^{(q+1)}(t_n) + h^{q+2}\sigma_n,$$

and since $|L(Z_n)-L(Z_{0n})|$ and $|K(Z_n)-K(Z_{0n})|$ can be bounded by Ch for some $C>0$ (see (3.6)), the bound required for $|\eta_n|$ in (3.11) follows as well.

We have thus proved that any Runge-Kutta method satisfying (2.3) and (2.4) is convergent on $\mathcal{S}(\alpha,\beta,\gamma)$ with order $p \geq q+1$ if $C(\mathcal{C})<\infty$, and $p \geq q$ otherwise. In order to show that the order cannot be larger than $q+1, q$, respectively, it is sufficient to deal with the case $\beta<0$. We will proceed as in [8] (cf. remark 3.4 below). We consider the model problems

$$u'(t) = \lambda u(t) + \{w'(t) - \lambda w(t)\}\ ,\ u(0) = w(0) \tag{3.12}$$

where $w(t)=t^{q+1}/(q+1)!$ and $\lambda \in \mathbb{C}$ with $\mathrm{Re}\lambda \leq \beta < 0$. The solution is $u=w$. For these problems (3.9) reduces to

$$\epsilon_{n+1} = K(z_0 I)\epsilon_n + L(z_0 I)h^{q+1} \quad \text{with } z_0 = h\lambda \tag{3.13}$$

If $L(\zeta_0 I) \neq 0$ for some $\zeta_0 \in \mathbb{C}^-$, we see from (3.13) with $n=0$ and $\lambda = h^{-1}\zeta_0 + \beta$, that the order is at most $q+1$.

Now assume $(1-K(\zeta_0 I))^{-1}L(\zeta_0 I)=\infty$ for some $\zeta_0 \in \mathbb{C}^- \cup \{\infty\}$. Suppose also $K(\zeta I) \not\equiv 1$ (otherwise the method cannot be convergent at all). Since $L(\zeta I)$ is uniformly bounded for $\zeta \in \mathbb{C}^-$, ζ_0 must be a zero of $1-K(\zeta I)$. From the A-stability assumption, $|K(\zeta I)| \leq 1$ on $\mathbb{C}^-$, it can be concluded that $1-K(\zeta I)$ can have only simple zeros on $\mathbb{C}^-$, see e.g. [12; lemma 2.2.2]. Also $\zeta=\infty$ can be at most a simple zero of $1-K(\zeta I)$, as can be seen by applying the same reasoning to $M(\zeta I)=1+b^T(\zeta I - A)^{-1}e$ $(=K(\zeta^{-1}I))$ near $\zeta=0$. Thus we have

$$K(\zeta_0 I) = 1\ ,\ L(\zeta_0 I) \neq 0. \tag{3.14}$$

We take in (3.12)

$$\lambda = h^{-1}\zeta_0 + \beta \quad \text{if } \zeta_0 \text{ is finite},$$

$$\lambda = -h^{-2} \quad \text{if } \zeta_0 = \infty.$$

Then we obtain for $z_0 = h\lambda$

$$L(z_0 I) = L(\zeta_0 I) + O(h)\ ,\ K(z_0 I) = 1 + \kappa_0 h + O(h^2) \quad (h \downarrow 0)$$

with $\kappa_0 = \beta K'(\zeta_0 I)$ if ζ_0 finite, and $\kappa_0 = -M'(0)$ if $\zeta_0 = \infty$ (in both cases $\kappa_0 \neq 0$). From (3.13) it follows that

$$\epsilon_n = (K(z_0 I)^n - 1)\ (K(z_0 I) - 1)^{-1} L(z_0 I) h^{q+1},$$

which leads to the asymptotic expression for $h \downarrow 0$, $t_n = nh$ fixed,

$$\epsilon_n \sim ((1+\kappa_0 h)^n - 1)(\kappa_0 h)^{-1} L(\zeta_0 I) h^{q+1} \sim$$
$$\sim \kappa_0^{-1} (\exp(\kappa_0 t_n) - 1) L(\zeta_0 I) h^q.$$

This completes the proof of theorem 2.1.

REMARK 3.4. It was already shown by DEKKER, KRAAIJEVANGER and SPIJKER [8; lemma 3.1] that we have order $p \leqslant q+1$ on $\mathcal{S}(\alpha,0,\gamma)$ if $L(\zeta I) \neq 0$. The above proof for $\beta < 0$ is a trivial extension of this. In the same paper, also lemma 3.1, it was proved that (3.14) implies that the order on $\mathcal{S}(\alpha,0,\gamma)$ cannot be larger than q.

REFERENCES

1. P. BRENNER, M. CROUZEIX and V. THOMÉE: *Single step methods for inhomogeneous linear differential equations in Banach space.* RAIRO Numer. Anal. 16 (1982), 5-26.
2. K. BURRAGE and J.C. BUTCHER: *Stability criteria for implicit Runge-Kutta methods.* SIAM J. Numer. Anal. 16 (1979), 46-57.
3. K. BURRAGE and W.H. HUNDSDORFER: *The order of B-convergence of algebraically stable Runge-Kutta methods.* BIT 27 (1987), 62-71.
4. K. BURRAGE, W.H. HUNDSDORFER and J.G. VERWER: *A study of B-convergence of Runge-Kutta methods.* Computing 36 (1986), 17-34.
5. M. CROUZEIX: *Sur la B-stabilité des méthodes de Runge-Kutta.* Numer. Math. 32 (1979), 75-82.
6. M. CROUZEIX and P.A. RAVIART: *Méthodes de Runge-Kutta.* Lecture Notes, Université de Rennes, 1980.
7. K. DEKKER and E. HAIRER: *A necessary condition for BSI-stability.* BIT 25 (1985), 35-41.
8. K. DEKKER, J.F.B.M. KRAAIJEVANGER and M.N. SPIJKER: *The order of B-convergence of the Gaussian Runge-Kutta methods.* Computing 36 (1986), 35-41.
9. K. DEKKER and J.G. VERWER: *Stability of Runge-Kutta methods of stiff nonlinear differential equations.* North-Holland, Amsterdam, 1984.
10. R. FRANK, J. SCHNEID and C.W. UEBERHUBER: *Stability properties of implicit Runge-Kutta methods.* SIAM J. Numer. Anal. 22 (1985), 497-514.
11. R. FRANK, J. SCHNEID and C.W. UEBERHUBER: *Order results for implicit Runge-Kutta methods applied to stiff systems.* SIAM J. Numer. Anal. 22 (1985), 515-543.
12. W.H. HUNDSDORFER: *The numerical solution of stiff initial value problems.* CWI Tract 12, Centre for Math. and Comp. Sc., Amsterdam, 1985.
13. J.F.B.M. KRAAIJEVANGER: *B-convergence of the implicit midpoint rule and the trapezoidal rule.* BIT 25 (1985), 652-666.
14. J.M. SANZ-SERNA, J.G. VERWER and W.H. HUNDSDORFER: *Convergence and order reduction of Runge-Kutta schemes applied to evolutionary problems in partial differential equations.* Numer. Math. 50 (1987), 405-418.
15. J. SCHNEID: *B-convergence of Lobatto III C formulas.* Numer. Math. 51 (1987), 229-235.
16. M.N. SPIJKER: *The relevance of algebraic stability in implicit Runge-Kutta methods.* Proc. Halle Conf., K. Strehmel (ed.), Teubner Verlag, Leipzig, 1986.
17. J.G. VERWER: *Convergence and order reduction of diagonally implicit Runge-Kutta methods.* Proc. Dundee Conf. 1985, D.F. Griffiths (ed.), Pitman Publ. Co., 1986.
18. J.G.VERWER and J.M. SANZ-SERNA: *Convergence of method of lines approximations to partial differential equations.* Computing 33 (1984), 297-313.