



ELSEVIER

Applied Numerical Mathematics 22 (1996) 309–325



APPLIED
NUMERICAL
MATHEMATICS

Error propagation in Runge–Kutta methods

M.N. Spijker¹

*Department of Mathematics and Computer Science, University of Leiden, Niels Bohrweg 1, 2333 CA Leiden,
The Netherlands*

Abstract

This paper deals with the general (explicit or implicit) Runge–Kutta method for the numerical solution of initial value problems. We consider how perturbations (like rounding errors), introduced in the consecutive time steps of the method, are propagated during the calculations.

The paper reviews various classical estimates, obtained since Runge (1905), of the accumulated error generated by all perturbations together. These estimates are compared with each other, regarding their behaviour as the step size h tends to zero.

Two new theorems are presented which extend and improve some of the classical results. The first of these theorems gives an upperbound for the norm of the accumulated error, whereas the second shows that this upperbound is maximally refined with respect to its behaviour as h tends to zero.

The paper also reviews various error estimates, obtained since the seventies, which are still relevant in some notorious (stiff) cases where h does not necessarily tend to zero and where the estimates mentioned above significantly overestimate the actual error.

The focus in the paper is on estimating the accumulated error with an arbitrary norm in \mathbb{R}^s (not necessarily generated by an inner product).

Keywords: Ordinary differential equations; Initial value problems; Runge–Kutta methods; Error propagation; Error bounds; Numerical stability

1. Introduction

1.1. The initial value problem

This paper deals with the numerical solution of the initial value problem

$$U'(t) = f(t, U(t)) \quad \text{for } 0 \leq t \leq T, \quad U(0) = u_0. \quad (1.1)$$

Here u_0 is a given vector in the s -dimensional real vector space \mathbb{R}^s , $s \geq 1$, and $U(t) \in \mathbb{R}^s$ is unknown for $0 < t \leq T$. Further, f denotes a given continuous mapping from $\mathbb{R} \times \mathbb{R}^s$ to \mathbb{R}^s .

¹ E-mail: spijker@wi.leidenuniv.nl.

Throughout the paper we denote by $|x|$ an arbitrary norm for the vectors $x \in \mathbb{R}^s$. If $\langle x, y \rangle$ is an inner product in \mathbb{R}^s , and $|x| = \sqrt{\langle x, x \rangle}$ (for all $x \in \mathbb{R}^s$), we say that the norm is *generated by an inner product*. In this paper we focus on the case where $|x|$ is *not* necessarily generated by an inner product.

We define L to be the smallest value with

$$|f(t, y) - f(t, x)| \leq L \cdot |y - x| \quad (\text{for all real } t \text{ and } x, y \in \mathbb{R}^s).$$

In the following this value is assumed to be finite, and is called the *Lipschitz constant* of f .

1.2. Runge–Kutta methods

Let N denote a positive integer. We consider the *step size* $h = T/N$ and corresponding *gridpoints* $t_n = nh$ for $n = 0, 1, \dots, N$. The *general Runge–Kutta method* can provide us with numerical approximations u_n to $U(t_n)$. For $1 \leq n \leq N$ the vector $u_n \in \mathbb{R}^s$ is obtained from u_{n-1} by the formula

$$u_n = u_{n-1} + h \sum_{j=1}^m b_j f(\tau_{n,j}, y_{n,j}),$$

where the vectors $y_{n,j} \in \mathbb{R}^s$ are computed from the relations

$$y_{n,i} = u_{n-1} + h \sum_{j=1}^m a_{ij} f(\tau_{n,j}, y_{n,j}) \quad (\text{for } i = 1, 2, \dots, m).$$

Here b_j, a_{ij} are real parameters specifying the method, with

$$b_1 + b_2 + \dots + b_m = 1.$$

The integer m is called the *number of stages* of the method. Further, for $1 \leq i \leq m$, we define

$$c_i = a_{i1} + a_{i2} + \dots + a_{im},$$

and we assume that

$$\tau_{n,i} = t_{n-1} + c_i h.$$

The so-called *classical Runge–Kutta method* is an instance of the above, with $m = 4$ and

$$\begin{aligned} a_{1,j} &= 0, \\ a_{2,1} &= 1/2, & a_{2,j} &= 0 \quad (j \neq 1), \\ a_{3,2} &= 1/2, & a_{3,j} &= 0 \quad (j \neq 2), \\ a_{4,3} &= 1, & a_{4,j} &= 0 \quad (j \neq 3), \\ b_1 &= b_4 = 1/6, & b_2 &= b_3 = 1/3. \end{aligned}$$

This famous choice for b_j, a_{ij} was made by W. Kutta [17], following ideas of C. Runge [19].

1.3. Error propagation

In the actual application of numerical methods one generally has to deal with *perturbations* occurring in the method, e.g., due to rounding-off. Therefore, suppose the general Runge–Kutta method is perturbed by some (small) quantities w_n , when the approximations u_n are computed from u_{n-1} .

It is an important question how such perturbations w_n are propagated in the subsequent applications of the Runge–Kutta method, and how they affect u_N . This question is particularly important in the case where $h > 0$ is small. In that case N is large, so that the perturbations occur in great numbers—and are propagated through many applications of the Runge–Kutta method.

We denote the approximation at $t = T$, obtained in the presence of all perturbations w_n , by \tilde{u}_N . The difference $\tilde{u}_N - u_N$ equals the accumulated effect of the perturbations, and will be called the *accumulated error*. We shall deal with estimating the norm $|\tilde{u}_N - u_N|$, in terms of the perturbations w_n , in particular for small values of $h > 0$. The purpose of the paper is to review and improve various estimates of $|\tilde{u}_N - u_N|$ which were obtained since an important work by Runge [20].

In analysing the relation between the accumulated error and the perturbations it is helpful to have concise formulations of the theoretical Runge–Kutta method and of its perturbed version. Assuming that the above system of equations for the vectors $y_{n,i}$ ($1 \leq i \leq m$) has a unique solution, the general Runge–Kutta process can be written as

$$u_n = u_{n-1} + h \cdot \Phi_n(u_{n-1}) \quad (n = 1, 2, \dots, N). \quad (1.2)$$

Here the function Φ_n depends on f , b_j , a_{ij} and h . Along with (1.2) we consider the perturbed process

$$\begin{aligned} \tilde{u}_n &= \tilde{u}_{n-1} + h \cdot [\Phi_n(\tilde{u}_{n-1}) + w_n] \quad (n = 1, 2, \dots, N), \\ \tilde{u}_0 &= u_0 + w_0, \end{aligned} \quad (1.3)$$

with perturbations w_0, w_1, \dots, w_N in \mathbb{R}^s .

As an illustration of (1.3) we consider the case where the vectors \tilde{u}_n denote the approximations to $U(t_n)$ actually obtained when the Runge–Kutta process is carried out on a computer (with finite-digit arithmetic). In this case w_0 denotes the rounding error introduced when the (true) initial vector u_0 is represented in the computer. Further, for $1 \leq n \leq N$ the occurrence of the vector w_n is caused by the round-off errors committed in the n th application of the Runge–Kutta method.

As a further illustration of (1.3) we consider the case where $\tilde{u}_n = U(t_n)$. Now $w_0 = 0$, and for $1 \leq n \leq N$ the vector w_n stands for the so-called *local discretization* error at t_n . The classical Runge–Kutta method has the well-known property that its local discretization errors satisfy $|w_n| = O(h^4)$, uniformly for $1 \leq n \leq N$ (assuming that $f \in C^4(\mathbb{R} \times \mathbb{R}^s)$). One usually refers to this property by saying that the classical method has an *order of accuracy* equal to 4.

Clearly, in both examples, upperbounds for $|\tilde{u}_N - u_N|$ in terms of w_n can reveal useful (apriori) information about the performance of the Runge–Kutta method.

1.4. Scope of the rest of the paper

In Section 2 we review various classical bounds for the accumulated error, obtained between 1905 and 1977. We compare these bounds with each other, and we identify one of them as being the most refined one with respect to its behaviour as $h \rightarrow 0$.

In Section 3 we note that in certain cases, where $\exp(LT)$ is much greater than 1, the classical bounds overestimate the actual accumulated error so strongly that the bounds become useless. We review various sharper error bounds, relevant to these cases, obtained since 1972.

In Section 4 we turn again to the situation where the classical bounds do *not* overestimate the actual error too much. We prove two new theorems which generalize and improve some of the material presented in Section 2.

2. Classical error bounds

2.1. Error estimates along the lines of Runge

Runge [20] essentially derived an estimate for $|\tilde{u}_N - u_N|$. He dealt with (what we now call) the classical Runge–Kutta method specified in Section 1.2, and was only concerned with the scalar case $s = 1$.

Runge's treatment of the classical Runge–Kutta method consists of two steps.

Step 1. First it is shown that

$$|\Phi_n(y) - \Phi_n(x)| \leq \left[1 + \frac{Lh}{2} + \frac{(Lh)^2}{6} + \frac{(Lh)^3}{24} \right] L|y - x|,$$

and from this inequality it is concluded that

$$|\Phi_n(y) - \Phi_n(x)| \leq \Lambda|y - x| \quad (\text{for } 0 < h \leq h_0).$$

One may choose, e.g.,

$$h_0 = 1/L \quad \text{and} \quad \Lambda = \left(1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24} \right) L = \frac{41}{24} L.$$

Step 2. By subtracting the relations (1.2) from the corresponding ones in (1.3) and by using the last upperbound for $|\Phi_n(y) - \Phi_n(x)|$ one obtains the inequality

$$|\tilde{u}_n - u_n| \leq (1 + h\Lambda)|\tilde{u}_{n-1} - u_{n-1}| + h|w_n|.$$

Applying this inequality for $n = N, N-1, \dots, 1$ and using $(1 + h\Lambda)^n \leq e^{nh\Lambda}$ one arrives at the error estimate

$$|\tilde{u}_N - u_N| \leq e^{\Lambda T} |w_0| + \frac{e^{\Lambda T} - 1}{\Lambda} \max_{1 \leq n \leq N} |w_n| \quad (0 < h \leq h_0). \quad (2.1)$$

Runge [20] used the last error estimate with w_n equal to the local discretization error (for $1 \leq n \leq N$) and $w_0 = 0$. In this fashion he obtained

$$|U(t_N) - u_N| = O(h^4).$$

Runge's derivation of (2.1) can be generalized to the case of arbitrary $s \geq 1$, $m \geq 1$, b_j , a_{ij} (see, e.g., [4,12] or Lemma 4.2 of the present paper). In that case the estimate (2.1) is still valid with $h_0 > 0$, $\Lambda = \sigma L$, $\sigma \geq 1$. Here h_0 depends only on L , a_{ij} ; and σ depends only on L , b_j , a_{ij} .

For the classical Runge–Kutta method, Bukovics [2] obtained better values for Λ , h_0 than Runge. For the case $s = 1$, Bukovics arrived at (2.1), with

$$\Lambda = L,$$

and with no restriction at all on $h > 0$.

We recall that the general Runge–Kutta method is said to have an *order of accuracy* p if the norm of the local discretization error is $O(h^p)$ (whenever $f \in C^p(\mathbb{R} \times \mathbb{R}^s)$). In [12] it was noted that, for $s \geq 1$, the improved value $\Lambda = L$ can be used in (2.1), for any $h > 0$, whenever the order of accuracy of the general Runge–Kutta method equals $m \geq 1$, and

$$\begin{aligned} b_j &\geq 0, & a_{ij} &\geq 0 & (\text{for } 1 \leq j \leq i \leq m), \\ a_{ij} &= 0 & & & (\text{for } 1 \leq i \leq j \leq m). \end{aligned}$$

2.2. More refined error estimates for $h \rightarrow 0$

Consider the estimate (2.1) for the case of arbitrary $s \geq 1$, $m \geq 1$, b_j, a_{ij} . Since

$$\frac{e^{\Lambda T} - 1}{\Lambda T} = 1 + \frac{\Lambda T}{2!} + \frac{(\Lambda T)^2}{3!} + \dots \leq e^{\Lambda T},$$

we obtain from (2.1)

$$|\tilde{u}_N - u_N| \leq \alpha \cdot [|w_0| + T \cdot \max_{1 \leq n \leq N} |w_n|] \quad (0 < h \leq h_0) \tag{2.2a}$$

with $\alpha = e^{\Lambda T}$.

In studying linear multistep methods, Dahlquist [8] and Hull and Luxemburg [14] dealt with error estimates that are of the form

$$|\tilde{u}_N - u_N| \leq \beta \cdot [|w_0| + h(|w_1| + |w_2| + \dots + |w_N|)] \quad (0 < h \leq h_0). \tag{2.2b}$$

Here β denotes a constant independent of h, w_n . The bound (2.2b) is also valid for the general Runge–Kutta method; it was used, e.g., by Stetter [27]. In fact, (2.2b) holds with $h_0 > 0$ depending only on L, a_{ij} ; and with $\beta = e^{\Lambda T}$, $\Lambda = \sigma L$ where $\sigma \geq 1$ depends only on L, b_j, a_{ij} as before.

The estimate (2.2b) is essentially *more refined* than (2.2a) in that (2.2b) implies (2.2a) (with $\alpha = \beta$) but, conversely, (2.2a) does *not* imply (2.2b) with fixed β (independent of h, w_n).

In order to illustrate that (2.2b) is more refined than (2.2a), we consider the special case where all $w_n = 0$ with the exception of $|w_1| = \varepsilon > 0$. Now (2.2a) only tells us that $|\tilde{u}_N - u_N| \leq \alpha T \varepsilon$. But, (2.2b) shows that $|\tilde{u}_N - u_N| \leq \beta \varepsilon \cdot h$, and accordingly that $|\tilde{u}_N - u_N| \rightarrow 0$ for $h \rightarrow 0$.

In Spijker [21] the bound (2.2b) was refined still further. It was proved that

$$|\tilde{u}_N - u_N| \leq \gamma \cdot \max_{0 \leq n \leq N} |w_0 + h(w_1 + w_2 + \dots + w_n)| \quad (0 < h \leq h_0), \tag{2.2c}$$

with a constant γ independent of h, w_n . In (2.2c), and in the subsequent, we use the convention that $(w_1 + w_2 + \dots + w_n) = 0$ if $n = 0$. We note that (2.2c) can be shown to hold with $\gamma = e^{\Lambda T}$, $\Lambda = \sigma L$, where $h_0 > 0$ and $\sigma \geq 1$ still only depend on the same parameters as above (for this result and for related material see also [11,12,22,23,27]).

Relation (2.2c) implies (2.2b) (with $\beta = \gamma$) but, conversely, (2.2b) does *not* imply (2.2c) with fixed γ (independent of h, w_n).

To illustrate that (2.2c) is more refined than (2.2b), we consider the case where $w_0 = 0, w_n = (-1)^n y$ ($1 \leq n \leq N$), $|y| = \varepsilon > 0$. Now (2.2b) only shows that $|\tilde{u}_N - u_N| \leq \beta T \varepsilon$. But, (2.2c) implies that $|\tilde{u}_N - u_N| \leq \gamma \varepsilon \cdot h$, and therefore that $|\tilde{u}_N - u_N| \rightarrow 0$ for $h \rightarrow 0$.

2.3. About the most refined estimate for $h \rightarrow 0$

The following question arises: Is (2.2c) maximally refined (in the sense of Section 2.2), or is it possible to refine this estimate still further?

In dealing with this question it is convenient to use the notation

$$v_n = w_0 + h(w_1 + w_2 + \cdots + w_n) \quad (\text{for } 0 \leq n \leq N).$$

From the material in [24] it immediately follows that

$$|\tilde{u}_N - u_N| \leq \delta \cdot \left[|v_N| + \frac{h}{T} (|v_0| + |v_1| + \cdots + |v_{N-1}|) \right] \quad (0 \leq h \leq h_0), \quad (2.2d)$$

with a constant δ independent of h , w_n . In Section 4.2 we shall give expressions for δ and h_0 .

The bound (2.2d) is still more refined than (2.2c). This is illustrated by the case where all $w_n = 0$ with the exception of w_2 and w_1 satisfying $w_2 = -w_1$, $|w_1| = \varepsilon h^{-1} > 0$. Now (2.2c) only implies that $|\tilde{u}_N - u_N| \leq \gamma\varepsilon$, whereas (2.2d) shows that $|\tilde{u}_N - u_N| \rightarrow 0$ for $h \rightarrow 0$.

We have now seen three consecutive improvements over (2.2a). One may hope that the third improvement did the trick, and that (2.2d) allows no further refinement.

From the material in [24] the following theorem follows. It deals with *Euler's method*, i.e., the case where $m = 1$, $a_{11} = 0$.

Theorem 2.1. *Suppose F is a function such that for Euler's method, and for any $f(t, x)$ as in Section 1.1, one has the error bound*

$$|\tilde{u}_N - u_N| \leq F(w_0, w_1, \dots, w_N; L, T, h) \quad (0 < h \leq h_0).$$

Then this function F necessarily satisfies, for $0 < h \leq h_0$,

$$F(w_0, w_1, \dots, w_N; L, T, h) \geq \lambda \cdot \left[|v_N| + \frac{h}{T} (|v_0| + |v_1| + \cdots + |v_{N-1}|) \right],$$

with $\lambda = LT/(1 + LT)$.

This theorem shows that, in the general situation with $L > 0$, the error bound (2.2d) for Euler's method cannot be refined (in the sense of the present section). For Euler's method, the bound (2.2d) may thus be considered to be *maximally refined*.

The estimate (2.2d) (for the general Runge–Kutta method) will be extended in Section 4 to the case of variable step sizes h_n . Moreover, in that section the bound (2.2d) will be proved to be maximally refined, not only for Euler's method but for any Runge–Kutta method.

3. Error bounds for special cases where $e^{LT} \gg 1$

3.1. The size of the error propagation factors

In Sections 2.2 and 2.3 we looked at the behaviour of the error estimates for $h \rightarrow 0$ and for fixed L and T . The size of the factors α , β , etc., was not taken into account in assessing the quality of the error bounds. In the present section we pay attention to the size of these factors.

For the ease of presentation, we consider throughout Section 3 only the case where

$$w_n = 0 \quad \text{for } n = 1, 2, \dots, N.$$

The estimate (2.2a) yields, for $0 < h \leq h_0$,

$$|\tilde{u}_N - u_N| \leq \alpha \cdot |\tilde{u}_0 - u_0|, \quad (3.1)$$

with $\alpha = e^{\Lambda T}$, $\Lambda = \sigma L \geq L$.

As an illustration we shall apply the bound (3.1) in the numerical solution of the test problem

$$U'(t) = -100 [U(t) - \frac{1}{3}] \quad (0 \leq t \leq 1), \quad U(0) = u_0.$$

The true solution is

$$U(t) = \frac{1}{3} [1 - \exp(-100t)] + u_0 \exp(-100t),$$

and the Lipschitz constant equals $L = 100$. Since $T = 1$, the factor α in (3.1) satisfies

$$\alpha \geq e^{LT} = e^{100} > 10^{43}.$$

We consider the numerical solution of our problem by the *implicit Euler method*, i.e., $m = 1$, $a_{11} = 1$. Suppose $u_0 = \frac{1}{3}$, and the (theoretical) numerical process is carried out instead with $\tilde{u}_0 = 0.333\ 333$. Since $|\tilde{u}_0 - u_0| = \frac{1}{3} \times 10^{-6}$, the estimate (3.1) reduces, for $0 < h \leq h_0$, to

$$|\tilde{u}_N - u_N| \leq E, \quad \text{with } E > 10^{36}.$$

But, it is easily verified that actually, for any $h > 0$,

$$|\tilde{u}_N - u_N| \leq |\tilde{u}_0 - u_0| = \frac{1}{3} \times 10^{-6}$$

In this example the magnitude of the factor α causes (3.1) to be a useless overestimation of the actual error at $T = 1$; and restricting h to some finite interval $(0, h_0]$ appears to be unnecessary.

An early improvement on the factor $\alpha \geq e^{LT}$ in (3.1) was obtained, for a large class of Runge–Kutta methods, by Galler and Rozenberg [10]. Following an idea of Carr [5], they established, under special assumptions about the function $f(t, x)$, an error propagation factor $\alpha < e^{LT}$. However, Galler and Rozenberg still restricted the step size to some interval $(0, h_0]$, where essentially $h_0 = O(1/L)$.

In the following Sections 3.2 and 3.3 we will shortly review more recent error bounds, with $\alpha < e^{LT}$ and without any restriction on $h > 0$. These bounds are very useful in some cases where $e^{LT} \gg 1$.

3.2. Contractivity

Consider the case where the function $f(t, x)$ and the norm $|x|$ on \mathbb{R}^s are such that

$$|W(T_1) - V(T_1)| \leq |W(T_0) - V(T_0)|,$$

whenever $T_1 > T_0$ and $V(t)$, $W(t)$ are any two solutions to our differential equation. We will refer to this situation by saying that the initial value problem (1.1) is *dissipative* (with respect to the given norm). The test problem of Section 3.1 is an example of a dissipative problem.

Under the assumption of dissipativity it is natural to ask for an estimate of the form

$$|\tilde{u}_N - u_N| \leq |\tilde{u}_0 - u_0|.$$

Definition 3.1. The Runge–Kutta method is *contractive* if the last inequality is valid for all $h > 0$, whenever $f(t, x)$ and the norm $|x|$ render problem (1.1) dissipative.

Clearly, contractivity is a useful demand on a Runge–Kutta method. It guarantees that a perturbation in the numerical process is propagated in a mild fashion—provided the initial value problem is dissipative.

Since the early seventies (cf. [9]) it has been known that the implicit Euler method has the favourable contractivity property. In a beautiful paper, Kraaijevanger [16] derived conditions on the coefficients b_j , a_{ij} of the general Runge–Kutta method that are necessary and sufficient in order that the method is (essentially) contractive in the above sense. We refer the reader to that paper for more details.

Unfortunately, all contractive Runge–Kutta methods are subject to a severe restriction on their order of accuracy. In [25] it was proved that the order p of any contractive method satisfies

$$p \leq 1.$$

Therefore, if one would insist on contractivity, many useful methods with an order $p > 1$ should be abandoned—which seems unrealistic. For that reason it makes sense to look for properties that are a bit weaker than contractivity.

One of these weaker properties studied in the literature is *B-stability*. This property still amounts to the inequality $|\tilde{u}_N - u_N| \leq |\tilde{u}_0 - u_0|$ (for all $h > 0$), but now only in those cases where (1.1) is dissipative with respect to a norm generated by an inner product in \mathbb{R}^s . Burrage and Butcher [3] and Crouzeix [6] succeeded in establishing very elegant conditions on b_j , a_{ij} which, in important cases, are necessary and sufficient for B-stability. For more details we refer to [4,13,15]. We note that there exist B-stable Runge–Kutta methods of any desired order $p \geq 1$.

Another weaker property dealt with in the literature is the relation

$$|\tilde{u}_N - u_N| \leq \beta \cdot |\tilde{u}_0 - u_0|, \quad (3.2)$$

(for all $h > 0$) with a *moderate* factor β . Here β is allowed to be greater than 1 (but of course essentially smaller than α as in (3.1)). In Section 3.3 we discuss estimates of the form (3.2), again for arbitrary norms in \mathbb{R}^s , not necessarily generated by an inner product.

3.3. Error propagation factors of moderate size

Establishing error bounds of the form (3.2) is highly nontrivial. Even for the linear problem

$$U'(t) = AU(t) \quad \text{for } 0 \leq t \leq T, \quad U(0) = u_0, \quad (3.3)$$

with constant $s \times s$ matrix A the analysis is far from easy. Below we shortly review a few results from the literature pertinent to (3.3).

A Runge–Kutta method is called *A-stable* if $|u_n| \leq |u_{n-1}|$ (for all $h > 0$), whenever the method is applied to any scalar test problem of the form $U'(t) = \lambda U(t)$, $U(0) = u_0$, with $u_0 \in \mathbb{C}$, $\lambda \in \mathbb{C}$, $\text{Re } \lambda \leq 0$. There exist A-stable Runge–Kutta methods of any desired order $p \geq 1$ (see, e.g., [4,13]).

In the rest of this section, problem (3.3) is assumed to be dissipative with respect to an arbitrary norm in \mathbb{R}^s , and the Runge–Kutta method is assumed to be A-stable. We consider two sequences of approximations $\{u_n\}$ and $\{\tilde{u}_n\}$ generated by applying the Runge–Kutta method to (3.3), with the same step size $h > 0$, starting from arbitrary initial vectors u_0 and \tilde{u}_0 , respectively, in \mathbb{R}^s .

From the material in the important paper [1] it follows that (3.2) holds, under the assumptions just stated, with

$$\beta = C_{\text{RK}} \cdot \sqrt{N} \quad (\text{for all } s \geq 1, N \geq 1, T > 0, h = T/N).$$

Here C_{RK} denotes a constant only depending on the coefficients b_j, a_{ij} of the Runge–Kutta method (and not on A or LT , with $L = \max\{|Ax|/|x|: x \neq 0\}$). For $e^{LT} \gg 1$ and realistic values of N , the estimate (3.2) with β as indicated is a great improvement over (3.1).

Another interesting improvement can be deduced from [18, Theorem 3.1]. Under the above assumptions, (3.2) also holds with

$$\beta = C_{\text{RK}} \cdot s \quad (\text{for all } s \geq 1, N \geq 1, T > 0, h = T/N),$$

where C_{RK} again denotes a constant only depending on b_j, a_{ij} .

We note that (3.3) can arise as outcome from semi-discretizing a given initial-boundary value problem for a linear partial differential equation (method of lines). In such a case the dimension s is related to the accuracy of the semi-discretization and can attain (arbitrarily) large values. For that reason, in [26], the question has been raised of whether the last formula for β can be improved into

$$\beta = C_{\text{RK}} \cdot s^q \quad (\text{for all } s \geq 1, N \geq 1, T > 0, h = T/N),$$

with some fixed $q < 1$. In the paper just mentioned it is conjectured that such an improvement is possible with $q = 1/2$. It is also conjectured that, under the quite general assumptions as stated above, an improvement with $q < 1/2$ is impossible. Up to now these questions are unresolved.

4. Error estimates which are maximally refined for $h \rightarrow 0$

4.1. Preliminaries

In the following we shall deal with error bounds which are maximally refined in the sense of the Sections 2.2 and 2.3. We assume that $|x|$ is an arbitrary norm in \mathbb{R}^s , and that $f(t, x)$, L are as in Section 1.1.

We consider a variable step size version of the general Runge–Kutta method. For $1 \leq n \leq N$ we specify step sizes h_n , gridpoints t_n and approximations u_n by

$$h_n > 0, \quad t_0 = 0, \quad t_n = t_{n-1} + h_n, \quad t_N = T, \quad (4.1)$$

and

$$u_n = u_{n-1} + h_n \sum_{j=1}^m b_j f(\tau_{n,j}, y_{n,j}), \quad (4.2a)$$

where the vectors $y_{n,j}$ satisfy

$$y_{n,i} = u_{n-1} + h_n \sum_{j=1}^m a_{ij} f(\tau_{n,j}, y_{n,j}) \quad (\text{for } 1 \leq i \leq m). \quad (4.2b)$$

We assume, for $1 \leq n \leq N$, $1 \leq i \leq m$, that

$$\tau_{n,i} = t_{n-1} + c_i h_n.$$

In Section 4.2 we shall give a generalization of (2.2d) for the case of the process (4.2), and provide an expression for the factor δ in (2.2d).

In Section 4.3 we present Theorem 4.7. This theorem consists in an improved version of Theorem 2.1, with the following two features: (1) It deals with the general process (4.2). (2) It reveals the optimality of bounds like (2.2d) already within the class of autonomous differential equations. In view of the last feature Theorem 4.7, when applied to Euler's method with constant step sizes, already amounts to a statement which is stronger than Theorem 2.1. The proof of Theorem 4.7 in Section 4.3 is essentially different from, and more simple than, the original proof of Theorem 2.1 along the lines of [24].

In the Sections 4.2 and 4.3 we shall use the subsequent Lemmas 4.1 and 4.2, which are based on ideas employed earlier by Crouzeix and Mignot [7, Section 5.5].

Lemma 4.1. *Let $p, q \in \mathbb{R}^m$, with components p_i, q_i , respectively. Let $Q = (q_{ij})$ be an $m \times m$ matrix with entries $q_{ij} \geq 0$, and let*

$$p_i \leq q_i + \sum_{j=1}^m q_{ij} p_j \quad (1 \leq i \leq m).$$

Assume the spectral radius $\rho(Q)$ of Q satisfies $\rho(Q) < 1$. Denote the entries of the matrix $R = (I - Q)^{-1}$ by r_{ij} . Then

$$p_i \leq \sum_{j=1}^m r_{ij} q_j \quad (1 \leq i \leq m).$$

Proof. Interpreting inequalities between vectors componentwise, we can write

$$p \leq q + Qp \leq q + Qq + \cdots + Q^k q + Q^{k+1} p, \quad \text{for any } k \geq 1.$$

By letting $k \rightarrow \infty$, we obtain $p \leq (I + Q + Q^2 + \cdots)q = Rq$, which proves the lemma. \square

We shall denote the $m \times m$ matrix, the entries of which are equal to the absolute values $|a_{ij}|$ of the coefficients in (4.2b), by $|A|$. The lemma below deals with a step size h_n satisfying

$$0 < h_n \leq h^*, \tag{4.3a}$$

where $h^* > 0$ is so small that

$$h^* L \cdot \rho(|A|) < 1. \tag{4.3b}$$

In the lemma we refer to the entries r_{ij} of the matrix

$$R = (r_{ij}) = (I - h^* L \cdot |A|)^{-1}. \tag{4.3c}$$

Lemma 4.2. *Assume (4.3). Then the conclusions (a), (b) are valid.*

(a) *For each $y \in \mathbb{R}^s$, the system of equations*

$$y_i = y + h_n \sum_{j=1}^m a_{ij} f(\tau_{n,j}, y_j) \quad (1 \leq i \leq m)$$

has a unique solution $y_i \in \mathbb{R}^s$ ($1 \leq i \leq m$).

(b) Denoting the above solution by $y_i = g_{n,i}(y)$, and defining

$$\Phi_n(y) = \sum_{j=1}^m b_j f(\tau_{n,j}, g_{n,j}(y)),$$

we have

$$|\Phi_n(y) - \Phi_n(x)| \leq A|y - x| \quad (\text{for all } x, y \in \mathbb{R}^s),$$

with

$$A = L \sum_{i,j=1}^m |b_i r_{ij}|. \tag{4.4}$$

Proof. (a) For any vector $X = (x_1, x_2, \dots, x_m)$, where $x_i \in \mathbb{R}^s$, we define the norm

$$|X| = \max \{|x_1|, |x_2|, \dots, |x_m|\},$$

and we write

$$G(X) = (G_1(X), G_2(X), \dots, G_m(X)), \quad \text{with } G_i(X) = y + h_n \sum_{j=1}^m a_{ij} f(\tau_{n,j}, x_j).$$

Let $X = (x_1, x_2, \dots, x_m)$, $Y = (y_1, y_2, \dots, y_m)$, $x_i \in \mathbb{R}^s$, $y_i \in \mathbb{R}^s$ be given. We put

$$(x_1^{(k)}, x_2^{(k)}, \dots, x_m^{(k)}) = G^k(X), \quad (y_1^{(k)}, y_2^{(k)}, \dots, y_m^{(k)}) = G^k(Y) \quad (\text{for } k = 0, 1, 2, \dots),$$

where G^k denotes the k th power of the application G , and G^0 is the identity. We define $p^{(k)}$ to be the vector in \mathbb{R}^m with components

$$p_i^{(k)} = |y_i^{(k)} - x_i^{(k)}| \quad (\text{for } 1 \leq i \leq m).$$

Putting

$$Q = h^* L \cdot |A|,$$

and interpreting inequalities between vectors in \mathbb{R}^m componentwise, we can write

$$p^{(k)} \leq Q p^{(k-1)} \leq \dots \leq Q^k p^{(0)}.$$

Since $\rho(Q) < 1$, it follows that there is a constant $\theta < 1$ and an integer $k \geq 1$ (both independent of X, Y) such that

$$p_i^{(k)} \leq \theta \cdot \max \{p_1^{(0)}, p_2^{(0)}, \dots, p_m^{(0)}\} \quad (\text{for } 1 \leq i \leq m).$$

Consequently, $|G^k(Y) - G^k(X)| = \max \{p_1^{(k)}, p_2^{(k)}, \dots, p_m^{(k)}\} \leq \theta \cdot |Y - X|$.

We have proved that G^k is a contraction in $(\mathbb{R}^s)^m$, so that there is a unique $Y = (y_1, y_2, \dots, y_m)$ with $G^k(Y) = Y$. We have

$$|G(Y) - Y| = |G^k(GY) - G^k(Y)| \leq \theta |G(Y) - Y|,$$

and therefore $G(Y) = Y$. Part (a) of the lemma has thus been proved.

(b) Assume, for $1 \leq i \leq m$,

$$x_i = x + h_n \sum_{j=1}^m a_{ij} f(\tau_{n,j}, x_j) \quad \text{and} \quad y_i = y + h_n \sum_{j=1}^m a_{ij} f(\tau_{n,j}, y_j).$$

Since

$$|y_i - x_i| \leq |y - x| + h^* L \sum_{j=1}^m |a_{ij}| |y_j - x_j| \quad (1 \leq i \leq m),$$

we obtain, using Lemma 4.1,

$$|y_i - x_i| \leq |y - x| \sum_{j=1}^m r_{ij},$$

with r_{ij} as in (4.3c). Hence,

$$|\Phi_n(y) - \Phi_n(x)| \leq L \sum_{i=1}^m |b_i| |y_i - x_i| \leq L |y - x| \sum_{i,j=1}^m |b_i r_{ij}|. \quad \square$$

As an illustration to Lemma 4.2, consider the classical Runge–Kutta method. For this method the spectral radius $\rho(|A|)$ is equal to zero, so that condition (4.3b) is satisfied for any $h^* > 0$. Applying Lemma 4.2, with $h_n = h = h^*$, we arrive at

$$|\Phi_n(y) - \Phi_n(x)| \leq \left[1 + \frac{hL}{2} + \frac{(hL)^2}{6} + \frac{(hL)^3}{24} \right] L |y - x|,$$

which proves Runge's upperbound for $|\Phi_n(y) - \Phi_n(x)|$ stated in Section 2.1 (at the beginning of Step 1).

4.2. Upper bounds for $|\tilde{u}_N - u_N|$

In this subsection we deal with the following variable step size versions of (1.2) and (1.3):

$$u_n = u_{n-1} + h_n \Phi_n(u_{n-1}) \quad (n = 1, 2, \dots, N), \quad (4.5)$$

$$\tilde{u}_n = \tilde{u}_{n-1} + h_n [\Phi_n(\tilde{u}_{n-1}) + w_n] \quad (n = 1, 2, \dots, N), \quad (4.6)$$

$$\tilde{u}_0 = u_0 + w_0.$$

We adjoin, to the above perturbations

$$w_n \in \mathbb{R}^s \quad (0 \leq n \leq N), \quad (4.7a)$$

the following vectors:

$$v_0 = w_0 \quad \text{and} \quad v_n = w_0 + h_1 w_1 + h_2 w_2 + \dots + h_n w_n \quad (1 \leq n \leq N). \quad (4.7b)$$

The theorem below applies to the situation where arbitrary constants A, A_0 exist such that, for all $x, y \in \mathbb{R}^s$ and $1 \leq n \leq N$,

$$|\Phi_n(y) - \Phi_n(x)| \leq \Lambda|y - x|, \tag{4.8a}$$

$$|\{y + h_n\Phi_n(y)\} - \{x + h_n\Phi_n(x)\}| \leq (1 + h_n\Lambda_0)|y - x|. \tag{4.8b}$$

Theorem 4.3. Assume (4.5)–(4.8). Then

$$|\tilde{u}_N - u_N| \leq |v_N| + \Lambda \sum_{n=1}^N \delta_n h_n |v_{n-1}|,$$

where

$$\delta_N = 1, \quad \delta_n = \prod_{i=n+1}^N (1 + h_i \Lambda_0) \quad (1 \leq n \leq N - 1).$$

Proof. We define

$$r_n = \tilde{u}_n - v_n \quad \text{and} \quad s_n = \Phi_n(\tilde{u}_{n-1}) - \Phi_n(r_{n-1}).$$

In view of (4.6) and (4.7b) we have

$$r_n = r_{n-1} + h_n [\Phi_n(r_{n-1}) + s_n].$$

By subtracting from this equality the relation

$$u_n = u_{n-1} + h_n \Phi_n(u_{n-1})$$

we obtain, by using (4.8b),

$$|r_n - u_n| \leq (1 + h_n \Lambda_0) |r_{n-1} - u_{n-1}| + h_n |s_n|.$$

Applying this inequality for $n = N, N - 1, \dots, 1$ there follows

$$|r_N - u_N| \leq \delta_N h_N |s_N| + \delta_{N-1} h_{N-1} |s_{N-1}| + \dots + \delta_1 h_1 |s_1|.$$

We have

$$|\tilde{u}_N - u_N| \leq |v_N| + |r_N - u_N|,$$

and in view of (4.8a)

$$|s_n| \leq \Lambda |v_{n-1}|.$$

Using the last three inequalities the desired result follows. \square

The following corollary to Theorem 4.3 shows, among other things, that (2.2d) holds with

$$h_0 = h^* \quad \text{and} \quad \delta = \max \{1, \Lambda T e^{\Lambda T}\};$$

here $h^* > 0$ is such that (4.3b) holds, and Λ is defined by (4.4) and (4.3c).

Corollary 4.4. Assume (4.3), and define $\Phi_n(y)$ as in Lemma 4.2. Suppose $u_n, \tilde{u}_n, w_n, v_n$ are as in (4.5)–(4.7). Then

$$|\tilde{u}_N - u_N| \leq |v_N| + \Lambda \sum_{n=1}^N \exp[\Lambda(T - t_n)] \cdot h_n |v_{n-1}|, \tag{4.9}$$

where Λ is defined by (4.4).

Proof. By Lemma 4.2 we have (4.8a) with Λ as in (4.4). Further, (4.8b) holds with $\Lambda_0 = \Lambda$. By applying Theorem 4.3 and noting that $\delta_n \leq \exp[\Lambda(T - t_n)]$ we obtain (4.9). \square

The following two corollaries to Theorem 4.3 specify conditions under which the exponential factors in (4.9) can be removed.

Corollary 4.5. *Suppose the numerical process (4.5) originates from the application of a contractive Runge–Kutta method to an initial value problem (1.1) which is dissipative (see Section 3.2). Then, under the sole assumption (4.8a), the relations (4.5)–(4.7) imply*

$$|\tilde{u}_N - u_N| \leq |v_N| + \Lambda \sum_{n=1}^N h_n |v_{n-1}|. \quad (4.10)$$

Proof. In view of the contractivity we have (4.8b) with $\Lambda_0 = 0$, so that Theorem 4.3 can be applied with $\delta_n = 1$. \square

Corollary 4.6. *Let the process (4.5) originate from the application of a B-stable Runge–Kutta method to a problem (1.1) which is dissipative (Section 3.2). Suppose the norm in \mathbb{R}^s is generated by an inner product. Then, assuming (4.8a), the relations (4.5)–(4.7) again imply (4.10).*

Proof. Similarly as above. \square

4.3. A lower bound for $|\tilde{u}_N - u_N|$

Along with (4.2) we consider, for $1 \leq n \leq N$, the perturbed process

$$\tilde{u}_n = \tilde{u}_{n-1} + h_n \sum_{j=1}^m b_j f(\tau_{n,j}, \tilde{y}_{n,j}) + h_n w_n, \quad (4.11a)$$

$$\tilde{y}_{n,i} = \tilde{u}_{n-1} + h_n \sum_{j=1}^m a_{ij} f(\tau_{n,j}, \tilde{y}_{n,j}) \quad (\text{for } 1 \leq i \leq m), \quad (4.11b)$$

with initial vector

$$\tilde{u}_0 = u_0 + w_0. \quad (4.11c)$$

In the theorem below we use the assumption

$$\gamma^* = \sum_{i,j=1}^m |b_i c_j r_{ij}|, \quad h^* \gamma^* L < 1. \quad (4.12)$$

Theorem 4.7. *Let a norm $|x|$ in \mathbb{R}^s and a constant $L > 0$ be given. Let $u_0 \in \mathbb{R}^s$, and assume (4.1) and (4.7). Then there exists a function $f_0: \mathbb{R}^s \rightarrow \mathbb{R}^s$ such that the following statements (a), (b) are valid.*

(a) *The function $f(t, x) \equiv f_0(x)$ has a Lipschitz constant equal to L ;*

(b) Whenever the relations (4.2) and (4.11) are satisfied with $f(t, x) \equiv f_0(x)$ and with h_n restricted according to (4.3) and (4.12) ($1 \leq n \leq N$), then

$$|\tilde{u}_N - u_N| \geq |v_N| + \frac{(1 - h^* \gamma^* L)L}{1 + (1 + h^* \gamma^* L)LT} \sum_{n=1}^N h_n |v_{n-1}|. \tag{4.13}$$

Proof. (1) We define the vector $v \in \mathbb{R}^s$ by $v = |v_N|^{-1}v_N$ if $v_N \neq 0$, and we choose v arbitrary with $|v| = 1$ if $v_N = 0$. We define, for $x \in \mathbb{R}^s, t \in \mathbb{R}$,

$$f_0(x) = L|x - u_0|v \quad \text{and} \quad f(t, x) = f_0(x).$$

Using $|f(t, y) - f(t, x)| = L||y - u_0| - |x - u_0|| \leq L|y - x|$, we obtain statement (a).

(2) Assume (4.3) and (4.12), and let u_n, \tilde{u}_n be any vectors satisfying (4.2) and (4.11), respectively.

By virtue of Lemma 4.2(a), we conclude that, for $1 \leq n \leq N, 1 \leq i \leq m$,

$$u_n = y_{n,i} = u_0 \quad \text{with} \quad f(\tau_{n,i}, y_{n,i}) = 0.$$

Defining $d_n = \tilde{u}_n - u_n, e_{n,i} = \tilde{y}_{n,i} - y_{n,i}$, we see from (4.11) that $d_0 = v_0$ and, for $1 \leq n \leq N$,

$$d_n = d_{n-1} + h_n L \sum_{j=1}^m b_j |e_{n,j}|v + h_n w_n,$$

$$e_{n,i} = d_{n-1} + h_n L \sum_{j=1}^m a_{ij} |e_{n,j}|v \quad (\text{for } 1 \leq i \leq m).$$

Consequently,

$$e_{n,i} - d_{n-1} = h_n c_i L |d_{n-1}|v + h_n L \sum_{j=1}^m a_{ij} (|e_{n,j}| - |d_{n-1}|)v,$$

and therefore

$$|e_{n,i} - d_{n-1}| \leq h_n |c_i| L |d_{n-1}| + h^* L \sum_{j=1}^m |a_{ij}| \cdot |e_{n,j} - d_{n-1}|.$$

Applying Lemma 4.1, with $Q = h^* L|A|$, we conclude that, for $1 \leq n \leq N, 1 \leq i \leq m$,

$$|e_{n,i} - d_{n-1}| \leq h_n L \left(\sum_{j=1}^m |r_{ij} c_j| \right) |d_{n-1}|.$$

(3) The vectors $z_n = d_n - v_n$ satisfy

$$z_n = z_{n-1} + h_n L \sum_{i=1}^m b_i |e_{n,i}|v = z_{n-1} + h_n L \sum_{i=1}^m b_i (|d_{n-1}|v + s_{n,i}),$$

where $s_{n,i} = (|e_{n,i}| - |d_{n-1}|)v$. Using the conclusion at the end of part (2) of the proof we have

$$|s_{n,i}| \leq h_n L \left(\sum_{j=1}^m |r_{ij} c_j| \right) |d_{n-1}|.$$

Consequently,

$$z_n = z_{n-1} + h_n L |d_{n-1}| (v + x_n) \quad \text{with } |x_n| \leq h_n L \sum_{i,j=1}^m |b_i c_j r_{ij}|.$$

Hence, for $1 \leq n \leq N$,

$$z_n = L \sum_{j=1}^n h_j |d_{j-1}| (v + x_j), \quad \text{with } |x_j| \leq h^* \gamma^* L.$$

(4) From the last expression for z_n we obtain, since $d_N = |v_N|v + z_N$, the lower bound

$$|d_N| \geq |v_N| + (1 - h^* \gamma^* L) L \sum_{j=1}^N h_j |d_{j-1}|.$$

Using that $v_n = d_n - z_n$, we also obtain, from the above expression for z_n ,

$$|v_n| \leq |d_n| + (1 + h^* \gamma^* L) L \sum_{j=1}^n h_j |d_{j-1}|.$$

We thus have the inequality

$$\sum_{n=1}^N h_n |v_{n-1}| \leq \sum_{n=1}^N h_n |d_{n-1}| + (1 + h^* \gamma^* L) L \sum_{n=1}^N \sum_{j=1}^{n-1} h_n h_j |d_{j-1}|,$$

which implies

$$\sum_{n=1}^N h_n |v_{n-1}| \leq [1 + (1 + h^* \gamma^* L) LT] \sum_{j=1}^N h_j |d_{j-1}|.$$

By using the last inequality, the above lower bound for $|d_N|$, and the assumption on h^* made in (4.12), we arrive at (4.13). \square

Consider the case where the Runge–Kutta method stands for Euler's method ($m = 1$, $a_{11} = 0$), and where $h_n \equiv h > 0$. Then $\gamma^* = 0$, so that (4.13) reduces to

$$|\tilde{u}_N - u_N| \geq |v_N| + (1 + LT)^{-1} Lh \sum_{n=1}^N |v_{n-1}|.$$

This implies

$$|\tilde{u}_N - u_N| \geq \lambda \cdot [|v_N| + \frac{h}{T} (|v_0| + |v_1| + \dots + |v_{N-1}|)],$$

with $\lambda = LT/(1 + LT)$. The last inequality yields the lower bound for $F(w_0, w_1, \dots, w_N; L, T, h)$ stated in Theorem 2.1, and thus proves that theorem.

In the general case one can use (4.13), similarly as above for Euler's method, to prove that (4.9) is maximally refined (in the sense used earlier).

References

- [1] P. Brenner and V. Thomeé, On rational approximations of semigroups, *SIAM J. Numer. Anal.* 16 (1979) 683–694.
- [2] E. Bukovics, Beiträge zur numerische Integration II, *Monatsh. Math.* 57 (1953) 333–350.
- [3] K. Burrage and J.C. Butcher, Stability criteria for implicit Runge–Kutta methods, *SIAM J. Numer. Anal.* 16 (1979) 46–57.
- [4] J.C. Butcher, *The Numerical Analysis of Ordinary Differential Equations* (Wiley, Chichester, 1987).
- [5] J.W. Carr III, Error bounds for the Runge–Kutta single-step integration process, *J. Assoc. Comput. Mach.* 5 (1958) 39–44.
- [6] M. Crouzeix, Sur la B -stabilité des méthodes de Runge–Kutta, *Numer. Math.* 32 (1979) 75–82.
- [7] M. Crouzeix and A.L. Mignot, *Analyse Numérique des Équations Différentielles* (Masson, Paris, 1984).
- [8] G. Dahlquist, Convergence and stability in the numerical integration of ordinary differential equations, *Math. Scand.* 4 (1956) 33–53.
- [9] C.A. Desoer and H. Haneda, The measure of a matrix as a tool to analyse computer algorithms for circuit analysis, *IEEE Trans. Circuit Theory* 19 (1972) 480–486.
- [10] B.A. Galler and D.P. Rozenberg, A generalization of a theorem of Carr on error bounds for Runge–Kutta procedures, *J. Assoc. Comput. Mach.* 7 (1960) 57–60.
- [11] R.D. Grigorieff, *Numerik gewöhnlicher Differentialgleichungen*, Vol. 2 (Teubner, Stuttgart, 1977).
- [12] E. Hairer, S.P. Nørsett and G. Wanner, *Solving Ordinary Differential Equations I* (Springer, Berlin, 1993).
- [13] E. Hairer and G. Wanner, *Solving Ordinary Differential Equations II* (Springer, Berlin, 1991).
- [14] T.E. Hull and W.A.J. Luxemburg, Numerical methods and existence theorems for ordinary differential equations, *Numer. Math.* 2 (1960) 30–41.
- [15] W.H. Hundsdorfer and M.N. Spijker, A note on B -stability of Runge–Kutta methods, *Numer. Math.* 36 (1981) 319–331.
- [16] J.F.B.M. Kraaijevanger, Contractivity of Runge–Kutta methods, *BIT* 31 (1991) 482–528.
- [17] W. Kutta, Beitrag zur näherungsweise Integration totaler Differentialgleichungen, *Z. Math. Phys.* 46 (1901) 435–453.
- [18] C. Lubich and O. Nevanlinna, On resolvent conditions and stability estimates, *BIT* 31 (1991) 293–313.
- [19] C. Runge, Ueber die numerische Auflösung von Differentialgleichungen, *Math. Ann.* 46 (1895) 167–178.
- [20] C. Runge, Ueber die numerische Auflösung von totaler Differentialgleichungen, *Göttinger Nachr.* (1905) 252–257.
- [21] M.N. Spijker, Stability and convergence of finite-difference methods, Thesis, Leiden University (1968).
- [22] M.N. Spijker, On the structure of error estimates for finite-difference methods, *Numer. Math.* 18 (1971) 73–100.
- [23] M.N. Spijker, On the possibility of two-sided error bounds in the numerical solution of initial value problems, *Numer. Math.* 26 (1976) 271–300.
- [24] M.N. Spijker, The behaviour of error bounds in the numerical solution of initial value problems when the step size h tends to zero, in: J.J.H. Miller, ed., *Topics in Numerical Analysis III* (Academic Press, London, 1977) 383–400.
- [25] M.N. Spijker, Contractivity in the numerical solution of initial value problems, *Numer. Math.* 42 (1983) 271–290.
- [26] M.N. Spijker, Stiffness in numerical initial value problems, *J. Comput. Appl. Math.* 72 (1996) 393–406.
- [27] H.J. Stetter, *Analysis of Discretization Methods for Ordinary Differential Equations* (Springer, Berlin, 1973).