Numer. Math. 26, 271–300 (1976)

On the Possibility of Two-Sided Error Bounds in the Numerical Solution of Initial Value Problems

M. N. Spijker

Received February 6, 1975

Summary. Let u denote the approximation produced by a finite-difference method for solving an initial value problem for a given differential equation. Suppose the finite-difference equation is perturbed by a quantity w, e.g. due to round-off or truncation errors. Then, instead of u, one obtains a solution which we denote by $\tilde{u}$.

In this paper a condition is presented which is necessary and sufficient for the existence of a two-sided estimate of the error $\tilde{u}-u$ in terms of the perturbation w. The paper is concluded with applications in the fields of ordinary and partial parabolic differential equations.

1. Introduction

In order to introduce the main problem that is treated in this paper we consider the familiar example of an initial value problem

$$(1.1) \qquad \frac{d}{dt} U(t) = f(t, U(t)) \quad (0 \leqq t \leqq T), \qquad U(0) = c$$

which is solved numerically by Euler's method

$$(1.2) \qquad h^{-1}(u_n - u_{n-1}) = f(t_{n-1}, u_{n-1}) \quad (n=1, 2, \ldots, N), \qquad u_0 = c.$$

In (1.2) u_n denotes an approximation of $U(t)$ at $t=t_n=nh$ and N is the greatest integer with $N h \leqq T$. Along with (1.2) we consider a perturbed version of Euler's method

$$(1.3) \qquad h^{-1}(\tilde{u}_n - \tilde{u}_{n-1}) = f(t_{n-1}, \tilde{u}_{n-1}) + w_n \quad (n=1, 2, \ldots, N), \qquad \tilde{u}_0 = c + w_0$$

where $\tilde{u}_n$ denotes the approximation of $U(t_n)$ obtained in the presence of some perturbations $w_0, w_1, \ldots, w_N$. For instance w_n may represent a truncation error arising in the computation of $f(t_{n-1}, \tilde{u}_{n-1})$. Likewise, the local perturbations w_n may be caused by rounding-off. Finally w_n may also be understood to be the local discretization error (we use here the terminology of [14]) of Euler's method, in which case we have $w_n = \frac{h}{2} \cdot \frac{d^2}{dt^2} U(\tau_n)$ (with $t_{n-1} < \tau_n < t_n$, and $n=1, 2, \ldots, N$) and $\tilde{u}_n = U(t_n)$. Clearly, in each of these cases it is desirable to have an error bound by means of which the effect of the perturbations w_n on the differences $\tilde{u}_n - u_n$ can be estimated.

Suppose we have an error bound which can be written in the form

$$(1.4) \qquad \max_{0 \leq n \leq N} |\tilde{u}_n - u_n| \leqq E_1$$

where $E_1 = E_1[w_0, w_1, \ldots, w_N; h]$ depends on the perturbations w_n and on $h > 0$. Then (1.4) is of particular interest if the factor by which the actual quantity $\max_{0 \leq n \leq N} |\tilde{u}_n - u_n|$ may be overestimated, is bounded uniformly for all w_n and all $h > 0$. Thus a basic requirement to be imposed on the error estimate (1.4) could be that its right-hand member E_1 divided by $\max_{0 \leq n \leq N} |\tilde{u}_n - u_n|$ is bounded by some fixed constant $\beta > 0$. Clearly, this requirement is fulfilled if an only if there exists a *two-sided error bound*

$$E_0 \leqq \max_{0 \leqq n \leqq N} |\tilde{u}_n - u_n| \leqq E_1 \tag{1.5}$$

with a left-hand member E_0 that can be written in the form $E_0 = \frac{1}{\beta} \cdot E_1$ where $\beta > 0$ is independent of $w_0, w_1, \ldots, w_N$ and h.

We assume that the partial derivative with respect to x of the function f appearing in the initial value problem (1.1), satisfies

$$\left| \frac{\partial}{\partial x} f(t, x) \right| \leqq L < \infty \tag{1.6}$$

uniformly for $0 \leqq t \leqq T$, $-\infty < x < \infty$. Using (1.6) it can be proved (cf. [11, 14, 15]) that the error $\tilde{u}_n - u_n$ caused by the local perturbations $w_0, w_1, \ldots, w_n$ in (1.3) admits the following two-sided error bound of type (1.5):

$$\gamma_0 \cdot \max_{0 \leqq n \leqq N} \left| w_0 + h \sum_{j=1}^{n} w_j \right| \leqq \max_{0 \leqq n \leqq N} |\tilde{u}_n - u_n| \leqq \gamma_1 \cdot \max_{0 \leqq n \leqq N} \left| w_0 + h \sum_{j=1}^{n} w_j \right| \tag{1.7}$$

(we use the convention $\sum_{j=a}^{b} \ldots = 0$ for $a > b$). For γ_0, γ_1 one may obtain the expressions $\gamma_0 = (1 + LT)^{-1}$, $\gamma_1 = \exp(LT)$. Further, if assumptions more refined than (1.6) are made, (1.7) can be shown to hold with a constant γ_1 smaller than the exponential factor $\exp(LT)$. For instance, if, in addition to (1.6), it is assumed that

$$\frac{\partial}{\partial x} f(t, x) \leqq M < 0$$

(uniformly for $0 \leqq t \leqq T$, $-\infty < x < \infty$), then, for h sufficiently small, we may put $\gamma_1 = \min\left[1 - \frac{L}{M}, \exp(LT)\right]$. We note that the two-sided error estimate (1.7) has the important property that, except for factors γ_0, γ_1 which are independent of $w_0, w_1, \ldots, w_N$ and h, the lower bound and upper bound for $\max_{0 \leqq n \leqq N} |\tilde{u}_n - u_n|$ *do not depend on the function f.*

The question arises whether results similar to the above still hold if the maximum-norm $\max_{0 \leqq n \leqq N} |\tilde{u}_n - u_n|$ is replaced throughout by some other seminorm, e.g. by the Euclidean norm

$$\left\{ h \sum_{n=0}^{N} (\tilde{u}_n - u_n)^2 \right\}^{\frac{1}{2}} \tag{1.8}$$

or by the important seminorm

$$|\tilde{u}_N - u_N|. \tag{1.9}$$

In this paper it will be shown that the answer to this question can be positive as well as negative depending on the seminorm chosen. Using the theory of this article e.g. the conclusion can be drawn that the two-sided error bound

$$\begin{aligned}\gamma_0 \cdot \left\{h \sum_{n=0}^{N} \left(w_0 + h \sum_{j=1}^{n} w_j\right)^2\right\}^{\frac{1}{2}} &\leq \left\{h \cdot \sum_{n=0}^{N} (\tilde{u}_n - u_n)^2\right\}^{\frac{1}{2}} \\ &\leq \gamma_1 \cdot \left\{h \sum_{n=0}^{N} \left(w_0 + h \sum_{j=1}^{n} w_j\right)^2\right\}^{\frac{1}{2}}\end{aligned} \tag{1.10}$$

holds, with the same γ_0, γ_1 as in (1.7). The lower bound and upper bound in (1.10) are again independent of the function f, except for factors independent of $w_0, w_1, \ldots, w_N$ and h. The results in this paper also imply that, on the other hand, there exists no two-sided error bound for $|\tilde{u}_N - u_N|$ of the form

$$\gamma_0 \cdot \phi[w_0, w_1, \ldots, w_N; h] \leq |\tilde{u}_N - u_N| \leq \gamma_1 \cdot \phi[w_0, w_1, \ldots, w_N; h] \tag{1.11}$$

with constants γ_0, γ_1 independent of $w_0, w_1, \ldots, w_N$ and h, and with a function ϕ independent of f.

We note that (1.11) is trivially fulfilled with $\gamma_0 = \gamma_1 = 1$ and ϕ defined by

$$\phi[w_0, w_1, \ldots, w_N; h] = |\tilde{u}_N - u_N|$$

where $\tilde{u}_N$ is computed from (1.3) and $|\tilde{u}_N - u_N|$ can thus be regarded as a function of $w_0, w_1, \ldots, w_N$ and h. Clearly, with this choice of γ_0, γ_1 and ϕ the error estimate (1.11) is useless, since the values $\phi[w_0, w_1, \ldots, w_N; h]$ depend on $w_0, w_1, \ldots, w_N$ in a manner which in general is untransparent—due to the possibly complicated structure of the (nonlinear) function f. It is in view of the existence of such trivial and simultaneously untransparent error estimates that we focus on two-sided estimates in which, apart from factors independent of w_n and h, the lower bound and upper bound are independent of f.

In Chapter 2 we introduce the class of finite-difference methods we shall deal with in the rest of this paper. It consists of step-by-step methods for solving initial value problems for semilinear (ordinary or partial) differential equations. Chapter 2 is concluded with Theorem 2 containing the main result of this article. This theorem gives a very simple condition on the seminorm which is necessary and sufficient for the existence of nontrivial two-sided error bounds.

In Chapter 3 we prove Theorem 2 using a series of lemmata proved at the beginning of Chapter 3. At the end of Chapter 3 (Section 3.5) we touch on some further applications of these lemmata, which have not been incorporated in Theorem 2 of Chapter 2. We prove the above statement about the non-existence of error bounds of type (1.11). Further, we indicate how the theorems which were stated without proof in [13] follow from the theory of Chapter 3.

Chapter 4 contains illustrations to the material of the Chapters 2, 3 in the fields of ordinary and partial differential equations. The error bounds (1.7), (1.10) as well as the expressions for γ_0, γ_1 given above easily follow from the examples treated in Section 4.2.

For further examples and applications of two-sided error bounds we refer to the publications [11—16].

2. A Theorem on the Existence of Two-Sided Error Bounds

2.1. The Finite-Difference Scheme

Let T and h_0 be given numbers with $0 < h_0 \leqq T$ and let H denote some subset of the interval $(0, h_0]$ with $\inf H = 0$. A number h belonging to H is called a *stepsize*. We shall deal with finite-difference methods producing approximations u_n to some solution $U(t)$ of a given differential equation at the *gridpoints* $t = t_n$ for $n = 0, 1, \dots, N$. The points t_n are defined by $t_n = n\, h$ (for $n = 0, 1, 2 \dots$) where h is a given stepsize, and N is the largest integer with $t_N \leqq T$. The approximations u_n, which depend on the stepsize h that is chosen, are assumed to belong to a real Banach space $\mathfrak{R}_h$, not necessarily of finite dimension. The norm in $\mathfrak{R}_h$ is denoted by $|a|_h$ or simply by $|a|$ for $a \in \mathfrak{R}_h$. It is assumed that $\mathfrak{R}_h$ contains a vector $e \neq 0$.

We assume that the approximations u_n are obtained by solving the finite-difference equation

$$(2.1.\text{a}) \qquad h^{-1} \sum_{i=0}^{k} A_{n,i}(u_{n-i}) = F_n(u_0, u_1, \dots, u_n; h) \quad (n = k, k+1, \dots, N)$$

using starting values $c_n \in \mathfrak{R}_h$:

$$(2.1.\text{b}) \qquad u_n = c_n \quad (n = 0, 1, \dots, k-1).$$

In (2.1) k denotes a fixed integer $\geqq 1$ and it is assumed that h_0 is so small that $k\, h_0 \leqq T$. Hence $N \geqq k$ for all $h \in H$. We assume that the operators $A_{n,i}$, F_n satisfy the following conditions.

Condition I. For each $h \in H$ and $i = 0, 1, \dots, k$, $n = k, k+1, \dots, N$ the $A_{n,i}$ are linear operators from $\mathfrak{R}_h$ into itself and the operators $A_{n,0}$ are invertible. The operators $A_{n,i}$ are allowed to depend on h but in order to avoid cumbersome notation we suppress a third index of $A_{n,i}$ indicating this dependence.

Condition II. For $h \in H$, $k \leqq n \leqq N$, $x_j \in \mathfrak{R}_h$ $(0 \leqq j \leqq n)$ the element $y = F_n(x_0, x_1, \dots, x_n; h)$ belongs to $\mathfrak{R}_h$. Further if x_j and $\tilde{x}_j$ are arbitrary vectors in $\mathfrak{R}_h$ then

$$|F_n(\tilde{x}_0, \dots, \tilde{x}_n; h) - F_n(x_0, \dots, x_n; h)| \leqq \sum_{i=0}^{n} \lambda_i |\tilde{x}_{n-i} - x_{n-i}|$$

where $\lambda_0, \lambda_1, \lambda_2, \dots$ are constants independent of h, n, x_j, $\tilde{x}_j$. It is assumed that the constants λ_i vanish for all $i < q$ and for all $i > r$. With q, r we denote arbitrary fixed integers with $0 \leqq q \leqq k$, $q \leqq r$.

Condition III. There is a constant $\alpha > 0$ such that whenever $h \in H$ and $x_k, x_{k+1}, \dots, x_N, y_0, y_1, \dots, y_N \in \mathfrak{R}_h$ satisfy

$$h^{-1} \sum_{i=0}^{k} A_{n,i}(y_{n-i}) = x_n \quad (k \leqq n \leqq N), \qquad y_n = 0 \quad (0 \leqq n \leqq k-1)$$

then

$$|y_n| \leqq \alpha \cdot h \sum_{j=k}^{n} |x_j| \quad (k \leqq n \leqq N).$$

Condition IV. For any $h\in H$, $k\leqq n\leqq N$ and any vectors $y, x_0, x_1, \ldots, x_{n-1}\in\mathfrak{R}_h$ the equation

$$h^{-1} A_{n,0}(x) - F_n(x_0, x_1, \ldots, x_{n-1}, x; h) = y$$

has a unique solution $x\in\mathfrak{R}_h$.

It is clear that Euler's method (1.2) is an example of the general finite-difference scheme (2.1) with $H=(0, T]$, $\mathfrak{R}_h=$ the set of real numbers $\mathbb{R}$, $A_{n,0}(x)\equiv x$, $A_{n,1}(x)\equiv -x$, $F_n(x_0, x_1, \ldots, x_n; h)\equiv f((n-1)h, x_{n-1})$, $k=1$, $\alpha=1$, $q=r=1$ and $\lambda_1=L$-assuming that $\frac{\partial}{\partial x} f$ satisfies the inequality (1.6). For further examples of (2.1) we refer to the Sections 4.2, 4.3.

From the conditions I, II it is clear that the finite-difference equation (2.1.a) consists of a linear part and a (nonlinear) part which is Lipschitz-continuous uniformly in the stepsize h. Therefore (2.1.a) is a *semilinear finite-difference equation*. We note that the right-hand member of the Lipschitz condition in condition II consists in a sum of at most $(r-q+1)$ terms.

Condition III requires that the linear part of the finite-difference equation (2.1.a) has a property which by many authors has been called *stability*.

Condition IV implies that for given $u_0, u_1, \ldots, u_{n-1}$ there is a unique u_n satisfying (2.1.a). Consequently there are unique vectors $u_0, u_1, \ldots, u_N$ satisfying (2.1).

2.2. The Perturbed Finite-Difference Scheme

Assume the finite-difference scheme is perturbed by quantities w_n, e.g. due to errors occurring in the actual calculation of F_n and c_n on a computer. Then instead of the u_n satisfying (2.1) we obtain vectors $\tilde{u}_n$ satisfying

$$\text{(2.2.a)}\qquad h^{-1}\sum_{i=0}^{k} A_{n,i}(\tilde{u}_{n-i}) = F_n(\tilde{u}_0, \tilde{u}_1, \ldots, \tilde{u}_n; h) + w_n \quad (n=k, k+1, \ldots, N),$$

$$\text{(2.2.b)}\qquad \tilde{u}_n = c_n + w_n \quad (n=0, 1, \ldots, k-1).$$

We note that condition IV ensures that for any $w_0, w_1, \ldots, w_N\in\mathfrak{R}_h$ there are unique $\tilde{u}_0, \tilde{u}_1, \ldots, \tilde{u}_N\in\mathfrak{R}_h$ satisfying (2.2).

Most methods of type (2.1) that are of practical value satisfy, in addition to the conditions I–IV, the following condition on the size of $|\tilde{u}_n - u_n|$.

Condition V. There is a constant $\gamma>0$ such that whenever $h\in H$ and u_n, $\tilde{u}_n$, $w_n\in\mathfrak{R}_h$ satisfy (2.1), (2.2) and $w_0=w_1=\cdots=w_{k-1}=0$ then

$$|\tilde{u}_n - u_n| \leqq \gamma\cdot h\sum_{j=k}^{n}|w_j| \quad (k\leqq n\leqq N).$$

This condition has a structure which is similar to that of the stability condition III. In fact, if F_n would vanish, condition V would reduce to condition III.

In the following theorem we formulate a simple condition under which condition V (and condition IV) is fulfilled. Although the proof of the theorem does not deviate substantially from proofs of related theorems to be found in the literature (see e.g. [14, p. 85]) we have included it, mainly because the proof reveals an inequality that is used in the subsequent.

Theorem 1. Assume the finite-difference scheme (2.1) satisfies the conditions I, II, III. Let h_0 be so small that

$$\alpha\,\lambda_0\,h_0 < 1. \tag{2.3}$$

Then the conditions IV, V are satisfied as well.

Proof. 1. In order to prove that condition IV is satisfied we rewrite the equation

$$h^{-1} A_{n,0}(x) = F_n(x_0, x_1, \ldots, x_{n-1}, x; h) + y \tag{2.4}$$

in the form $x=G(x)$ where $G(x)$ denotes the element into which the right-hand member of (2.4) is transformed by the operator $h(A_{n,0})^{-1}$. From condition III it follows that the norm of the operator $(A_{n,0})^{-1}$ induced by the norm in $\mathfrak{R}_h$, satisfies the inequality $|(A_{n,0})^{-1}| \leqq \alpha$. Hence by using the Lipschitz condition appearing in condition II we obtain $|G(\tilde{x})-G(x)| \leqq h\,\alpha\lambda_0 \cdot |\tilde{x}-x| \leqq \alpha\lambda_0 h_0 \cdot |\tilde{x}-x|$ for arbitrary x, $\tilde{x} \in \mathfrak{R}_h$. In view of (2.3) G is a contraction from $\mathfrak{R}_h$ into itself and the equation $x=Gx$ has a unique solution in $\mathfrak{R}_h$ (cf. [7]). It follows that condition IV is fulfilled.

2. In order to prove that condition V is satisfied we subtract the relations appearing in (2.1) from the corresponding ones in (2.2). Writing $d_n = \tilde{u}_n - u_n$ and using that $w_0 = w_1 = \cdots = w_{k-1} = 0$ we thus have

$$h^{-1}\sum_{i=0}^{k} A_{n,i}(d_{n-i}) = [F_n(\tilde{u}_0, \ldots, \tilde{u}_n; h) - F_n(u_0, \ldots, u_n; h)] + w_n \quad (k \leqq n \leqq N),$$
$$d_n = 0 \quad (0 \leqq n \leqq k-1).$$

By virtue of the conditions III, II we obtain

$$|d_n| \leqq \alpha\,h \cdot \left\{\lambda_0 |d_n| + \lambda \sum_{j=k}^{n-1} |d_j| + \sum_{j=k}^{n} |w_j|\right\} \quad (k \leqq n \leqq N)$$

where

$$\lambda = \sum_{i=q}^{r} \lambda_i. \tag{2.5}$$

Since $h \leqq h_0$ and h_0 is so small that (2.3) holds, we have

$$|d_n| \leqq \beta\,\lambda \cdot h \sum_{j=k}^{n-1} |d_j| + \beta \cdot h \sum_{j=k}^{n} |w_j| \quad (k \leqq n \leqq N)$$

with

$$\beta = \alpha(1 - \alpha\,\lambda_0\,h_0)^{-1}. \tag{2.6}$$

From this inequality there follows, by induction with respect to n and by noting that $|\tilde{u}_n - u_n| = |d_n|$,

$$|\tilde{u}_n - u_n| \leqq \beta \cdot h \sum_{j=k}^{n} (1 + \beta\,\lambda \cdot h)^{n-j} |w_j| \quad (k \leqq n \leqq N). \tag{2.7}$$

It follows that condition V is fulfilled with

$$\gamma = \beta \cdot \exp(\beta\,\lambda\,T). \tag{2.8}$$

We note that the assumption (2.3) of the theorem can always be fulfilled by restricting our considerations to stepsizes h which are small enough. Moreover, if $q \geqq 1$, we have $\lambda_0 = 0$ and condition (2.3) is trivially fulfilled, independently of the values α, h_0. This is the case e.g. for Euler's method (1.2).

On the other hand, in the numerical solution of so-called stiff initial value problems (cf. [14]) the assumption (2.3) may be nonrealistic as well as superfluous—see [5] for a finite-difference method satisfying the conditions I–V while (2.3) is violated.

2.3. Two-Sided Error Bounds

In order to derive and to investigate bounds for the errors $\tilde{u}_n - u_n$, caused by the local perturbations w_n in (2.2), it is appropriate to introduce the vectorspace

$$X_h = (\mathfrak{R}_h)^{N+1} = \{x \mid x = (x_0, x_1, \ldots, x_N) \quad \text{with all } x_n \in \mathfrak{R}_h\}$$

in which addition and multiplication with real numbers are defined coordinatewise. We define the vectors $w = (w_0, w_1, \ldots, w_N)$, $u = (u_0, u_1, \ldots, u_N)$ and $\tilde{u} = (\tilde{u}_0, \tilde{u}_1, \ldots, \tilde{u}_N)$, which belong to the space X_h. For each $h \in H$ we denote by $\|x\|_h$, or simply by $\|x\|$, an arbitrary seminorm on X_h.

Definition. Let ϕ_h be a real functional defined on X_h (for each $h \in H$) and let γ_0, γ_1 be positive constants (independent of h). If for all $h \in H$ and all $w_0, w_1, \ldots, w_N \in \mathfrak{R}_h$ the relations (2.1), (2.2) imply that

$$\gamma_0 \cdot \phi_h[w] \leqq \|\tilde{u} - u\|_h \leqq \gamma_1 \cdot \phi_h[w], \tag{2.9}$$

then (2.9) is called a *two-sided error bound* for the finite-difference scheme (2.1).

(1.7) provides an example of a two-sided error bound of type (2.9) with $X_h = \mathbb{R}^{N+1}$ and

$$\|x\|_h = \max_{0 \leqq n \leqq N} |x_n|, \tag{2.10}$$

$$\phi_h[w] = \max_{0 \leqq n \leqq N} \left| w_0 + h \sum_{j=1}^{n} w_j \right| \tag{2.11}$$

for $x = (x_0, x_1, \ldots, x_N)$, $w = (w_0, w_1, \ldots, w_N) \in X_h$. Similarly (1.10) is an example of (2.9) with

$$\|x\|_h = \left\{ h \sum_{n=0}^{N} (x_n)^2 \right\}^{\frac{1}{2}}, \tag{2.12}$$

$$\phi_h[w] = \left\{ h \sum_{n=0}^{N} \left(w_0 + h \sum_{j=1}^{n} w_j \right)^2 \right\}^{\frac{1}{2}}. \tag{2.13}$$

Finally, it is clear that (1.11) corresponds to the case where

$$\|x\|_h = |x_N| \tag{2.14}$$

and $\phi_h[w] = \phi[w_0, w_1, \ldots, w_N; h]$.

Definition. The seminorm $\|x\|_h$ on X_h is *absolute* if for all $x = (x_0, x_1, \ldots, x_N)$, $y = (y_0, y_1, \ldots, y_N) \in X_h$ with $|x_n| = |y_n|$ $(n = 0, 1, \ldots, N)$ we have $\|x\|_h = \|y\|_h$.

In the rest of this paper we shall confine ourselves mainly to absolute seminorms $\|x\|_h$ on X_h. Note that (2.10), (2.12) and (2.14) are examples of absolute seminorms.

We conclude this section with the remark that the concept of a two-sided error bound, as defined above, is closely related to the concept of a minimal stability functional defined in [11]. To see this, associate with (2.1) the operator C_h, mapping X_h into itself, defined by

$$(C_h\, x)_n = x_n - c_n \quad (n=0, 1, \ldots, k-1),$$
$$(C_h\, x)_n = h^{-1} \sum_{i=0}^{k} A_{n,i}(x_{n-i}) - F_n(x_0, x_1, \ldots, x_n; h) \quad (n=k, k+1, \ldots, N),$$

where $x=(x_0, x_1, \ldots, x_N)$ denotes an arbitrary vector in X_h. It is easily verified that, given a finite-difference scheme (2.1) and a functional ϕ_h, we have a two-sided error bound (2.9) if an only if ϕ_h is a minimal stability functional for the operator C_h associated with the given finite-difference scheme.

2.4. Formulation of the Main Result

In the introduction we showed that two-sided error bounds for Euler's method with a functional ϕ_h which is allowed to depend on the right-hand member f of the finite-difference method (1.2) can be simultaneously trivial and useless. In order to avoid such trivialities in the investigation of the more general finite-difference scheme (2.1) we focus on error bounds for (2.1) with a functional ϕ_h which is independent of the function F_n appearing in the right-hand member of (2.1.a). The following theorem gives a condition on the seminorm $\|x\|_h$ (viz. statement 3 in Theorem 2) which is necessary and sufficient in order that such non-trivial two-sided error bounds exist for all finite-difference schemes belonging to a specific class $\mathscr{K}$.

In order to define the class $\mathscr{K}$ we assume that $\{\mathfrak{R}_h\}$ is a fixed family of Banach-spaces (h varies through H) and that $k \geqq 1$, $q \geqq 0$ are fixed integers with $q \leqq k$. With $\mathscr{K}$ we denote the class of all finite-difference schemes of type (2.1) satisfying the conditions I–V of the Sections 2.1, 2.2 with the given k, q and $\{\mathfrak{R}_h\}$. It should be understood that two finite-difference schemes belonging to $\mathscr{K}$ are considered to be different from each other if an only if for some $h \in H$ their corresponding operators C_h, as introduced in Section 2.3, are not identical.

Theorem 2. Let $\|x\|_h$ be an absolute seminorm on X_h. Then the three following statements are equivalent.

1. For each finite-difference scheme of the class $\mathscr{K}$ there exists a two-sided error bound of type (2.9) with a functional ϕ_h that is independent of the function F_n appearing in (2.1.a).

2. For each finite-difference scheme of the class $\mathscr{K}$ there exists a two-sided error bound of the form

$$\gamma_0 \cdot \|v\|_h \leqq \|\tilde{u} - u\|_h \leqq \gamma_1 \cdot \|v\|_h \tag{2.15}$$

where γ_0, γ_1 are positive constants independent of h, w_n and where

$v=(v_0, v_1, \ldots, v_N)$ is defined by

$$h^{-1}\sum_{i=0}^{k} A_{n,i}(v_{n-i})=w_n \quad (k\leq n\leq N), \qquad v_n=w_n \quad (0\leq n\leq k-1).$$

3. There is a constant $\delta>0$ such that for all $h\in H$ and $x=(x_0, x_1, \ldots, x_N)\in X_h$, $y=(y_0, y_1, \ldots, y_N)\in X_h$ with

$$y_n=0 \quad (n=0, 1, \ldots, k-1), \qquad y_n=h\sum_{j=0}^{n-q} x_j \quad (n=k, k+1, \ldots, N)$$

the following inequality holds

$$\|y\|_h\leq\delta\cdot\|x\|_h.$$

In order to illustrate Theorem 2 we consider again Euler's method (1.2). A little calculation shows that the seminorms (2.10), (2.12) satisfy condition 3 of the above theorem with $X_h=\mathbb{R}^{N+1}$, $k=q=1$, $\delta=T$. Consequently, in case (2.10) or (2.12) is used, the statements 1 and 2 of the theorem hold and the functional $\phi_h[w]=\|v\|_h$ appearing in statement 2 equals (2.11) or (2.13), respectively. However, if the seminorm (2.14) is chosen, condition 3 of the above theorem is violated.

3. Necessary and Sufficient Conditions for the Existence of Two-Sided Error Bounds

3.1. Introduction

The main purpose of the present chapter is to prove Theorem 2. Before turning to the proof proper we present conditions for the existence of a two-sided error bound of the special form (2.15). In Section 3.2 we give sufficient conditions and in Section 3.3 there are given two necessary conditions for the existence of such an error bound. Then using these results we prove in Section 3.4 a key lemma from which Theorem 2 easily follows.

The following simple lemma will be used repeatedly in the following sections.

Lemma 1. Let the seminorm $\|x\|_h$ on X_h be absolute. Let

$$x(m)=(x_0(m), x_1(m), \ldots, x_N(m))\in X_h$$

denote vectors (for $m=0, 1, \ldots, M$) with

$$|x_n(0)|\leq\sum_{m=1}^{M}|x_n(m)| \quad (n=0, 1, \ldots, N).$$

Then we have

$$\|x(0)\|_h\leq\sum_{m=1}^{M}\|x(m)\|_h.$$

Proof. 1. Let $x(m)=(x_0(m), x_1(m), \ldots, x_N(m))\in X_h$ $(m=0, 1, \ldots, M)$ and $|x_n(0)|\leq\sum_{m=1}^{M}|x_n(m)|$ $(n=0, 1, \ldots, N)$. Define vectors

$$\bar{x}(m)=(\bar{x}_0(m), \bar{x}_1(m), \ldots, \bar{x}_N(m))\in X_h$$

by

$$\bar{x}_n(m)=|x_n(m)|\cdot e \quad (0\leq n\leq N, 0\leq m\leq M),$$

where e is some vector in $\Re_h$ with norm $|e|=1$. We have

$$|x_n(0)|=|\bar{x}_n(0)| \quad \text{and} \quad \sum_{m=1}^{M} |x_n(m)| = \left| \sum_{m=1}^{M} \bar{x}_n(m) \right|,$$

Consequently

$$|\bar{x}_n(0)| \leqq \left| \sum_{m=1}^{M} \bar{x}_n(m) \right| \quad (0 \leqq n \leqq N).$$

Defining the vectors $y=(y_0, y_1, \dots, y_N)$, $z=(z_0, z_1, \dots, z_N) \in X_h$ by

$$y=\bar{x}(0), \qquad z=\sum_{m=1}^{M} \bar{x}(m)$$

we thus have

$$(3.1) \qquad |y_n| \leqq |z_n| \quad (0 \leqq n \leqq N).$$

It will be shown below that

$$(3.2) \qquad \|y\| \leqq \|z\|.$$

Hence $\|\bar{x}(0)\| \leqq \left\| \sum_{m=1}^{M} \bar{x}(m) \right\| \leqq \sum_{m=1}^{M} \|\bar{x}(m)\|$. Since $\|\bar{x}(m)\| = \|x(m)\|$ $(0 \leqq m \leqq M)$ we thus have $\|x(0)\| \leqq \sum_{m=1}^{M} \|x(m)\|$.

2. Now we shall prove (3.2). Define $v=(v_0, v_1, \dots, v_N)$ with

$$v_n = \alpha_n \cdot z_n + \beta_n \cdot (-z_n)$$

where the coefficients α_n, β_n are chosen in such a way that

$$\alpha_n \geqq 0, \ \beta_n \geqq 0, \ \alpha_n + \beta_n = 1, \ (\alpha_n - \beta_n) \cdot |z_n| = |y_n|$$

(this is possible in view of (3.1)). It is easily verified that for arbitrary $r_i \in \Re_h$ we have

$$\|(r_0, \dots, r_{n-1}, v_n, r_{n+1}, \dots, r_N)\| \leqq \|(r_0, \dots, r_{n-1}, z_n, r_{n+1}, \dots, r_N)\|.$$

Applying this inequality successively with $n=0, 1, \dots, N$ and $r_i = z_i$ $(i<n)$, $r_i = v_i$ $(i>n)$ we obtain

$$\|v\| \leqq \|z\|.$$

By definition of v_n we have $|v_n| = |(\alpha_n - \beta_n)\, z_n| = |y_n|$. Consequently

$$\|v\| = \|y\|$$

It follows that (3.2) holds and the lemma has thus been proved.

The above lemma is a slight extension of a result contained in [1] on absolute norms in finite-dimensional vector spaces. The above proof is more elementary than the one in [1].

3.2. Sufficient Conditions for the Existence of Two-Sided Error Bounds

We shall first derive error bounds for finite-difference schemes (2.1) that satisfy conditions which are a bit more general than the conditions III, V of Chapter 2. Assume $\psi_{n,h}[x]$ is a given absolute seminorm on X_h (for each $h \in H$, $k \leqq n \leqq N$). Then our generalized conditions are as follows.

Condition III'. Whenever $h \in H$ and $x_k, x_{k+1}, \ldots, x_N, \; y_0, y_1, \ldots, y_N \in \mathfrak{R}_k$ satisfy

$$h^{-1} \sum_{i=0}^{k} A_{n,i}(y_{n-i}) = x_n \quad (k \leqq n \leqq N), \qquad y_n = 0 \quad (0 \leqq n \leqq k-1),$$

then

$|y_n| \leqq \psi_{n,h}[x]$ $(k \leqq n \leqq N)$ where $x \in X_h$ is defined by

$$x = (0, \ldots, 0, x_k, x_{k+1}, \ldots, x_N).$$

Condition V'. Whenever $h \in H$ and $u_n, \tilde{u}_n, w_n \in \mathfrak{R}_k$ satisfy (2.1), (2.2) and $w_0 = w_1 = \cdots = w_{k-1} = 0$, then

$$|\tilde{u}_n - u_n| \leqq \psi_{n,h}[w] \quad (k \leqq n \leqq N) \quad \text{where} \quad w = (0, \ldots, 0, w_k, w_{k+1}, \ldots, w_N).$$

Clearly the conditions III', V' reduce to the conditions III, V if

$$\psi_{n,h}[x] = \alpha \cdot h \sum_{j=k}^{n} |x_j| \tag{3.3.a}$$

or

$$\psi_{n,h}[x] = \gamma \cdot h \sum_{j=k}^{n} |x_j|, \tag{3.3.b}$$

respectively, for $x = (x_0, x_1, \ldots, x_N) \in X_h$.

In order to formulate concisely a basic assumption that will be made in the subsequent, we introduce the notations

$$\begin{aligned} (E^i x)_n &= x_{n+i} \quad &&(\text{for } 0 \leqq n \leqq N,\ 0 \leqq n+i \leqq N), \\ (E^i x)_n &= 0 &&(\text{for } 0 \leqq n \leqq N,\ n+i < 0 \text{ or } n+i > N), \\ E^i x &= ((E^i x)_0, (E^i x)_1, \ldots, (E^i x)_N) \end{aligned}$$

for $x = (x_0, x_1, \ldots, x_N) \in X_h$ and arbitrary integers i. In the subsequent lemma we make the following assumption about the seminorm $\|x\|_h$.

(3.4) There is a constant $\mu > 0$ such that whenever $q \leqq i \leqq r$, $h \in H$ and the vectors $x = (x_0, x_1, \ldots, x_N)$, $y = (y_0, y_1, \ldots, y_N) \in X_h$ satisfy

$$y_n = 0 \quad (0 \leqq n \leqq k-1), \qquad |y_n| \leqq \psi_{n,h}[E^{-i} x] \quad (k \leqq n \leqq N),$$

then: $\|y\|_h \leqq \mu \cdot \|x\|_h$.

Lemma 2. Consider a finite-difference scheme of type (2.1) satisfying the conditions I, II, IV of Chapter 2 and define

$$\lambda = \sum_{i=q}^{r} \lambda_i. \tag{3.5}$$

Let $\psi_{n,h}[x]$ be an absolute seminorm on X_h and let $\|x\|_h$ be an arbitrary seminorm on X_h. Suppose $\mu > 0$ is a constant with the properties required in (3.4).

Let $u_n, \tilde{u}_n, w_n \in \mathfrak{R}_k$ satisfy (2.1), (2.2) and put $v = (v_0, v_1, \ldots, v_N)$ where

$$h^{-1} \sum_{i=0}^{k} A_{n,i}(v_{n-i}) = w_n \quad (k \leqq n \leqq N), \qquad v_n = w_n \quad (0 \leqq n \leqq k-1). \tag{3.6}$$

a) If condition III′ holds, then $\gamma_0 \cdot \|v\|_h \leqq \|\tilde{u} - u\|_h$ with $\gamma_0 = (1 + \lambda\mu)^{-1}$.

b) If condition V′ holds, then $\|\tilde{u} - u\|_h \leqq \gamma_1 \cdot \|v\|_h$ with $\gamma_1 = (1 + \lambda\mu)$.

Proof. a. By subtracting (3.6) from (2.2) we obtain for vectors r_n defined by

$$r_n = \tilde{u}_n - v_n$$

the following relations

$$(3.7) \quad h^{-1} \sum_{i=0}^{k} A_{n,i}(r_{n-i}) = F_n(\tilde{u}_0, \ldots, \tilde{u}_n; h) \quad (k \leqq n \leqq N), \qquad r_n = c_n \quad (0 \leqq n \leqq k-1).$$

By subtracting (2.1) from (3.7) we obtain

$$(3.8) \quad \begin{aligned} &h^{-1} \sum_{i=0}^{k} A_{n,i}(z_{n-i}) = F_n(\tilde{u}_0, \ldots, \tilde{u}_n; h) - F_n(u_0, \ldots, u_n; h) \quad (k \leqq n \leqq N), \\ &z_n = 0 \quad (0 \leqq n \leqq k-1), \end{aligned}$$

where the vectors z_n are defined by

$$z_n = r_n - u_n.$$

We define the vector $p = (p_0, p_1, \ldots, p_N) \in X_h$ by

$$p_n = 0 \quad (0 \leqq n \leqq k-1), \qquad p_n = F_n(\tilde{u}_0, \ldots, \tilde{u}_n; h) - F_n(u_0, \ldots, u_n; h) \quad (k \leqq n \leqq N).$$

By applying condition III′ (with $y_n = z_n$, $x_n = p_n$) to (3.8) we obtain

$$(3.9) \qquad |z_j| \leqq \psi_{j,h}[p] \quad (k \leqq j \leqq N).$$

In view of condition II we have

$$|p_n| \leqq \sum_{i=0}^{n} \lambda_i |\tilde{u}_{n-i} - u_{n-i}| \quad (0 \leqq n \leqq N).$$

We define the vector $d = (d_0, d_1, \ldots, d_N) \in X_h$ by

$$d_n = \tilde{u}_n - u_n.$$

Since $\lambda_i = 0$ for $i < q$ and $i > r$ (cf. condition II) it thus follows that

$$|p_n| \leqq \sum_{i=q}^{r} |(\lambda_i \cdot E^{-i} d)_n| \quad (0 \leqq n \leqq N).$$

An application of Lemma 1 (with $\|x\|_h$ replaced by $\psi_{j,h}[x]$) yields

$$\psi_{j,h}[p] \leqq \sum_{i=q}^{r} \psi_{j,h}[\lambda_i \cdot E^{-i} d].$$

Defining

$$\alpha_{j,i} = 0 \quad (0 \leqq j \leqq k-1, q \leqq i \leqq r), \qquad \alpha_{j,i} = \psi_{j,h}[E^{-i} d] \quad (k \leqq j \leqq N, q \leqq i \leqq r)$$

we get, in view of (3.9), the inequalities

$$|z_j| \leqq \sum_{i=q}^{r} \lambda_i \alpha_{j,i} \quad (0 \leqq j \leqq N).$$

We put

$$\sigma_j = \sum_{i=q}^{r} \lambda_i \alpha_{j,i} \quad (0 \leqq j \leqq N)$$

and define vectors $g(i) = (g_0(i), g_1(i), \ldots, g_N(i))$ by

$$g_j(i) = \frac{\alpha_{j,i}}{\sigma_j} \cdot z_j \quad (\text{if } \sigma_j \neq 0), \qquad g_j(i) = 0 \quad (\text{if } \sigma_j = 0).$$

Since $|z_j| \leqq \sigma_j$ there follows $|g_j(i)| \leqq \alpha_{j,i}$ $(0 \leqq j \leqq N,\ q \leqq i \leqq r)$. In view of the definition of $\alpha_{j,i}$ we may apply condition (3.4) with $y = g(i)$, $x = d$. It follows that

$$\|g(i)\| \leqq \mu \cdot \|d\| \quad (q \leqq i \leqq r).$$

From the definition of $g(i)$ we have for the vector z defined by

$$z = (z_0, z_1, \ldots, z_N)$$

the representation

$$z = \sum_{i=q}^{r} \lambda_i\, g(i).$$

Hence $\|z\| \leqq \sum_{i=q}^{r} \lambda_i \cdot \|g(i)\| \leqq \sum_{i=q}^{r} \lambda_i \cdot \mu \|d\| = \lambda\mu \cdot \|d\|$. Since $v = d - z$ there follows $\|v\| \leqq \|d\| + \|z\| \leqq (1 + \lambda\mu) \cdot \|d\| = (1 + \lambda\mu) \cdot \|\tilde{u} - u\|$. This proves part a) of the lemma.

b. In order to prove part b) of the lemma we define the vector

$$s = (s_0, s_1, \ldots, s_N) \in X_h$$

by

$$s_n = 0 \quad (0 \leqq n \leqq k-1), \qquad s_n = F_n(\tilde{u}_0, \ldots, \tilde{u}_n; h) - F_n(r_0, \ldots, r_n; h) \quad (k \leqq n \leqq N)$$

and rewrite (3.7) in the form

$$(3.10) \qquad \begin{aligned} h^{-1} \sum_{i=0}^{k} A_{n,i}(r_{n-i}) &= F_n(r_0, \ldots, r_n; h) + s_n \quad (k \leqq n \leqq N), \\ r_n &= c_n + s_n \quad (0 \leqq n \leqq k-1). \end{aligned}$$

An application of condition V′ (with $\tilde{u}_n$, w_n replaced by r_n and s_n, respectively) shows that the vectors r_n form (3.10) satisfy the inequalities

$$|r_j - u_j| \leqq \psi_{j,h}[s] \quad (k \leqq j \leqq N).$$

Since $z_j = r_j - u_j$ we thus have

$$|z_j| \leqq \psi_{j,h}[s] \quad (k \leqq j \leqq N).$$

By arguments similar to those following (3.9) and by noting that $\tilde{u}_n - r_n = v_n$ we arrive at the inequality

$$\|z\| \leqq \lambda\mu \cdot \|v\|.$$

Consequently $\|\tilde{u} - u\| = \|v + z\| \leqq (1 + \lambda\mu) \cdot \|v\|$, which proves part b) of the lemma.

We next turn to the case where the finite-difference scheme (2.1) satisfies the original conditions III, V of Chapter 2. For such a finite-difference scheme we shall obtain a two-sided error bound under a condition on the seminorm $\|x\|_h$ which is more transparent than (3.4). It is assumed that

(3.11) There is a constant $\delta > 0$ such that for all $h \in H$ and

$$x = (x_0, x_1, \ldots, x_N), \quad y = (y_0, y_1, \ldots, y_N) \in X_h$$

with

$$y_n = 0 \quad (0 \leqq n \leqq k-1), \quad |y_n| \leqq h \sum_{j=0}^{n-q} |x_j| \quad (k \leqq n \leqq N)$$

the following inequality holds: $\|y\|_h \leqq \delta \cdot \|x\|_h$.

Condition (3.11) is slightly stronger than the requirement that statement 3 (of Theorem 2) holds. On the other hand in the following lemma it is not required that the seminorm $\|x\|_h$ be absolute.

Lemma 3. Let $\|x\|_h$ be an arbitrary seminorm on X_h satisfying condition (3.11) and consider a finite-difference scheme of type (2.1) satisfying conditions I–V of Chapter 2. Let u_n and $\tilde{u}_n$ satisfy (2.1), (2.2), respectively. Then

$$\gamma_0 \cdot \|v\|_h \leqq \|\tilde{u} - u\|_h \leqq \gamma_1 \cdot \|v\|_h \tag{3.12}$$

where $v = (v_0, v_1, \ldots, v_N)$ is defined by (3.6). Further, if λ is defined by (3.5) and α, γ, δ are as in the conditions III, V, (3.11), respectively, then γ_0 and γ_1 are given by

$$\gamma_0 = (1 + \alpha \lambda \delta)^{-1}, \tag{3.13.a}$$

$$\gamma_1 = (1 + \gamma \lambda \delta). \tag{3.13.b}$$

Proof. a. In order to prove the first of the two inequalities in (3.12) we define $\psi_{n,h}[x]$ by (3.3.a). From (3.11) it easily follows that (3.4) holds with $\mu = \alpha \delta$. Further condition III′ is fulfilled since it is equivalent to condition III. By virtue of Lemma 2 (part a)) we have the first inequality in (3.12) with

$$\gamma_0 = (1 + \lambda \mu)^{-1} = (1 + \alpha \lambda \delta)^{-1}.$$

b. In order to prove the second inequality in (3.12) we define $\psi_{n,h}[x]$ by (3.3.b). Now (3.4) holds with $\mu = \gamma \delta$ and the proof is again completed by an application of Lemma 2 (part b)).

We end this section with the remark that Lemma 3 will be essential in our proof of Theorem 2, while (the more general) Lemma 2 will be used later on in Section 4.2.3.

3.3. Necessary Conditions for the Existence of Two-Sided Error Bounds

In this section we denote by $A_{n,i}$ arbitrary but fixed operators satisfying the conditions I, III of Chapter 2. The function F_n is defined by

$$F_n(x_0, x_1, \ldots, x_n; h) = \beta \cdot (x_{n-k} + x_{n-k+1} + \cdots + x_{n-q}).$$

The finite-difference scheme (2.1) thus reduces to

$$h^{-1} \sum_{i=0}^{k} A_{n,i}(u_{n-i}) = \beta \cdot (u_{n-k} + u_{n-k+1} + \cdots + u_{n-q}) \quad (k \leqq n \leqq N), \tag{3.14}$$

$$u_n = c_n \quad (0 \leqq n \leqq k-1),$$

and the perturbed version (2.2) takes the form

$$(3.15)\quad \begin{aligned} &h^{-1}\sum_{i=0}^{k} A_{n,i}(\tilde{u}_{n-i}) = \beta\cdot(\tilde{u}_{n-k}+\tilde{u}_{n-k+1}+\cdots+\tilde{u}_{n-q})+w_n \quad (k\leqq n\leqq N),\\ &\tilde{u}_n = c_n + w_n \quad (0\leqq n\leqq k-1).\end{aligned}$$

Throughout this section the constant $\beta \neq 0$ is assumed to be chosen in such a way that the finite-difference scheme (3.14) satisfies all of the conditions I–V of Chapter 2. In view of Theorem 1 this can be achieved by choosing β sufficiently close to zero.

The following Lemma's 4, 5 show that a condition on the seminorm $\|x\|_h$ similar to (3.11) must be fulfilled when the following statement (3.16) about a (one-sided) error bound holds for our finite-difference scheme (3.14).

(3.16) There is a constant $\gamma_0 > 0$ such that whenever $h\in H$ and u_n, $\tilde{u}_n$, w_n satisfy (3.14), (3.15) then

$$\gamma_0\cdot\|v\|_h \leqq \|\tilde{u}-u\|_h$$

where $v=(v_0, v_1, \ldots, v_N)$ is defined by

$$h^{-1}\sum_{i=0}^{k} A_{n,i}(v_{n-i}) = w_n \quad (k\leqq n\leqq N), \qquad v_n = w_n \quad (0\leqq n\leqq k-1).$$

Lemma 4. Let $\|x\|_h$ be an arbitrary seminorm on X_h. Assume that (3.16) holds. Then the seminorm $\|x\|_h$ has the following property (3.17).

(3.17) There is a constant $\delta_1 > 0$ such that for all $h\in H$ and $x=(x_0, x_1, \ldots, x_N)$, $y=(y_0, y_1, \ldots, y_N)\in X_h$ with

$$h^{-1}\sum_{i=0}^{k} A_{n,i}(y_{n-i}) = (x_{n-k}+x_{n-k+1}+\cdots+x_{n-q}) \quad (k\leqq n\leqq N),$$

$y_n = 0$ $(0\leqq n\leqq k-1)$, the following inequality holds:

$$\|y\|_h \leqq \delta_1\cdot\|x\|_h.$$

Proof. 1. Using the notations introduced in the proof of Lemma 2 and applying formula (3.8) to the finite-difference scheme (3.14) under consideration we obtain

$$(3.18)\quad h^{-1}\sum_{i=0}^{k} A_{n,i}(z_{n-i}) = \beta\cdot\sum_{i=n-k}^{n-q} d_i \quad (k\leqq n\leqq N), \qquad z_n = 0 \quad (0\leqq n\leqq k-1).$$

Assumption (3.16) implies the inequality $\|v\|_h \leqq (\gamma_0)^{-1}\|d\|_h$ for some $\gamma_0 > 0$. Since $\|z\|_h = \|d-v\|_h \leqq \|d\|_h + \|v\|_h$ we have

$$(3.19)\quad \|z\|_h \leqq [1+(\gamma_0)^{-1}]\cdot\|d\|_h.$$

2. Let x and y satisfy the relations

$$(3.20)\quad h^{-1}\sum_{i=0}^{k} A_{n,i}(y_{n-i}) = \sum_{i=n-k}^{n-q} x_i \quad (k\leqq n\leqq N), \qquad y_n = 0 \quad (0\leqq n\leqq k-1),$$

occurring in (3.17). We define u_n by (3.14) and $\tilde{u}_n$ by $\tilde{u}_n = u_n + \frac{1}{\beta}\cdot x_n$. Further w_n is defined by

$$w_n = \tilde{u}_n - c_n \quad (0\leqq n\leqq k-1), \qquad w_n = h^{-1}\sum_{i=0}^{k} A_{n,i}(\tilde{u}_{n-i}) - \beta\cdot\sum_{i=n-k}^{n-q}\tilde{u}_i \quad (k\leqq n\leqq N).$$

Since $\beta \cdot d = x$ it follows in view of (3.18), (3.20) that $y_n = z_n$ $(0 \leqq n \leqq N)$. By virtue of (3.19) we have

$$\|y\|_h \leqq [1 + (\gamma_0)^{-1}] \cdot \left\| \frac{1}{\beta} \cdot x \right\|_h$$

and (3.17) thus holds with $\delta_1 = [1 + (\gamma_0)^{-1}] \cdot |\beta|^{-1}$. The lemma has thus been proved.

In the next lemma it is assumed that the operators $A_{n,i}$ occurring in (3.14) are such that the following condition (3.21) is fulfilled.

(3.21) For each $h \in H$ there exists a vector $e \in \mathfrak{R}_h$ with norm $|e| = 1$ such that for each j with $k \leqq j \leqq N$ the solution $y_n = y_{n,j}$ of

$$\sum_{i=0}^{k} A_{n,i}(y_{n-i}) = 0 \quad (j < n \leqq N), \qquad y_0 = y_1 = \cdots = y_{j-1} = 0,\ y_j = (A_{j,0})^{-1} e$$

is of the form $y_{n,j} = |y_{n,j}| \cdot e$ and

$$\theta = \inf \{|y_{n,j}| : h \in H;\ j = k, k+1, \ldots, N;\ n = j, j+1, \ldots, N\} > 0.$$

In order to illustrate condition (3.21) we consider the case where $\mathfrak{R}_h = \mathbb{R}$, $k = 1$, $A_{n,0} = -A_{n,1} =$ the identity I, which corresponds to Euler's method (1.2). It is easily verified that (3.21) holds with $e = 1$ and $y_{n,j} = 0$ $(n < j)$, $y_{n,j} = 1$ $(n \geqq j)$, $\theta = 1$. For further examples cf. [13] and Section 3.5, note a.

Lemma 5. Let $\|x\|_h$ be an absolute seminorm on X_h. Assume the operators $A_{n,i}$ appearing in (3.14) satisfy condition (3.21). Then statement (3.16) implies statement 3 of Theorem 2.

Proof. 1. Assuming (3.16) we may conclude from Lemma 4 that the seminorm $\|x\|_h$ satisfies (3.17).

Let $\xi_0, \xi_1, \ldots, \xi_N$ be arbitrary real numbers $\geqq 0$ and define

(3.22) $$\bar{x}_n = \xi_n \cdot e \quad (n = 0, 1, \ldots, N).$$

Let $\bar{x} = (\bar{x}_0, \bar{x}_1, \ldots, \bar{x}_N)$ and define $\bar{y} = (\bar{y}_0, \bar{y}_1, \ldots, \bar{y}_N)$ by requiring that

(3.23) $$h^{-1} \sum_{i=0}^{k} A_{n,i}(\bar{y}_{n-i}) = (\bar{x}_{n-k} + \bar{x}_{n-k+1} + \cdots + \bar{x}_{n-q}) \quad k \leqq n \leqq N),$$
$$\bar{y}_n = 0 \quad (0 \leqq n \leqq k-1).$$

Applying (3.17) with $x = \bar{x}$, $y = \bar{y}$ we obtain

(3.24) $$\|\bar{y}\|_h \leqq \delta_1 \cdot \|\bar{x}\|_h.$$

Since the vectors $y_{n,j}$ $(j = k, k+1, \ldots, N)$ appearing in (3.21) satisfy

$$\sum_{i=0}^{k} A_{n,i}(y_{n-i,j}) = \delta_{n,j} \cdot e \quad (k \leqq n \leqq N), \qquad y_{n,j} = 0 \quad (0 \leqq n \leqq k-1)$$

($\delta_{n,j}$ denoting the Kronecker delta), the solution $\bar{y}_n$ of (3.23) can be written in the form

$$\bar{y}_n = \sum_{j=k}^{N} \sigma_j \cdot y_{n,j} \quad (n = 0, 1, \ldots, N)$$

with

$$\sigma_j = h \cdot (\xi_{j-k} + \xi_{j-k+1} + \cdots + \xi_{j-q}) \quad (j = k, k+1, \ldots, N).$$

Hence for $k \leqq n \leqq N$ we have

$$|\bar{y}_n| = \left| \sum_{j=k}^{n} \sigma_j \cdot y_{n,j} \right| = \sum_{j=k}^{n} \sigma_j |y_{n,j}| \geqq \theta \sum_{j=k}^{n} \sigma_j \geqq \theta \cdot h \sum_{j=0}^{n-q} \xi_j.$$

Since $\xi_j = |\bar{x}_j|$ (cf. (3.22)) we thus obtain

$$(3.25) \qquad |\bar{y}_n| = 0 \quad (0 \leqq n \leqq k-1), \qquad |\bar{y}_n| \geqq \theta \cdot h \sum_{j=0}^{n-q} |\bar{x}_j| \quad (k \leqq n \leqq N).$$

2. In order to prove statement 3 of Theorem 2 we assume that

$$x = (x_0, x_1, \ldots, x_N), \qquad y = (y_0, y_1, \ldots, y_N)$$

are given vectors in X_h with

$$y_n = 0 \quad (0 \leqq n \leqq k-1), \qquad y_n = h \sum_{j=0}^{n-q} x_j \quad (k \leqq n \leqq N).$$

We define numbers ξ_n by

$$\xi_n = |x_n| \quad (n = 0, 1, \ldots, N).$$

Defining $\bar{x}_n$ and $\bar{y}_n$ by (3.22), (3.23) we may apply the results (3.24), (3.25) of part 1. Since

$$(3.26) \qquad |x_n| = |\bar{x}_n| \quad (n = 0, 1, \ldots, N)$$

we have for $k \leqq n \leqq N$

$$|y_n| = h \left| \sum_{j=0}^{n-q} x_j \right| \leqq h \sum_{j=0}^{n-q} |\bar{x}_j|.$$

Using (3.25) there follows

$$|y_n| \leqq \theta^{-1} \cdot |\bar{y}_n| \quad (k \leqq n \leqq N).$$

Since $|y_n| = |\theta^{-1} \bar{y}_n| = 0$ $(0 \leqq n \leqq k-1)$ we obtain in view of Lemma 1 the inequality

$$\|y\|_h \leqq \theta^{-1} \cdot \|\bar{y}\|_h.$$

An application of (3.24) and (3.26) now shows that

$$\|y\|_h \leqq \theta^{-1} \cdot \delta_1 \|x\|_h,$$

which proves statement 3 with $\delta = \theta^{-1} \delta_1$.

3.4. The Proof of Theorem 2

We first state a lemma which is much similar to Theorem 2, the only difference being that the class $\mathscr{K}$ appearing in Theorem 2 is replaced here by an arbitrary class $\mathscr{K}'$ satisfying the conditions i), ii), iii) listed below. The proof of the lemma is based on the Lemmata 1, 3, 5. Next Theorem 2 is proved by verifying that the class $\mathscr{K}$ itself satisfies the conditions i), ii), iii) and by applying the lemma with $\mathscr{K}' = \mathscr{K}$.

The conditions i), ii), iii) are as follows:

i) $\mathscr{K}'$ is a subset of the set $\mathscr{K}$ defined in Section 2.4,

ii) $\mathcal{K}'$ contains a finite-difference scheme of the form (3.14) with $\beta \neq 0$ and with $A_{n,i}$ satisfying (3.21),

iii) If (2.1) is any given finite-difference scheme in K', then there also is a scheme in K' with the same operators $A_{n,i}$ as in the given scheme but with $F_n(x_0, x_1, \ldots, x_n; h) \equiv 0$.

Lemma 6. Let $\|x\|_h$ be an absolute seminorm on X_h. Let $\mathcal{K}'$ denote an arbitrary class of finite-difference schemes (2.1) which satisfies the conditions i), ii), iii). Then the following three propositions are equivalent.

1. For each finite-difference scheme of the class $\mathcal{K}'$ there exists a two-sided error bound of the general type (2.9) with a functional ϕ_h that is independent of the function F_n appearing in (2.1.a).

2. For each finite-difference scheme of the class $\mathcal{K}'$ there exists a two-sided error bound of the special form

$$\gamma_0 \cdot \|v\|_h \leqq \|\tilde{u}-u\|_h \leqq \gamma_1 \cdot \|v\|_h$$

where $v=(v_0, v_1, \ldots, v_N)$ is defined by

$$h^{-1} \sum_{i=0}^{k} A_{n,i}(v_{n-i}) = w_n \quad (k \leqq n \leqq N), \qquad v_n = w_n \quad (0 \leqq n \leqq k-1).$$

3. There is a constant $\delta > 0$ such that for all $h \in H$ and $x=(x_0, x_1, \ldots, x_N) \in X_h$, $y=(y_0, y_1, \ldots, y_N) \in X_h$ with

$$y_n = 0 \quad (n=0, 1, \ldots, k-1), \qquad y_n = h \sum_{j=0}^{n-q} x_j \quad (n=k, k+1, \ldots, N)$$

the following inequality holds

$$\|y\|_h \leqq \delta \cdot \|x\|_h.$$

Proof. We shall prove the lemma by showing successively that statement m of the lemma implies statement $m+1$ (for $m=1, 2$) and that statement 3 implies statement 1.

1. Let statement 1 hold. Consider an arbitrary finite-difference scheme, denoted by C, which belongs to the class $\mathcal{K}'$ and let C_h denote the operator from X_h into itself associated with C as indicated in Section 2.3. Denote by A the scheme, referred to in condition iii), with the same operators $A_{n,i}$ as in C but with $F_n = 0$ and let A_h denote the operator associated with A, i. e.

$$(A_h x)_n = \begin{cases} x_n - c_n & (0 \leqq n \leqq k-1), \\ h^{-1} \sum_{i=0}^{k} A_{n,i}(x_{n-i}) & (k \leqq n \leqq N) \end{cases}$$

for $x=(x_0, x_1, \ldots, x_N) \in X_h$. In view of statement 1 and property iii) there exist two-sided error bounds of type (2.9) for C and A, respectively in which one and the same functional, say ψ_h, occurs.

Let $h \in H$, $w \in X_h$ and assume

$$C_h u = 0, \qquad C_h \tilde{u} = w,$$
$$A_h x = 0, \qquad A_h \tilde{x} = w.$$

Then we have

$$\beta_0 \cdot \psi_h[w] \leqq \|\tilde{u}-u\|_h \leqq \beta_1 \cdot \psi_h[w],$$
$$\alpha_0 \cdot \psi_h[w] \leqq \|\tilde{x}-x\|_h \leqq \alpha_1 \cdot \psi_h[w]$$

for some positive constants $\beta_0, \beta_1, \alpha_0, \alpha_1$, which are independent of h and w. There follows

$$\frac{\beta_0}{\alpha_1} \cdot \|\tilde{x}-x\|_h \leqq \|\tilde{u}-u\|_h \leqq \frac{\beta_1}{\alpha_0} \cdot \|\tilde{x}-x\|_h.$$

Since $\tilde{x}-x=v$ we thus have the two-sided error estimate appearing in statement 2 with $\gamma_0=\beta_0/\alpha_1$, $\gamma_1=\beta_1/\alpha_0$.

2. Let statement 2 hold. Then, in view of property ii), there exists a finite-difference scheme of the form (3.14) for which $\beta \neq 0$ and (3.21), (3.16) hold. An application of Lemma 5 proves statement 3.

3. Let statement 3 hold. The proof of statement 1 is given in two steps.

a. We shall prove that condition (3.11) is fulfilled. Let $x=(x_0, x_1, \ldots, x_N)$, $y=(y_0, y_1, \ldots, y_N)$ be given vectors with

$$y_n=0 \quad (0 \leqq n \leqq k-1), \qquad |y_n| \leqq h \sum_{j=0}^{n-q} |x_j| \quad (k \leqq n \leqq N).$$

Assume e is a vector in $\mathfrak{R}_h$ with norm $|e|=1$. Define

$$\bar{x}_n=|x_n| \cdot e \quad (n=0, 1, \ldots, N),$$

$$\bar{y}_n=0 \quad (0 \leqq n \leqq k-1), \qquad \bar{y}_n=h \sum_{j=0}^{n-q} \bar{x}_j \quad (k \leqq n \leqq N).$$

In view of statement 3 we have for the vectors

$$\bar{x}=(\bar{x}_0, \bar{x}_1, \ldots, \bar{x}_N), \qquad \bar{y}=(\bar{y}_0, \bar{y}_1, \ldots, \bar{y}_N)$$

the inequality

$$\|\bar{y}\|_h \leqq \delta \cdot \|\bar{x}\|_h.$$

Since $|\bar{x}_n|=|x_n|$ $(0 \leqq n \leqq N)$ we have

$$\|\bar{x}\|_h=\|x\|_h.$$

From the equality

$$|\bar{y}_n|=h \sum_{j=0}^{n-q} |x_j| \quad (k \leqq n \leqq N)$$

there follows $|y_n| \leqq |\bar{y}_n|$ $(0 \leqq n \leqq N)$, which shows that

$$\|y\|_h \leqq \|\bar{y}\|_h$$

(cf. Lemma 1). Hence $\|y\|_h \leqq \delta \cdot \|x\|_h$ and (3.11) has thus been proved.

b. Consider any finite-difference scheme belonging to $\mathscr{K}'$. In view of property i) the scheme satisfies the conditions I–V of Chapter 2 and Lemma 3 can be applied. It follows that statement 1 holds with

$$\phi_h[w]=\|v\|_h$$

where $v=(v_0, v_1, \ldots, v_N)$ is defined by (3.6). This completes the proof of Lemma 6.

We finally turn to the proof of Theorem 2.

Proof of Theorem 2. In view of Lemma 6 it is sufficient to show that $\mathscr{K}'=\mathscr{K}$ satisfies the conditions i), ii), iii). Condition i) is trivially fulfilled.

In order to prove ii) we define $A_{n,0}=I$, $A_{n,1}=-I$, $A_{n,i}=0$ $(1<i\leqq k)$, I denoting the identity and 0 the zero-operator. The factor β is defined by $\beta=\frac{1}{2T}$. With these definitions the scheme (3.14) satisfies the conditions I, II, III of Chapter 2 with $\alpha=1$, $\lambda_0\leqq\frac{1}{2T}$. Since $\alpha\lambda_0 h_0\leqq\frac{h_0}{2T}\leqq\frac{1}{2}$ the inequality (2.3) is valid. By Theorem 1 our scheme (3.14) thus belongs to $\mathcal{K}$. Further it satisfies (3.21) with $\theta=1$. Consequently condition ii) is fulfilled.

Finally, an inspection of the conditions I–V of Chapter 2 shows that condition iii) is fulfilled.

3.5. Notes

a) We mention two applications of Lemma 6 where $\mathcal{K}'\neq\mathcal{K}$.

Suppose first that the class $\mathcal{K}'$ consists of all finite-difference schemes of the form (1.2) where $f(t,x)$ is a real function defined and differentiable with respect to x on $[0,T]\times\mathbb{R}$ and where $\frac{\partial}{\partial x}f(t,x)$ satisfies (1.6) (for some $L>0$). Clearly, $\mathcal{K}'$ satisfies the conditions i), ii), iii) of Section 3.4. Defining $\|x\|_h$ by $\|x\|_h=|x_N|$ it follows that statement 3 (with $k=q=1$) of Lemma 6 is violated. Hence by Lemma 6 we may conclude that statement 1 is violated as well. This proves the statement made in Chapter 1 about the non-existence of error bounds of type (1.11).

In a similar fashion Lemma 6 can be used to prove a result contained in [13, Theorem 1]. (Note that assumption 3 of section 2.1 in [13] ensures that condition (3.21) of the present article is fulfilled with $A_{n,0}=I$, $A_{n,1}=-K$).

b) Assume statement 3 of Theorem 2 is true.

Using Lemma 3 and the result contained in part 3.a of the proof of Lemma 6 we find the values $\gamma_0=(1+\alpha\delta\lambda)^{-1}$, $\gamma_1=(1+\gamma\delta\lambda)$ for the coefficients appearing in statement 2 of Theorem 2. Here λ is defined by (3.5), δ is as in statement 3 and α,γ are from the conditions III, V, respectively.

Combining these expressions for γ_0,γ_1 with (2.8), (2.6) we obtain a proof of a two-sided error bound stated in [13, Theorem 2].

c) Let $q=0$ and let $\|x\|_h$ denote an absolute seminorm satisfying condition 3 of Theorem 2.

Let (2.1) be a finite-difference scheme satisfying the conditions I–V of Chapter 2. Then for the errors $\tilde{u}_n-u_n$, caused by arbitrary perturbations w_n in (2.2) with $w_0=w_1=\cdots=w_{k-1}=0$, we have the estimate

$$\|\tilde{u}-u\|_h\leqq\gamma\delta\cdot\|w\|_h \tag{3.27}$$

of $\|\tilde{u}-u\|_h$ in terms of $\|w\|_h$. The estimate (3.27) follows by combining the inequality from condition V with the fact, proved in part 3.a of the proof of Lemma 6, that (3.11) holds.

4. Examples and Applications

4.1. Weighted l_p-norms

In this chapter we deal with seminorms $\|x\|_h=\|x\|_{p,h}$ for $x=(x_0,x_1,\ldots,x_N)\in X_h$ that are given by

$$\|x\|_{p,h}=\left\{h\sum_{i=0}^{N}[\delta_i(h)]^p\cdot|x_i|^p\right\}^{1/p} \tag{4.1.a}$$

for $1 \leqq p < \infty$, and by

(4.1.b) $$\|x\|_{p,h} = \max_{0 \leqq i \leqq N} \delta_i(h) \cdot |x_i|$$

for $p = \infty$. In (4.1) the $\delta_i(h)$ are arbitrary given weights $\geqq 0$. For instance in case $\mathfrak{R}_h = \mathbb{R}$, $\delta_i(h) = 0$ $(0 \leqq i \leqq N-1)$, $\delta_N(h) = h^{-1/p}$, the seminorms (4.1) reduce to (2.14). We use the conventions

$$\alpha/\infty = 0, \quad \alpha/0 = \infty \text{ (for } \alpha > 0), \quad \alpha/0 = 0 \text{ (for } \alpha = 0).$$

The following theorem gives conditions on the weights $\delta_i(h)$ under which the crucial condition 3 (appearing in Theorem 2) is fulfilled by the seminorm (4.1).

Theorem 3. 1. For $p = 1$ condition 3 of Theorem 2 is fulfilled if and only if

(4.2) There is a constant $\delta < \infty$ such that

$$\max_{0 \leqq j \leqq N-q} h \sum_{i=\max(k, j+q)}^{N} \delta_i(h)/\delta_j(h) \leqq \delta \quad \text{(for all } h \in H).$$

2. For $p = \infty$ condition 3 of Theorem 2 is fulfilled if and only if

(4.3) There is a constant $\delta < \infty$ such that

$$\max_{k \leqq i \leqq N} h \sum_{j=0}^{i-q} \delta_i(h)/\delta_j(h) \leqq \delta \quad \text{(for all } h \in H).$$

3. For $1 < p < \infty$ condition 3 of Theorem 2 is fulfilled if (4.2), (4.3) hold.

Proof. 1. Let p be any given number, $1 \leqq p \leqq \infty$, and let $\delta_i(h) \geqq 0$ be given weights.

Suppose condition 3 is fulfilled. Applying the inequality $\|y\|_{p,h} \leqq \delta \cdot \|x\|_{p,h}$ (cf. condition 3) with $x_j = \tau_j \cdot e$, where e is a vector with norm $|e| = 1$ and $\tau_j \geqq 0$, we easily obtain

(4.4.a) $$\left\{h \sum_{i=k}^{N} \left[\delta_i(h) \cdot h \sum_{j=0}^{i-q} \tau_j\right]^p\right\}^{1/p} \leqq \delta \cdot \left\{h \sum_{i=0}^{N} [\delta_i(h)\, \tau_i]^p\right\}^{1/p} \quad \text{(for all } \tau_i \geqq 0)$$

if $1 \leqq p < \infty$, and

(4.4.b) $$\max_{k \leqq i \leqq N} \delta_i(h) \cdot h \sum_{j=0}^{i-q} \tau_j \leqq \delta \cdot \max_{0 \leqq i \leqq N} \delta_i(h)\, \tau_i \quad \text{(for all } \tau_i \geqq 0)$$

if $p = \infty$. Conversely, by applying (4.4) with $\tau_j = |x_j|$, it follows that (4.4) implies condition 3.

We denote by $n = n(h)$ the largest integer $\leqq N$ such that $\delta_i(h) \neq 0$ $(0 \leqq i \leqq n)$ and we put $M = \min(N, n+q)$.

If $M < N$, we have $\delta_{n+1}(h) = 0$ and, applying (4.4) with $\tau_i = 0$ $(i \neq n+1)$, $\tau_i = 1$ $(i = n+1)$, there follows

(4.5) $$\delta_i(h) = 0 \quad \text{(for } k \leqq i,\ M+1 \leqq i \leqq N)$$

and consequently

(4.6.a) $$\left\{\sum_{i=k}^{M} \left[\delta_i(h) \cdot h \sum_{j=0}^{i-q} \tau_j\right]^p\right\}^{1/p} \leqq \delta \cdot \left\{\sum_{i=0}^{M} [\delta_i(h)\, \tau_i]^p\right\}^{1/p} \text{(for all } \tau_i \geqq 0)$$

if $1 \leqq p < \infty$, and

$$\max_{k \leqq i \leqq M} \delta_i(h) \cdot h \sum_{j=0}^{i-q} \tau_j \leqq \delta \cdot \max_{0 \leqq i \leqq M} \delta_i(h)\, \tau_i \quad \text{(for all } \tau_i \geqq 0) \tag{4.6.b}$$

if $p=\infty$. Conversely, (4.6) in combination with (4.5) yields (4.4). Hence condition 3 is equivalent to the requirements (4.5), (4.6).

2. Let $\xi_0, \xi_1, \ldots, \xi_M$ be arbitrary real numbers. Applying (4.6) with

$$\tau_i = |\xi_i| / \delta_i(h) \quad (0 \leqq i \leqq n), \qquad \tau_i = 0 \quad (i > n)$$

there follows

$$\left\{\sum_{i=k}^{M} \left| \sum_{j=0}^{i-q} [h\,\delta_i(h)/\delta_j(h)] \cdot \xi_j \right|^p \right\}^{1/p} \leqq \delta \cdot \left\{\sum_{i=0}^{M} |\xi_i|^p \right\}^{1/p} \tag{4.7.a}$$

if $1 \leqq p < \infty$, and

$$\max_{k \leqq i \leqq M} \left| \sum_{j=0}^{i-q} [h\,\delta_i(h)/\delta_j(h)] \cdot \xi_j \right| \leqq \delta \cdot \max_{0 \leqq i \leqq M} |\xi_i| \tag{4.7.b}$$

if $p=\infty$. Defining the operator D from $\mathbb{R}^{M+1}$ into itself by $D\xi = \eta$ where

$$\xi = (\xi_0, \xi_1, \ldots, \xi_M), \qquad \eta = (\eta_0, \eta_1, \ldots, \eta_M),$$

$$\eta_i = 0 \quad (0 \leqq i \leqq k-1), \qquad \eta_i = \sum_{j=0}^{i-q} [h\,\delta_i(h)/\delta_j(h)] \cdot \xi_j \quad (k \leqq i \leqq M)$$

the relations (4.7) can be written as

$$\|D\xi\|^{(p)} \leqq \delta \cdot \|\xi\|^{(p)} \tag{4.8}$$

where $\|\cdot\|^{(p)}$ stands for the usual l_p-norm in $\mathbb{R}^{M+1}$ (cf. [7]). Since (4.8) holds for all $\xi \in \mathbb{R}^{M+1}$ there follows

$$\|D\|^{(p)} \leqq \delta \tag{4.9}$$

where $\|D\|^{(p)}$ denotes the norm of D induced by the norm in $\mathbb{R}^{M+1}$. Conversely, (4.9) implies (4.6). Hence we may conclude that condition 3 is equivalent to (4.5), (4.9).

3. Using standard expressions for $\|D\|^{(p)}$ if $p=1$ or $p=\infty$ we have

$$\max_{0 \leqq j \leqq M-q} \sum_{i=\max(k,\, j+q)}^{M} h\,\delta_i(h)/\delta_j(h) = \|D\|^{(1)}, \tag{4.10}$$

$$\max_{k \leqq i \leqq M} \sum_{j=0}^{i-q} h\,\delta_i(h)/\delta_j(h) = \|D\|^{(\infty)}. \tag{4.11}$$

Applying the conventions for a division by zero stated above it thus follows that (4.2) is equivalent to (4.5), (4.9) (with $p=1$) and that (4.3) is equivalent to (4.5), (4.9) (with $p=\infty$).

Since condition 3 is equivalent to (4.5), (4.9), the statements 1, 2 of Theorem 3 have thus been proved.

Let $1 < p < \infty$ and assume (4.2), (4.3) hold. It follows that (4.5) and (4.9) (with $p=1$ and with $p=\infty$) hold. By virtue of the Riesz convexity theorem (cf. [10])

we have

$$\|D\|^{(p)} \leqq \{\|D\|^{(1)}\}^{1/p} \cdot \{\|D\|^{(\infty)}\}^{1-1/p}$$

and consequently

$$\|D\|^{(p)} \leqq \delta.$$

Hence (4.5), (4.9) (with the given p, $1<p<\infty$) hold. Thus condition 3 is fulfilled and the theorem has been proved.

4.2. Systems of Ordinary Differential Equations

4.2.1. General Multistep Methods

Let there be given an initial value problem for a system of s ordinary differential equations which, by using vector notation, can be written in the form

$$\frac{d}{dt} U(t) = f(t, U(t)) \quad (0 \leqq t \leqq T), \qquad U(0) = c. \tag{4.12}$$

In (4.12) c denotes a given vector in the s-dimensional real vectorspace $\mathbb{R}^s$ and f is a given mapping from $[0, T] \times \mathbb{R}^s$ into $\mathbb{R}^s$. Assume the Jacobian matrix of the function $f(t, x)$, denoted by $J(t, x) = \frac{\partial}{\partial x} f(t, x)$, exists and is continuous on $[0, T] \times \mathbb{R}^s$. Let $|J(t, x)|$ denote the norm of the matrix $J(t, x)$ subordinate to some given vector norm $|x|$ in $\mathbb{R}^s$ and assume

$$|J(t, x)| \leqq L \quad (0 \leqq t \leqq T,\ x \in \mathbb{R}^s), \tag{4.13}$$

where L denotes a given positive constant.

We consider the approximation of a solution $U(t)$ to (4.12) by the general multistep method

$$\begin{aligned} &h^{-1}(\alpha_k u_n + \alpha_{k-1} u_{n-1} + \cdots + \alpha_0 u_{n-k}) \\ &\qquad = f_n(u_{n-k}, \ldots, u_{n-1}, u_n; h) \quad (n = k, k+1, \ldots, N), \end{aligned} \tag{4.14.a}$$

$$u_n = c_n \quad (n = 0, 1, \ldots, k-1). \tag{4.14.b}$$

The c_n are starting vectors in $\mathbb{R}^s$ found e.g. by a Taylor expansion of $U(t)$ at $t=0$. In (4.14.a) the α_i denote real constants with $\alpha_k \neq 0$ and the vectorvalued function $f_n(x_0, x_1, \ldots, x_k; h)$—which depends on the given function f—is defined for $x_i \in \mathbb{R}^s$, $h \in H = (0, h_0] \subset (0, T/k]$, $n = k, k+1, \ldots, N$ and is assumed to satisfy a Lipschitz condition

$$|f_n(\tilde{x}_0, \tilde{x}_1, \ldots, \tilde{x}_k; h) - f_n(x_0, x_1, \ldots, x_k; h)| \leqq \sum_{i=0}^{k} \lambda_i |\tilde{x}_{k-i} - x_{k-i}| \tag{4.15}$$

where $\lambda_0, \lambda_1, \ldots, \lambda_k$ are constants independent of h, n, $\tilde{x}_i$ and x_i.

It is easily verified that many well known methods (general linear multistep methods, Runge-Kutta methods as well as generalizations of these methods, cf. [14]) are step-by-step methods of the form (4.14) satisfying the Lipschitz condition (4.15).

Clearly (4.14) is an example of the general finite-difference scheme (2.1) with $\mathfrak{R}_h = \mathbb{R}^s$, $A_{n,i}(x) \equiv \alpha_{k-i} x$,

$$F_n(x_0, x_1, \ldots, x_n; h) \equiv f_n(x_{n-k}, \ldots, x_{n-1}, x_n; h).$$

The general conditions I, II of Chapter 2 are satisfied here with $q=0$, $r=k$ and in order also to ensure that condition III is fulfilled we assume that all (complex) roots ζ of the equation

$$\alpha_k \zeta^k + \cdots + \alpha_1 \zeta + \alpha_0 = 0$$

have a modulus $|\zeta| \leqq 1$ and that the roots with $|\zeta| = 1$ are simple (cf. [14]). Finally we assume that the conditions IV, V of Chapter 2 are satisfied as well. By Theorem 1 this last assumption doesn't impose any new requirements on the multistep method (4.14) whenever $\lambda_0 = 0$ or, in case $\lambda_0 \neq 0$, the maximal stepsize h_0 is sufficiently small.

4.2.2. The Existence of Two-Sided Error Bounds

Since the conditions I–V are satisfied the multistep method (4.14) belongs to the class $\mathscr{K}$ introduced in Section 2.4 with $\mathfrak{R}_h = \mathbb{R}^s$, $H = (0, h_0]$, $q = 0$.

With the above definitions of $A_{n,i}$, F_n the perturbed finite-difference scheme (2.2) reduces to the following perturbed version of (4.14):

$$\text{(4.16.a)} \quad \begin{aligned} &h^{-1}(\alpha_k \tilde{u}_n + \alpha_{k-1} \tilde{u}_{n-1} + \cdots + \alpha_0 \tilde{u}_{n-k}) \\ &= f_n(\tilde{u}_{n-k}, \ldots, \tilde{u}_{n-1}, \tilde{u}_n; h) + w_n \quad (n = k, k+1, \ldots, N), \end{aligned}$$

$$\text{(4.16.b)} \quad \tilde{u}_n = c_n + w_n \quad (n = 0, 1, \ldots, k-1).$$

An application of Theorem 2 to the step-by-step method (4.14) yields in combination with Theorem 3 the following

Theorem 4. Let $1 \leqq p \leqq \infty$ and let $\|x\|_{p,h}$ be defined by (4.1). Assume the weights $\delta_i(h)$ satisfy (4.2) or (4.3) (with $q=0$) if $p=1$ or $p=\infty$, respectively and assume both (4.2) and (4.3) hold (with $q=0$) if $1 < p < \infty$. Then the errors $\tilde{u}_n - u_n$ caused by perturbations w_n occurring in (4.16) satisfy

$$\text{(4.17)} \quad \gamma_0 \cdot \|v\|_{p,h} \leqq \|\tilde{u} - u\|_{p,h} \leqq \gamma_1 \cdot \|v\|_{p,h}$$

where $v = (v_0, v_1, \ldots, v_N)$ is defined by the relations

$$\text{(4.18)} \quad \begin{aligned} \alpha_k v_n + \alpha_{k-1} v_{n-1} + \cdots + \alpha_0 v_{n-k} &= h\, w_n \quad (k \leqq n \leqq N), \\ v_n &= w_n \quad (0 \leqq n \leqq k-1) \end{aligned}$$

and γ_0, γ_1 are positive constants independent of h and $w_0, w_1, \ldots, w_N$.

Example 1. With $1 \leqq p \leqq \infty$, $\delta_i(h) \equiv 1$ the assumptions of Theorem 4 are fulfilled and (4.17) provides a straightforward generalization to general l_p-norms with $1 \leqq p \leqq \infty$ of a result for $p = \infty$ contained in [11].

Example 2. Let $\alpha_k = 1$, $\alpha_{k-1} = -1$, $\alpha_i = 0$ $(i < k-1)$, which is the case e.g. if (4.14) stands for a Runge-Kutta method or for a predictor-corrector scheme with an Adams-Moulton corrector formula. Now (4.17) takes on a very attractive form since v_n (cf. (4.18)) is simply given by the expressions

$$\text{(4.19)} \quad v_n = w_n \quad (0 \leqq n \leqq k-1), \qquad v_n = w_{k-1} + h \sum_{j=k}^{n} w_j \quad (k \leqq n \leqq N).$$

If the coefficients α_i are not of the simple form indicated, there are still cases where (4.18) may be replaced by (4.19) (cf. [11] where this is done for the case $p=\infty$, $\delta_i(h)\equiv 1$).

Example 3. Let $p=\infty$, $\delta_i(h)=h^{\varepsilon/(i+1)}$ where ε is a real parameter. It may be verified that (4.3) holds if an only if $\varepsilon\leqq 1$. Applying theorem 4 with any $\varepsilon\leqq 1$ and observing that $|\tilde{u}_N-u_N|\leqq\max_n[\delta_n(h)/\delta_N(h)]\cdot|\tilde{u}_n-u_n|$ there follows

$$|\tilde{u}_N-u_N|\leqq\gamma_1\cdot\max_{0\leqq n\leqq N}h^{\varepsilon\cdot[1/(n+1)-1/(N+1)]}\cdot|v_n|. \tag{4.20}$$

For $\varepsilon=1$ the estimate (4.20) is essentially more refined than for $\varepsilon=0$, since the factors

$$h^{[1/(n+1)-1/(N+1)]}$$

tend to zero if $h\to 0$ while n remains sufficiently small. As the choice $\varepsilon=0$ corresponds to the standard l_∞-norm, this example clearly shows that weights $\delta_i(h)\not\equiv 1$ in (4.1) can lead to an improved estimate of $|\tilde{u}_N-u_N|$.

4.2.3. The Constants γ_0, γ_1

In this section we indicate how expressions can be obtained for the constants γ_0, γ_1 appearing in (4.17). We confine our considerations to the norm $\|x\|_h=\|x\|_{p,h}$ defined by

$$\|x\|_{p,h}=\left\{h\sum_{i=0}^{N}|x_i|^p\right\}^{1/p}\quad(\text{if } 1\leqq p<\infty),$$

$$\|x\|_{p,h}=\max_{0\leqq i\leqq N}|x_i|\quad(\text{if } p=\infty),$$

which, according to Example 1 of Section 4.2.2, satisfies the assumptions of Theorem 4.

From Lemma 3 it is clear that expressions easily follow from (3.13) by computing a number δ with the properties required in (3.11) (cf. also Section 3.5, note b)). However, it turns out that in general better values for γ_0, γ_1 are obtained by using Lemma 2.

We first consider the computation of γ_0. We shall apply Lemma 2 (part a)) with $\psi_{n,h}[x]$ defined by (3.3.a) so as to ensure that condition III' is fulfilled. In order to find a number μ with the properties required in (3.4), we define for $0\leqq i\leqq k$, $0<h\leqq h_0$ the operator $S_{i,h}$ from $\mathbb{R}^{N+1}$ into itself by $S_{i,h}[\xi]=\eta$ where $\xi=(\xi_0,\xi_1,\ldots,\xi_N)$, $\eta=(\eta_0,\eta_1,\ldots,\eta_N)$ and

$$\eta_n=0\quad(0\leqq n\leqq k-1),\qquad \eta_n=\alpha\cdot h\sum_{j=k-i}^{n-i}\xi_j\quad(k\leqq n\leqq N).$$

It is easily verified that a number μ is consistent with the requirements of (3.4) whenever

$$\mu\geqq\sup_{i,h}\|S_{i,h}\|^{(p)}.$$

Here $\|S\|^{(p)}$ denotes the norm of S induced by the usual l_p-norm in $\mathbb{R}^{N+1}$ (cf. [7]). Since $\|S_{i,h}\|^{(\infty)}\leqq\alpha\,T$, $\|S_{i,h}\|^{(1)}\leqq\alpha\,T$, we obtain by using the Riesz convexity theorem (cf. [10]),

$$\mu=\alpha\,T. \tag{4.21}$$

From Lemma 2 we thus have

$$\gamma_0 = (1+\alpha\lambda T)^{-1} \quad (\text{for } 1 \leqq p \leqq \infty), \tag{4.22}$$

where

$$\lambda = \lambda_0 + \lambda_1 + \cdots + \lambda_k$$

and α is the constant appearing in condition III.

We next turn to the computation of γ_1. Let us first assume that $\alpha\lambda_0 h_0 < 1$. From (2.7) it then follows that condition V′ holds with

$$\psi_{n,h}[x] = \beta \cdot h \sum_{j=k}^{n} (1+\beta\lambda \cdot h)^{n-j} |x_j| \quad (k \leqq n \leqq N). \tag{4.23}$$

It may be shown, by an argument similar to the one leading to (4.21), that the requirements of (3.4) are fulfilled with $\psi_{n,h}[x]$ from (4.23) and

$$\mu = \frac{1}{\lambda} \cdot [\exp(\beta\lambda T) - 1]. \tag{4.24}$$

By Lemma 2 (part b)) we thus have

$$\gamma_1 = \exp(\beta\lambda T) \quad (\text{for } 1 \leqq p \leqq \infty), \tag{4.25}$$

where $\beta = \alpha \cdot (1 - \alpha\lambda_0 h_0)^{-1}$ and α, λ are as in (4.22).

In order to illustrate (4.22), (4.25), we assume (4.14) stands for Euler's method, i.e. $k=1$, $\alpha_1 = 1$, $\alpha_0 = -1$, $f_n(x_0, x_1; h) = f((n-1)h, x_0)$. We then have $\lambda_0 = 0$, $\lambda = \lambda_1 = L$, $\beta = \alpha = 1$ and (4.22), (4.25) reduce to

$$\gamma_0 = (1+LT)^{-1}, \quad \gamma_1 = \exp(LT) \quad (\text{for } 1 \leqq p \leqq \infty). \tag{4.26}$$

For $p=1$ and $p=\infty$ the expressions (4.26) have also been derived in [6]. Further it has been shown in [6] that, if $p=1$ or $p=\infty$, there exist no constants $\gamma_0 = \gamma_0(L, T)$, $\gamma_1 = \gamma_1(L, T)$ which are larger or smaller, respectively than those given in (4.26) and for which, at the same time, the corresponding two-sided error bound (4.17) remains valid for *all* initial value problems (4.12) satisfying (4.13).

On the other hand, if one is willing to make more assumptions about the initial-value problem (4.12) or the multistep method (4.14) than were made in deriving (4.25) one can easily obtain expressions for γ_1 that are smaller than (4.25). In fact, for many actual initial value problems of type (4.12) and corresponding multistep methods (4.14) condition V′ is known to be fulfilled with a seminorm $\psi_{n,h}[x]$ that is essentially smaller than the one defined in (4.23) (cf. [2–5, 8, 9]). For such seminorms it may be possible to satisfy the requirements of (3.4) with a constant μ which is smaller than (4.24). In these cases an application of Lemma 2 (part b)) obviously yields a smaller constant γ_1 than (4.25).

As an illustration of such an improvement on (4.25) we first consider the application of Euler's method to the initial value problem (4.12) where the Jacobian matrix satisfies, in addition to (4.13),

$$\mu[J(t, x)] \leqq M \quad (0 \leqq t \leqq T, \ x \in \mathbb{R}^s). \tag{4.27}$$

Here M is any constant (which may be negative) and $\mu[A] = \lim_{h \downarrow 0} h^{-1} \cdot \{|I + hA| - 1\}$ stands for the logarithmic norm of the matrix A (cf. [3], [5]). Condition V′ is fulfilled with

$$\psi_{n,h}[x] = h \sum_{j=1}^{n} [1 + h \cdot (M+\varepsilon)]^{n-j} |x_j| \qquad (1 \leqq n \leqq N) \tag{4.28}$$

where $\varepsilon = \varepsilon(h)$ is a quantity satisfying $\varepsilon(h) \downarrow 0$ (for $h \downarrow 0$) (see [9], [3, p. 11]). It follows that the requirements in (3.4) are fulfilled with

$$\mu = \Theta \cdot \frac{1}{M} \{\exp(MT) - 1\}$$

where $\Theta = \Theta(h_0)$ depends on h_0 in such a way that

$$\lim_{h_0 \downarrow 0} \Theta(h_0) = 1.$$

An application of Lemma 2 (part b)) yields

$$\gamma_1 = 1 + \Theta \cdot \frac{L}{M} \cdot \{\exp(MT) - 1\} \quad (\text{for } 1 \leqq p \leqq \infty). \tag{4.29}$$

In case $M=0$, we have to replace the expression $\frac{1}{M} \cdot \{\exp(MT) - 1\}$ by T. Clearly, for $M < L$ and h_0 sufficiently small, the present value for γ_1 is smaller than the one in (4.26).

We state a similar result for the *backward Euler method* (i.e. (4.14) with $k=1$, $\alpha_1 = 1$, $\alpha_0 = -1$, $f_n(x_0, x_1; h) = f(nh, x_1)$). Using the error bound of type V′ derived in [5] we obtain, for $M < 0$, the expression

$$\gamma_1 = 1 + \frac{L}{M} \cdot \{\exp(MT) - 1\} \qquad (\text{for } 1 \leqq p \leqq \infty). \tag{4.30}$$

We note that in the derivation of (4.30) we have not assumed that $\alpha \lambda_0 h_0 = L h_0 < 1$, as was necessary in obtaining (4.25).

We end this section by indicating how Lemma 2 can be applied to the error bounds derived in [3, 4]. In these error bounds a $\psi_{n,h}[x]$ occurs that is generally smaller than (4.23), but these bounds do not fit in the framework of our condition V′. The fact is that they do not apply to the difference $\tilde{u}_n - u_n$ but to $\tilde{u}_n - \bar{u}_n$ instead. Here $\bar{u}_n$ denotes the value of (a smooth function closely related to) $U(t)$ at the point $t = t_n$. This formal difficulty in the application of Lemma 2 can be overcome e.g. by defining $\bar{f}_n(x_0, x_1, \ldots, x_k; h) = f_n(x_0, x_1, \ldots, x_k; h) + \overline{w}_n$ where $\overline{w}_n$ equals the perturbation occurring if $\bar{u}_n$ is substituted into (4.14.a). A combination of the results in [3], [4] with Lemma 2 (where now u_n, f_n are to be replaced by $\bar{u}_n, \bar{f}_n$) then yields two-sided error bounds for $\|\tilde{u} - \bar{u}\|$ in terms of $w_n - \overline{w}_n$ with a constant γ_1 which is smaller than (4.25).

4.3. A Parabolic Differential Equation

In this section we briefly indicate how the concepts of the preceding chapters can be applied in the numerical solution of partial differential equations.

We consider the numerical solution of the simple semilinear parabolic initial-value problem

$$\begin{aligned} &\frac{\partial}{\partial t} U(s,t) - g(s,t) \cdot \frac{\partial^2}{\partial s^2} U(s,t) = f(s,t,U(s,t)), \\ &U(s,0) = c(s), \\ &-\infty < s < \infty, \quad 0 \leqq t \leqq T \end{aligned} \tag{4.31}$$

by the finite-difference scheme

$$
\begin{aligned}
&h^{-1}[u_n(s)-u_{n-1}(s)]-(\Delta s)^{-2}g(s,t_{n-1})\cdot[u_{n-1}(s-\Delta s)-2u_{n-1}(s)\\
&+u_{n-1}(s+\Delta s)]=\beta_n\cdot f(s,t_{n-1},u_{n-1}(s)),\\
&u_0(s)=c_0(s).
\end{aligned}
\tag{4.32}
$$

h and Δs denote positive increments of the variables t and s, respectively such that $0<h\leqq T$, $h/(\Delta s)^2=\varrho$, ϱ denoting some positive constant. In (4.32) the integer n takes the values $1, 2, \ldots, N$, the number s varies through the grid

$$G_h=\{s|s=m\Delta s;\ m=0,\ \pm 1,\ \pm 2, \ldots\}$$

and c_0 denotes the restriction of the given function c to G_h. With β_n we denote coefficients, to be specified below, with

$$B=\sup_n|\beta_n|<\infty.$$

Further $u_n(s)$ denotes an approximation of $U(s, t)$ at the point

$$(s, t)=(s, t_n)=(m\Delta s,\ nh).$$

With $\mathfrak{R}_h$ we denote the vectorspace consisting of all real functions defined and bounded on G_h and for $a\in\mathfrak{R}_h$ we define the norm

$$|a|=\sup\{|a(s)|: s\in G_h\}. \tag{4.33}$$

Assuming that the given functions c, g, f (cf. (4.31)) are bounded we may write the scheme (4.32) in the general form (2.1) with $k=1$, $A_{n,0}=$ the identity I and with operators $A_{n,1}$, F_n defined by

$$
\begin{aligned}
(A_{n,1}[a])\,(s)=&-\varrho g(s,t_{n-1})\cdot a(s-\Delta s)-(1-2\varrho g(s,t_{n-1}))\cdot a(s)\\
&-\varrho g(s,t_{n-1})\cdot a(s+\Delta s),\\
F_n(x_0, x_1, \ldots, x_n;\ h)\,(s)=&\ \beta_n\cdot f(s,t_{n-1},x_{n-1}(s))
\end{aligned}
$$

for $s\in G_h$ and $a, x_0, x_1, \ldots, x_n\in\mathfrak{R}_h$. Further we assume

$$0<\alpha_0\leqq g(s,t)\leqq\alpha_1,$$

and

$$|f(s,t,\tilde{\xi})-f(s,t,\xi)|\leqq L\cdot|\tilde{\xi}-\xi|$$

for some constants α_0, α_1, L independent of $s, t, \tilde{\xi}, \xi$. It is easily verified that the general conditions I, II, IV are now satisfied with $q=r=1$, $\lambda_1=BL$, $\lambda_0=0$. Finally we require the ratio ϱ to satisfy

$$\varrho\leqq 1/(2\alpha_1),$$

so as to ensure that condition III is fulfilled (with $\alpha=1$). In view of Theorem 1, condition V is fulfilled as well.

With the above definitions (4.32) thus provides an example of the general scheme (2.1) satisfying the conditions I–V of Chapter 2. In our example the perturbed scheme (2.2) takes the form

$$(4.34)\qquad \begin{aligned} &h^{-1}[\tilde{u}_n(s)-\tilde{u}_{n-1}(s)]-(\Delta s)^{-2}g(s,t_{n-1})\cdot[\tilde{u}_{n-1}(s-\Delta s)-2\tilde{u}_{n-1}(s)\\ &\quad+\tilde{u}_{n-1}(s+\Delta s)]=\beta_n\cdot f(s,t_{n-1},\tilde{u}_{n-1}(s))+w_n(s),\\ &\tilde{u}_0(s)=c_0(s)+w_0(s).\end{aligned}$$

Applying the Theorems 2, 3 (with $p=\infty$, $\delta_i(h)\equiv 1$) we obtain the estimate

$$(4.35)\qquad \gamma_0\cdot\max_{0\leqq n\leqq N}|v_n|\leqq\max_{0\leqq n\leqq N}|\tilde{u}_n-u_n|\leqq\gamma_1\cdot\max_{0\leqq n\leqq N}|v_n|$$

for the errors $\tilde{u}_n(s)-u_n(s)$ caused by the perturbations w_n in (4.34). Proceeding as in Section 4.2 (see (4.22), (4.25)) we have

$$\gamma_0=(1+BLT)^{-1},\qquad \gamma_1=\exp(BLT).$$

Using the definition of v_n (in statement 2 of Theorem 2) it follows easily that

$$(4.36)\quad v_n=h\cdot(w_n+A_nw_{n-1}+A_nA_{n-1}w_{n-2}+\cdots+A_nA_{n-1}\ldots A_2w_1)+A_nA_{n-1}\ldots A_1w_0$$

where A_n stands for $A_n=-A_{n,1}$.

We conclude this section with two applications of (4.35). It will be assumed that all functions appearing in (4.31) have bounded partial derivatives of sufficiently high order.

Application 1. Let $\beta_n\equiv 1$. It is well known that

$$(4.37)\qquad \sup\{|U(s,t)-u_n(s)|:0\leqq t=t_n\leqq T,\ s\in G_h\}\leqq\gamma\cdot h$$

for some constant γ independent of $h\in(0,T]$. We assume that the expression

$$\frac{1}{2}\cdot\frac{\partial^2}{\partial t^2}U(s,t)-\frac{1}{12\varrho}g(s,t)\cdot\frac{\partial^4}{\partial s^4}U(s,t)$$

(which after a multiplication by h stands for a first order approximation of the truncation error of method (4.32)) does not vanish identically for $-\infty<s<\infty$, $0\leqq t\leqq T$. Then, by using the inequality $\max_n|\tilde{u}_n-u_n|\geqq\gamma_0\cdot\max_n|v_n|$ (cf. (4.33), (4.35), (4.36)) with $\tilde{u}_n$ defined by $\tilde{u}_n(s)=U(s,t_n)$, it can be proved that also

$$(4.38)\qquad \sup\{|U(s,t)-u_n(s)|:0\leqq t=t_n\leqq T,\ s\in G_h\}\geqq\gamma'\cdot h$$

(for h sufficiently small). In (4.38) γ' denotes some positive constant independent of h. The inequality (4.38) proves that in general the error estimate (4.37) cannot essentially be improved.

Application 2. Let $\beta_n=0$ (if n is odd), $\beta_n=2$ (if n is even). An application of the upper bound for $\max_n|\tilde{u}_n-u_n|$ appearing in (4.35), with $\tilde{u}_n$ again defined by $\tilde{u}_n(s)=U(s,t_n)$, shows that still with this choice of β_n the estimate

$$(4.39)\qquad \sup\{|U(s,t)-u_n(s)|:0\leqq t=t_n\leqq T,\ s\in G_h\}=\mathcal{O}(h)$$

holds. We note that the number of evaluations of the function f in the finite-difference scheme (4.32) now has been reduced by a factor 1/2.

More generally, using (4.35), it can be proved that (4.39) holds whenever the coefficients β_n appearing in (4.32) satisfy

$$\sup_n \left| \sum_{j=1}^{n} (1-\beta_j) \right| < \infty.$$

References

1. Bauer, F.L., Stoer, J., Witzgall, C.: Absolute and monotonic norms. Numer. Math. **3**, 257–264 (1961)
2. Carr, J.W.: Error bounds for the Runge-Kutta single-step integration process. J. ACM **5**, 39–44 (1958)
3. Dahlquist, G.: Stability and error bounds in the numerical integration of ordinary differential equations. Trans. Roy. Inst. Technol., Stockholm, Nr. 130 (1959)
4. Dahlquist, G.: Error analysis for a class of methods for stiff nonlinear initial value problems. Report TRITA-NA-7511. Dept. of Information Processing, Roy. Inst. Technol., Stockholm (1975)
5. Desoer, C.A., Haneda, H.: The measure of a matrix as a tool to analyze computer algorithms for circuit analysis. IEEE Trans. Circuit Theory **19**, 480–486 (1972)
6. Didelot, G.: Sur l'existence d'estimations optimales d'erreur dans la résolution numérique d'équations différentielles. Rapport de D.E.A., Analyse Numérique. Université Scientifique et Médicale de Grenoble (1975)
7. Edwards, R.E.: Functional analysis. New York: Holt, Rinehart and Winston 1965
8. Galler, B.A., Rozenberg, D.P.: A generalization of a theorem of Carr on error bounds for Runge-Kutta procedures. J. ACM **7**, 57–60 (1960)
9. Henrici, P.: Problems of stability and error propagation in the numerical integration of ordinary differential equations. In: Proceedings of the International Congress of Mathematicians 1962. Djursholm: Institut Mittag-Leffler 1963
10. Riesz, M.: Sur les maxima des formes bilinéaires et sur les fonctionnelles linéaires. Acta Mathematica **46**, 465–497 (1926)
11. Spijker, M.N.: On the structure of error estimates for finite-difference methods. Numer. Math. **18**, 73–100 (1971)
12. Spijker, M.N.: Optimum error estimates for finite-difference methods. Acta Universitatis Carolinae-Mathematica et Physica **15**, 159–164 (1974)
13. Spijker, M.N.: Two-sided error estimates in the numerical solution of initial value problems. In: Numerische Behandlung nichtlinearer Integro-differential- und Differentialgleichungen, Lecture Notes in Mathematics 395, pp. 109–122. Berlin: Springer 1974
14. Stetter, H.J.: Analysis of discretization methods for ordinary differential equations. Berlin: Springer-Verlag 1973
15. Stummel, F.: Approximation methods in analysis. Lecture Notes Series of Aarhus Universitet (1973)
16. Stummel, F.: Biconvergence, bistability and consistency of one-step methods for the numerical solution of initial value problems in ordinary differential equations. In: Proceedings of the Conference on Numerical Analysis, Dublin 1974. London-New York: Academic Press 1975

Prof. Dr. M.N. Spijker
Mathematisch Instituut
Rijksuniversiteit Leiden
Wassenaarseweg 80
Leiden
Nederland