EQUIVALENCE THEOREMS FOR NONLINEAR FINITE-DIFFERENCE METHODS

by

M.N. Spijker

1. INTRODUCTION

Assume U is a solution of a given nonlinear (integro-)differential equation and assume u to be a corresponding solution of a finite-difference equation approximating the original infinitesimal problem. Let the finite-difference equation and its solution u depend on the so-called mesh-width $h > 0$. In this paper we deal with the following general problem: What conditions on the finite-difference equation are necessary and at the same time sufficient for the convergence of u to the true solution U as $h \to 0$?

For linear and semilinear differential equations there exists a series of equivalence theorems in which such necessary and sufficient conditions for convergence are given (see [2], [6], [7], [12], [13], [16]). For differential equations which are not semilinear there seem to exist no such equivalence theorems with any general applicability. Still for arbitrary differential equations which need not to be semilinear Stetter [15] and Ansorge [1] obtained a general condition that is sufficient for convergence - see also Kinnebrock [11]. Their sufficient condition essentially is that the linearization of the finite-difference scheme at the true solution U be stable. Their results only apply in full if the order of accuracy of the finite-difference scheme is high enough. Furthermore, if the linearization of the finite-difference scheme at U is only stable in the sense of Forsythe and Wasow (cf. [12, p. 95]) their results do not apply either.

In this paper an attempt is made to get rid of the limitations of the theories mentioned above. We shall present equivalence theorems giving conditions which are both necessary and sufficient for convergence. These theorems can be applied to differential equations that are not semilinear and furthermore they can be used in a number of cases where the theories of [1], [11], [15] mentioned above do not apply.

In this article we shall proceed in an abstract setting in order to make it possible to apply our equivalence theorems to different kinds of (integro-)differential equations (e.g. of hyperbolic, parabolic or elliptic type). The framework in which we proceed is similar to the one used by Stetter [15].

Part of the proof of the equivalence theorems will essentially be based on an idea used by Strang in [17] for proving convergence under conditions that are not subject to the limitations mentioned above in the case of nonlinear hyperbolic differential equations. As a consequence of this method of proof we are naturally led to a condition under which not only the global discretization error $u - U$ tends to zero as $h \to 0$ but also $u - U$ admits an expansion in powers of h. In this way we obtain the same result as derived in [14] but under weaker conditions than those stated there.

In chapter 2 we have collected a few results from functional analysis which are needed in the following chapters. In chapter 3 we derive the equivalence theorems as well as a number of closely related theorems. In chapter 4 we discuss the application of the abstract considerations of chapter 3 to a nonlinear parabolic differential equation and we display a numerical example in which repeated (Richardson) extrapolation to $h = 0$ is performed.

2. SOME RESULTS FROM FUNCTIONAL ANALYSIS

In this chapter we review some simple results from functional analysis. The proofs of the following lemma's have been included for completeness and because the Lipschitz condition being imposed here on the (n-th order) derivative of the operator Φ is slightly weaker than required in most textbooks on functional analysis.

In the following lemma a version of Taylor's formula in normed vector spaces is presented. The proof will be given along the lines of Collatz [4, p. 223].

LEMMA 1. Let R and S be real normed vector spaces and let E be an open subset of R. Let Φ be a mapping from E into S. Assume that the points x_o, x_1 and the segment $x_o + t\cdot(x_1-x_o)$ $(0 < t < 1)$ belong to E. Assume that the n-th order Fréchet derivative of Φ exists in all points of E and that it satisfies the Lipschitz condition

$$(2.1) \qquad || \Phi^{(n)}(x) - \Phi^{(n)}(x_o) || \leq \mu\cdot||x-x_o||$$

for all x of the form $x = x_o + t\cdot(x_1-x_o)$ with $0 < t < 1$. Then

$$(2.2) \qquad \begin{cases} \Phi(x_1) = \Phi(x_o) + \Phi'(x_o)(x_1-x_o) + \dots + \frac{1}{n!}\Phi^{(n)}(x_o)(x_1-x_o)^n + e \quad \text{where} \\ ||e|| \leq \frac{\mu}{(n+1)!}||x_1-x_o||^{n+1} \end{cases}$$

(for $n = 1,2,3,\dots$).

Proof. There exists a real-valued linear continuous operator L with domain S such that $||L|| = 1$, $L(e) = ||e||$ (see [10, p. 142]). We define $f(t) = L\,\Phi\,[x_o + t\cdot(x_1-x_o)]$ for $0 \leq t \leq 1$. f is continuous on $[0,1]$ and the derivatives $f^{(j)}$ exist and satisfy $f^{(j)}(t) = L\,\Phi^{(j)}[x_o+t(x_1-x_o)]\,(x_1-x_o)^j$ $(j = 1,2,\dots,n;\ 0 \leq t \leq 1)$. In view of (2.1) we have

$$(2.3) \qquad |f^{(n)}(t) - f^{(n)}(0)| \leq \mu\cdot t\cdot||x_1-x_o||^{n+1} \quad (0 < t < 1).$$

From the first line of (2.2) defining the vector e we have $||e|| = L(e) = f(1) - f(0) - f'(0) - \dots - \frac{1}{n!}\cdot f^{(n)}(0)$. Hence the function g defined by $g(t) = f(t) - f(0) - t\cdot f'(0) - \dots - \frac{t^n}{n!}\cdot f^{(n)}(0) - t^{n+1}\cdot||e||$ vanishes at $t = 1$. Since also $g(0) = 0$ we have that $g'(t_1) = 0$ for some $t_1 \in (0,1)$. Since $g'(0) = 0$ it follows that $g''(t_2) = 0$ for some $t_2 \in (0, t_1)$. Continuing in this way it follows easily that $g^{(n)}(t_n) = 0$ for some $t_n \in (0,1)$. Hence $f^{(n)}(t_n) - f^{(n)}(0) - (n+1)!\,t_n\cdot||e|| = 0$ and by using (2.3) with $t = t_n$ the inequality in (2.2) follows easily and the lemma has thus been proved.

The following lemma consists in a quantitative formulation of the inverse function theorem of function space theory (see [10, p. 687], [15]). We also note the strong relationship of this lemma to the well-known Newton-Kantorovich theorem (see [10, p. 708]).

LEMMA 2. Let R and S be real normed vector spaces and let R be complete (i.e. R is a Banach space). Let Φ be a mapping from the open subset $E \subset R$ into S. Let $x_o \in E$ and let $\Phi x_o = y_o$. Assume that the Fréchet derivative $\Phi'(x)$ exists on E and that it satisfies the following two conditions

(2.4) The inverse $\Phi'(x_o)^{-1}$ mapping S onto R exists and is bounded, $||\Phi'(x_o)^{-1}|| \leq \delta$,

(2.5) There is a constant $\mu > 0$ such that the open ball with center x_o and radius $1/\mu\delta$ belongs to E and $||\Phi'(x) - \Phi'(x_o)|| \leq \mu \cdot ||x-x_o||$ for all x in this ball.

Then for all $y \in S$ with $||y-y_o|| < 1/2\delta^2\mu$ there is a unique $x = x(y) \in R$ with $||x-x_o|| < 1/\delta\mu$ and $\Phi x = y$. Furthermore this $x(y)$ satisfies

(2.6) $$||x(y) - x_o|| \leq \frac{1}{\delta\mu}\cdot(1 - \sqrt{1-2\delta^2\mu||y-y_o||}) \leq 2\delta\cdot||y-y_o|| .$$

Proof. Choose any $y \in S$ with $||y-y_o|| < 1/2\delta^2\mu$ and define the operator G by $Gx = x - \Phi'(x_o)^{-1}[\Phi(x)-y]$. Then fixed points of G are solutions of the equation $\Phi x = y$. The Fréchet derivative of G exists, $G'(x) = \Phi'(x_o)^{-1}[\Phi'(x_o) - \Phi'(x)]$ and by (2.4), (2.5)

(2.7) $$||G'(x)|| \leq \delta\mu\cdot||x-x_o|| \quad (\text{for } ||x-x_o|| < 1/\delta\mu) .$$

Using (2.5) and lemma 1 with $n = 1$, $x_1 = x$ it follows that for $||x-x_o|| < 1/\delta\mu$ we have

$$\begin{aligned} ||Gx-x_o|| &= ||x-x_o - \Phi'(x_o)^{-1}[\Phi(x_o) + \Phi'(x_o)(x-x_o) + e - y]|| \\ &= ||\Phi'(x_o)^{-1}[y_o + e - y]|| \\ &\leq ||\Phi'(x_o)^{-1}|| \cdot \{||y-y_o|| + ||e||\} \\ &\leq \delta||y-y_o|| + \frac{\delta\mu}{2}\cdot||x-x_o||^2 . \end{aligned}$$

Furthermore it is easily verified that a number r satisfies the inequality $\delta||y-y_o|| + \frac{\delta\mu}{2}\cdot r^2 \leq r$ whenever $r_o \leq r \leq r_1$ where

$$r_o = \frac{1}{\delta\mu}\cdot(1 - \sqrt{1-2\delta^2\mu||y-y_o||}), \quad r_1 = \frac{1}{\delta\mu}\cdot(1 + \sqrt{1-2\delta^2\mu||y-y_o||}) .$$

Consequently, if we choose any r with $r_o \leq r < 1/\delta\mu$ and define D to be the closed ball with center x_o and radius r then G is a mapping from D into D. By (2.7) G is a contracting mapping from D into D.

Hence there is a unique x with $x \in D$, $Gx = x$ (see e.g. [10, p. 627]). By letting $r \uparrow 1/\delta\mu$ it follows that this x is also unique in the open ball with center x_o and radius $1/\delta\mu$ while by taking $r = r_o$ it follows that this x satisfies $||x-x_o|| \leq r_o$. Since the fixed points of G are just the solutions of the equation $\Phi x = y$ the lemma has thus been proved.

The following lemma, which is a partial converse of the preceding lemma shows that a condition of type (2.4) is indispensable for the result of lemma 2 to be valid.

LEMMA 3. Let R and S be real normed vector spaces and let Φ be a mapping from the open subset $E \subset R$ into S. Let $x_o \in E$ and $\Phi x_o = y_o$. Assume that for all $x \in E$ the Fréchet derivative $\Phi'(x)$ exists, that $\Phi'(x_o)$ is a one-to-one mapping from R onto S and that

$$(2.8) \qquad ||\Phi'(x) - \Phi'(x_o)|| \leq \mu \cdot ||x-x_o||$$

provided $||x-x_o||$ is small enough. Furthermore assume that there exist constants $\delta, \sigma, \tau > 0$ such that

1. for all $y \in S$ with $||y-y_o|| < \tau$ there is a unique $x = x(y) \in R$ with $||x-x_o|| < \sigma$ and $\Phi x = y$, and
2. This $x(y)$ satisfies $||x(y) - x_o|| \leq \delta \cdot ||y-y_o||$.

Then the inverse $\Phi'(x_o)^{-1}$ is bounded and $||\Phi'(x_o)^{-1}|| \leq \delta$.

Proof. Choose an arbitrary vector $z \in S$ with $||z|| = 1$ and define $v \in R$ by

$$\Phi'(x_o)\, v = z .$$

We shall show that $||v|| \leq \delta$, from which the lemma immediately follows.

Choose a constant $r > 0$ such that the following conditions (a) - (d) are fulfilled:

(a) the open ball with center x_o and radius r is contained in E,

(b) $||\Phi'(x) - \Phi'(x_o)|| \leq \mu \cdot ||x-x_o||$ for $||x-x_o|| < r$,

(c) $r < \sigma$,

(d) $\qquad ||\Phi(x) - y_o|| < \tau$ if $||x-x_o|| < r$.

Let t be a real parameter and define x by the relation $\Phi'(x_o)(x-x_o) = t\cdot z$. Hence $tv = x-x_o$. Now we take $0 < t < r/||v||$ so as to ensure that $||x-x_o|| < r$. We define y by $\Phi x = y$. Since (by (d)) $||y-y_o|| < \tau$ and $\Phi x = y$ we have by the assumption of the lemma that $||x-x_o|| \le \delta\cdot||y-y_o||$. Hence by (b) and lemma 1 it follows that

$$\begin{aligned} ||x-x_o|| &\le \delta\cdot\{||\Phi'(x_o)(x-x_o) + e||\} \\ &\le \delta\cdot\{||tz|| + ||e||\} \\ &\le \delta\cdot\{t + \tfrac{\mu}{2}\cdot||x-x_o||^2\} \\ &= t\delta\cdot\{1 + \tfrac{\mu}{2}\cdot t\cdot||v||^2\} . \end{aligned}$$

Since $||x-x_o|| = t\cdot||v||$ we have $||v|| \le \delta\cdot\{1 + t\,\frac{\mu}{2}\cdot||v||^2\}$. By letting $t \downarrow 0$ it follows that $||v|| \le \delta$ and the lemma has thus been proved.

3. A THEORY FOR NONLINEAR FINITE-DIFFERENCE EQUATIONS

In section 3.1 of this chapter we shall introduce the notations and definitions by means of which the equivalence theorems will be formulated. In section 3.2 we shall derive conditions which are necessary and sufficient for stability of a finite-difference scheme and in section 3.3 we shall state necessary and sufficient conditions for consistency. In section 3.4 we shall finally prove two equivalence theorems the last of which implies that a finite-difference method is convergent if and only if it is both stable and consistent (see theorem 5).

3.1 BASIC DEFINITIONS

3.1.1. The finite-difference equation

Assume A and B are real normed vector spaces. Let F be a given mapping with an open domain $D \subset A$ and with range in B . Let the vector $U \in D$ satisfy the equation

(3.1) $\qquad F(U) = 0$.

In the applications (3.1) will stand for a given differential or integro-differential equation with solution U . Initial or boundary conditions supplementing the differential equation are assumed incorporated in the mapping F . For an example we refer to chapter 4.

Let H be a subset of the real interval $(0, h_o]$ with $\inf H = 0$ and let A_h and B_h denote real normed vector spaces depending on the so-called mesh-width parameter $h \in H$. There are given families of operators Δ_h and L_h mapping A into A_h and B into B_h , respectively. It is assumed that the operators Δ_h and L_h are linear and bounded uniformly for $h \in H$. The space A_h is assumed to be complete for all $h \in H$.

We consider the approximation of U by solving instead of (3.1) the equation

(3.2) $\Phi_h(u) = 0$

where Φ_h is a mapping from an open set $D_h \subset A_h$ into B_h . We assume that Φ_h is Fréchet differentiable on D_h for each $h \in H$. Furthermore it is assumed that Δ_h maps the set D into D_h .

In the applications (3.2) will stand for a finite-difference equation with discrete initial or boundary conditions. The solution $u = u_h$ of (3.2) depends on h and will denote an approximation to $\Delta_h U$. For an example see chapter 4.

$$\begin{array}{ccccc} A & \supset & D & \xrightarrow{F} & B \\ \downarrow & \Delta_h & \downarrow & & \downarrow L_h \\ A_h & \supset & D_h & \xrightarrow{\Phi_h} & B_h \end{array}$$

3.1.2. Convergence

For the difference $u_h - \Delta_h U$ where $u = u_h$ and U satisfy (3.2) and (3.1), respectively the term *global* or ***accumulated discretization error*** **is used.** The finite-difference method (3.2) (or briefly: the operator Φ_h) is called *convergent* if for h sufficiently small a solution $u = u_h$ to (3.2) exists with $\lim_{h\to 0} ||u_h - \Delta_h U|| = 0$. In actual numerical applications it is not only desirable that

Φ_h is convergent according to this definition but it is also highly desirable that it satisfies the following two requirements:

1. Perturbations w in the finite-difference scheme should not destroy the convergence, at least if $||w||$ is of some order $O(h^\delta)$ - this is desirable since perturbations like round-off errors cannot be avoided in actual computations,

2. The global discretization error should admit an expansion in powers of h - this is desirable since such an expansion makes Richardson extrapolation possible.

In order to formulate a concept of convergence in which these two requirements are included we give the following definition.

DEFINITION 1. Let δ and q be real numbers > 0 and let $z_h \in A_h$ for each $h \in H$. Then Φ_h is said to be ***convergent** with an accuracy of order* q *at* z_h *under perturbations of order* δ if the following statement is true: for each family of vectors $w_h \in B_h$ depending on h in such a way that $||w_h|| = O(h^\delta)$ (for $h \to 0$), there exists a number $h_1 > 0$ with the property that for all $h \in H$ with $h \leq h_1$ there is a $v_h \in D_h$ with

(3.3) $\quad \Phi_h v_h = w_h ,$ and

(3.4) $\quad ||v_h - z_h|| = O(h^q)$ (for $h \in H$, $h \leq h_1$) .

We illustrate this definition by choosing for q an integer ≥ 1 and by defining

(3.5) $$z_h = \Delta_h \{U + \sum_{i=1}^{q-1} h^i U_i\}$$

where the U_i denote fixed elements of A. Suppose Φ_h is convergent with an accuracy of order q at this z_h under perturbations of order δ. It then follows from (3.4) that a solution v_h (see (3.3)) of the finite-difference equation in presence of a perturbation $w = w_h$ with $||w_h|| = O(h^\delta)$ satisfies $\lim_{h\to 0} ||v_h - \Delta_h U|| = 0$. Hence the requirement 1 stated above is fulfilled. Further, by choosing $w_h = 0$ we have $v_h = u_h$ and consequently

$$u_h - \Delta_h U = \sum_{i=1}^{q-1} h^i \Delta_h U_i + O(h^q)$$

which is the content of requirement 2.

3.1.3. Consistency

Throughout this chapter 3 r denotes a fixed integer ≥ 1 and $F_o, F_1, \ldots, F_{r-1}$ denote given operators from D into B. We assume that the operator Φ_h applied to elements $\Delta_h X$ with $X \in D$ can be expanded in powers of h in the following way:

(3.6) $$\Phi_h \Delta_h X = \sum_{j=0}^{r-1} h^j L_h F_j(X) + O(h^r) \quad (\text{for } h \in H)$$

DEFINITION 2. Let p be an integer with $1 \leq p \leq r$. Then the operator Φ_h is said to be *consistent of order p* if

$$F_o(U) = F_1(U) = \ldots. = F_{p-1}(U) = 0 .$$

For the element $\Phi_h(\Delta_h U)$ the term *local discretization error* is used. From (3.6) and definition 2 it follows that the local discretization error is $O(h^p)$ if Φ_h is consistent of order p. Hence the higher the order of consistency p, the smaller is the difference between $\Phi_h(u)$ and $\Phi_h(\Delta_h U)$ as $h \to 0$ (see (3.2)). Further it is clear from (3.1) that Φ_h is consistent of an order ≥ 1 if $F_o = F$.

3.1.4. Stability

In most current definitions of the concept of stability of finite-difference operators Φ_h it is required that the difference between solutions v and $\tilde{v}$ of the finite-difference equation obtained in presence of two different perturbations w and $\tilde{w}$, respectively can be estimated by an inequality of the form

(3.7) $$||v-\tilde{v}|| \leq \gamma\, h^{-\alpha} \cdot ||w-\tilde{w}||$$

where $\gamma > 0$ and $\alpha \geq 0$ are independent of h. The case $\alpha = 0$ corresponds to stability in the sense of Lax and Richtmyer [12, p. 45], while $\alpha > 0$ corresponds to stability in the sense of Forsythe and Wasow [12, p. 95]. For nonlinear problems stability is a local property of Φ_h and in general (3.7) does not hold *for all* w and $\tilde{w} \in B_h$. The following definition is similar to the one used by Stetter [15].

DEFINITION 3. Let m and α be real numbers ≥ 0 and let $z_h \in D_h$ (for $h \in H$). Then Φ_h is said to be *m-restricted stable of order* α *at* z_h if there exist constants $\beta, \gamma, h_1 > 0$ such that:

1. if $h \leq h_1$, $w \in B_h$ and $||w - \Phi_h z_h|| < \beta h^m$ then there is a unique $v = v(w) \in D_h$ with $\Phi_h v = w$ and $||v - z_h|| < \gamma\beta h^{m-\alpha}$, and
2. $v = v(w)$ satisfies $||v - z_h|| \leq \gamma h^{-\alpha} \cdot ||\Phi_h v - \Phi_h z_h||$.

We illustrate this definition by choosing $z_h = \Delta_h U$ (see (3.1)). Assume Φ_h is m-restricted stable of order α at this z_h and is consistent of an order p satisfying

$$p > m, \quad p > \alpha . \tag{3.8}$$

By choosing $w = 0$ in definition 3 it follows easily that $||w - \Phi_h z_h|| = \mathcal{O}(h^p) < \beta h^m$ for h sufficiently small. Hence there exists an element $u = v(0) \in D_h$ with $\Phi_h u = 0$ and $||u - \Delta_h U|| \leq \gamma h^{-\alpha} \cdot \mathcal{O}(h^p) = \mathcal{O}(h^{p-\alpha})$. Consequently Φ_h is convergent. Similarly it can be shown that Φ_h is convergent with an accuracy of order $p - \alpha$ at $\Delta_h U$ under perturbations of order p.

We shall call Φ_h stable of order α at z_h if it is 0-restricted stable of order α at z_h. Φ_h will be said to be stable of order α if it is stable of order α at each family z_h with z_h in the domain of definition of Φ_h (for each $h \in H$).

3.2 STABILITY THEOREMS

THEOREM 1 *(Stability of linear operators)*. Let Ψ_h be a linear bounded operator from A_h into B_h (for each $h \in H$). Then Ψ_h is stable of order α at $z_h \in A_h$ if and only if:

1) Ψ_h is a one-to-one mapping of A_h onto B_h,

2) the inverse Ψ_h^{-1} is bounded and

3) $||\Psi_h^{-1}|| \leq \gamma h^{-\alpha}$ for some constant γ , provided $h \leq$ some h_1 .

Proof. Assume Ψ_h is stable of order α at z_h . Let $y \in B_h$ with $||y|| = 1$. Put $w = \Psi_h z_h + t \cdot y$ with $0 < t < \beta$ where β is the constant of definition 3. Hence for $h \in H$, $h \leq h_1$ there is a unique $v \in A_h$ with $\Psi_h v = w$, $||v-z_h|| < \gamma\beta h^{-\alpha}$. Writing $x = v - z_h$ it follows that there is a unique $x \in A_h$ with $\Psi_h x = ty$, $||x|| < \gamma\beta h^{-\alpha}$. By defining $\hat{x} = t^{-1} \cdot x$ it follows that there is a unique $\hat{x} \in A_h$ with $\Psi_h \hat{x} = y$, $||\hat{x}|| < \gamma\beta h^{-\alpha} \cdot t^{-1}$. Since t may tend to zero it is clear that $\hat{x}$ satisfying $\Psi_h \hat{x} = y$ is also unique in the whole of A_h . Hence Ψ_h is a one-to-one mapping of A_h onto B_h . Furthermore $||\Psi_h^{-1} y|| = ||\hat{x}|| = t^{-1} \cdot ||x|| = t^{-1} \cdot ||v-z_h|| \leq t^{-1} \cdot \gamma h^{-\alpha} \, ||ty|| = \gamma h^{-\alpha}$. Thus the conditions 1), 2), 3) of theorem 1 are fulfilled.

Conversely, assume 1), 2), 3) to be true. Then $||\Psi_h^{-1} w - \Psi_h^{-1}(\Psi_h z_h)|| \leq \gamma h^{-\alpha} \, ||w-\Psi_h z_h||$ from which it follows easily that Ψ_h is stable of order α at z_h .

THEOREM 2 *(Stability of nonlinear operators)* . Let $\rho > 0$ and $c \geq 0$ be given constants and let $z_h \in D_h$ (for $h \in H$) . Assume that for each $h \in H$ the open ball with center z_h and radius ρh^c is contained in D_h and that for x in this ball

(3.9) $$||\Phi_h'(x) - \Phi_h'(z_h)|| \leq \lambda h^{-c} \, ||x-z_h||$$

where the constant λ is independent of h and x . Assume that $\Phi_h'(z_h)$ is a one-to-one mapping from A_h onto B_h (for $h \in H$) . Then Φ_h is $(c+2\alpha)$ - restricted stable of order α at z_h if and only if the linear operator $\Psi_h = \Phi_h'(z_h)$ is stable of order α .

Proof. Assume $\Phi_h'(z_h)$ is stable of order α . Applying theorem 1 it follows that

$$||\Phi_h'(z_h)^{-1}|| \leq \gamma h^{-\alpha} \; (h \leq h_1, \; h \in H) .$$

We shall apply lemma 2 with $R = A_h$, $S = B_h$, $E = D_h$, $\Phi = \Phi_h$, $x_o = z_h$, $\delta = \gamma h^{-\alpha}$. Condition (2.5) is fulfilled whenever μ satisfies $\mu \geq \lambda h^{-c}$ and

$h^{\alpha}\ \gamma^{-1}\ \mu^{-1} = \delta^{-1}\ \mu^{-1} \leq \rho h^{c}$. Therefore we choose $\mu = M\ h^{-c}$ with $M = \max(\lambda,\ \gamma^{-1}\ \rho^{-1}\ h_o^{\alpha})$ and lemma 2 may be applied. It follows that for all $w \in B_h$ with $||w-\Phi_h(z_h)|| < (2\delta^2\mu)^{-1} = (2\gamma^2 M)^{-1}\ h^{2\alpha+c}$ there is a unique $v \in A_h$ with $||v-z_h|| < (\delta\mu)^{-1} = (\gamma M)^{-1}\cdot h^{\alpha+c}$ and $\Phi_h(v) = w$. Furthermore (see (2.6)) this v satisfies $||v-z_h|| \leq 2\delta\ ||w-\Phi_h(z_h)|| = 2\gamma h^{-\alpha}||\Phi_h(v) - \Phi_h(z_h)||$. Hence Φ_h is $(2\alpha+c)$ - restricted stable of order α at z_h .

Now assume Φ_h is $(2\alpha+c)$ - restricted stable of order α at z_h . Lemma 3 may be applied with $R = A_h$, $S = B_h$, $E = D_h$, $\Phi = \Phi_h$, $x_o = z_h$, $\mu = \lambda h^{-c}$, $\tau = \beta h^{m}$, $\sigma = \gamma\beta h^{m-\alpha}$, $\delta = \gamma h^{-\alpha}$, $m = 2\alpha+c$, $h \leq h_1$ (see definition 3) . It follows that $\Phi_h'(z_h)^{-1}$ is bounded and $||\Phi_h'(z_h)^{-1}|| \leq \delta = \gamma h^{-\alpha}$ (for $h \leq h_1$) . By theorem 1 $\Phi_h'(z_h)$ is stable of order α . The theorem has thus been proved.

We note that theorem 1 is related to a result in [5, p. 106] and that theorem 2 is an extension of a result in [15, p. 114].

3.3 SOLVING THE EQUATION $\Phi_h u = 0$ WITH AN ERROR $= O(h^r)$

3.3.1. Preliminaries

In this section 3.3 we assume that the operators F_j appearing in (3.6) satisfy the following three conditions.

Condition I. For each closed ball C of finite radius contained in D and for $h \in H$ the relation (3.6) holds uniformly for X in C .

Condition II. The $(r-j)$-th order Fréchet derivative $F_j^{(r-j)}(X)$ exists for $X \in D$ and is continuous on D (for $j = 0,1,\ldots,r-1$) . Furthermore there is a unique $V_o \in D$ with $F_o(V_o) = 0$ and if $r \geq 2$ the derivative $F_o'(V_o)$ is a one-to-one mapping from A into B .

Condition III. If $r \geq 2$ the operator $F_o'(V_o)$ is a mapping onto B .

We note that from condition II and definition 2 it follows that Φ_h is consistent of order $p \geq 1$ if and only if $V_o = U$.

We define the function G mapping $D \times [-h_o,h_o]$ into B by

$G(X, h) = \sum_{j=0}^{r-1} h^j F_j(X)$. Now formula (3.6) can be written as

(3.10) $\Phi_h \Delta_h X = L_h G(X, h) + \mathcal{O}(h^r)$.

From the conditions II and III we have $G(V_o, 0) = 0$, and $\frac{\partial}{\partial X} G(V_o, 0) = F_o'(V_o)$ is a one-to-one mapping from A onto B (if $r \geq 2$) . In view of the implicit function theorem (see [10, p. 687]) we expect the existence of a function $X(h)$ with $G(X(h), h) = 0$, $X(0) = V_o$. Since the functions F_j are smooth we also expect a truncated Taylor series $\sum_{i=0}^{r-1} h^i V_i$ of $X(h)$ to satisfy $G(\sum_{i=0}^{r-1} h^i V_i, h) = \mathcal{O}(h^r)$. Since Δ_h maps D into D_h and in view of (3.10) the function

(3.11) $$V(h) = \sum_{i=0}^{r-1} h^i V_i$$

is thus expected to satisfy condition (3.12):

(3.12) $\begin{cases} \text{For sufficiently small } h \in H \text{ the element } \Delta_h V(h) \text{ belongs to } D_h \text{ and} \\ \text{satisfies } \Phi_h \Delta_h V(h) = \mathcal{O}(h^r) . \end{cases}$

The reason for considering a $V(h)$ satisfying (3.12) is that when $V_o = U$ it is more appropriate to prove convergence of Φ_h by comparing u_h with $\Delta_h V(h)$ rather than with $\Delta_h U$. The advantage is that for $\Delta_h V(h)$ the perturbation to the difference equation $\Phi_h u_h = 0$ can be made arbitrarily small by taking r sufficiently large. The need for this arises when one wants to treat the case where (3.8) is violated with $m = \rho + 2\alpha$ (see theorem 2). We note that the idea of comparing u_h with $\Delta_h V(h)$ was used with succes by Strang [16], [17].

In the following we shall prove (3.11), (3.12) rigorously and show that Φ_h is consistent of order p if and only if (3.15) holds.

3.3.2. *Recurrence relations for* $V_1, V_2, \ldots, V_{r-1}$

We put $V = \sum_{i=0}^{r-1} h^i V_i$ where the elements $V_1, V_2, \ldots, V_{r-1}$ in A are still to be determined. We choose h_1 with $0 < h_1 \leq h_o$ so small that the closed ball with center V_o and radius $\sigma = \sum_{i=1}^{r-1} h_1^i \|V_i\|$ is contained in D . For

$0 < h \leq h_1$ we have $V \in D$ and $G(V, h) = \sum_{j=0}^{r-1} h^j F_j(V_o + \sum_{i=1}^{r-1} h^i V_i) =$

$$\sum_{j=0}^{r-1} h^j \left\{ F_j(V_o) + \frac{1}{1!} F_j^{(1)}(V_o)(\sum_{i=1}^{r-1} h^i V_i) + \ldots + \frac{1}{(r-1-j)!} F_j^{(r-1-j)}(V_o)(\sum_{i=1}^{r-1} h^i V_i)^{r-1-j} \right\}$$

$+ W$. An application of lemma 1 with $R = A$, $S = B$, $E = D$, $x_o = V_o$, $x_1 = V$, $n = r-1-j$, $\mu = \sup ||F_j^{(r-j)}(Z)|| < \infty$ where the supremum of the continuous function $||F_j^{(r-j)}(Z)||$ is over all Z of the form $Z = V_o + t \cdot \sum_{i=1}^{r-1} h^i V_i$ $(0 \leq t \leq 1,\ 0 \leq h \leq h_1)$ (see condition II and [10, p. 660]), yields the result that $||W|| = \sum_{j=0}^{r-1} h^j \cdot O(h^{r-j}) = O(h^r)$. Since $F_o(V_o) = 0$ (see condition II) we thus have

$$G(V, h) = \sum_{i=1}^{r-1} h^i W_i + O(h^r) \tag{3.13}$$

where the $W_i \in B$ are defined by

$$W_i = \sum \frac{1}{M!} F_j^{(M)}(V_o)\, V_{i_1} V_{i_2} \cdots V_{i_M}$$

the summation being for all integers M, j, i_m with $0 \leq M$, $0 \leq j$, $1 \leq i_m$ $(1 \leq m \leq M)$, $j + i_1 + i_2 + \cdots + i_M = i$.

The largest index i_m appearing in this summation is $i_m = i$ and this value is obtained when $j = 0$, $M = 1$, $i_1 = i$. Further the only term in the sum that is independent of the V_{i_m} is $F_i^{(o)}(V_o) = F_i(V_o)$. Hence

$$W_i = F_o'(V_o) V_i + F_i(V_o) + Y_i$$

where Y_i is an expression which only contains V_k with $k < i$ and which vanishes when all V_k (with $1 \leq k < i$) equal zero. It follows from (3.13) that $G(V(h), h) = O(h^r)$ $(0 < h \leq h_1)$ if $V(h)$ is defined by (3.11) and the V_i satisfy

$$F_o'(V_o) V_i + F_i(V_o) + Y_i = 0 \quad (i = 1,2,\ldots,r-1) . \tag{3.14}$$

From the conditions II, III it follows that (3.14) is a series of recurrence relations defining the V_i uniquely. Using the relations (3.14) we shall prove:

THEOREM 3. Let the conditions I, II, III stated above be fulfilled. Then there is an element V_o in D and a series of elements $V_1, V_2, \ldots, V_{r-1}$ in A with the following two properties:

1. If $V(h)$ is defined by (3.11) then (3.12) holds,
2. Let p be an integer with $1 \leq p \leq r$. Then Φ_h is consistent of order p if and only if

$$V_o = U,\ V_1 = V_2 = \cdots = V_{p-1} = 0 . \tag{3.15}$$

Proof. We choose for V_o the element defined by condition II and for V_i $(1 \leq i \leq r-1)$ the elements defined by (3.14).

1. For $h \leq h_1$, $h \in H$ we have $G(V(h), h) = O(h^r)$ and $||V(h) - V_o|| \leq \sigma$ where σ is defined above. In view of condition I the relation (3.10) holds uniformly for X in the closed ball with center V_o and radius σ Hence (3.12) is fulfilled.

2. By examining the relations (3.14) it follows in view of condition II that $V_1 = \ldots = V_{p-1} = 0$ if and only if $F_1(V_o) = \ldots = F_{p-1}(V_o) = 0$. Hence $F_j(U) = 0$ $(0 \leq j \leq p-1)$ if and only if $U = V_o$, $V_1 = \ldots = V_{p-1} = 0$. The theorem has thus been proved.

Remark. In the above condition III has only been used to prove that there exist elements $V_1, V_2, \ldots, V_{r-1}$ satisfying (3.14). Therefore theorem 3 remains valid if condition III is replaced by the following weaker requirement III^*:

Condition III^*. If $r \geq 2$ there exist elements $V_1, V_2, \ldots, V_{r-1}$ in A satisfying (3.14).

Furthermore the uniformity of (3.6) on each closed ball in D has only been used to prove that (3.10) holds with X replaced by $V(h)$. Consequently the result of theorem 3 still holds if the conditions I and III are replaced simultaneously by I^* and III^* where condition I^* is as follows:

Condition I^*. $V(h) = \sum_{i=0}^{r-1} h^i V_i$ (where V_i are the elements in A from the conditions II and III^*) satisfies

$$\Phi_h \Delta_h V(h) = \sum_{j=0}^{r-1} h^j \cdot L_h F_j(V(h)) + O(h^r) \text{ (for } h \to 0)$$

3.4 EQUIVALENCE THEOREMS

Throughout this section 3.4 we assume the following condition to be fulfilled:

Condition IV. V_o is an element in D and $V_1, V_2, \ldots, V_{r-1}$ are elements in A with the properties 1 and 2 stated in theorem 3. Furthermore $H \subset (0, h_o]$

where h_o is so small that the closed ball with center V_o and radius $\sigma = \sum_{i=1}^{r-1} h_o^i \, ||V_i||$ is contained in D .

Since Δ_h maps D into D_h there follows from condition IV that the element $\Delta_h \{\sum_{i=0}^{q-1} h^i V_i\}$ belongs to D_h whenever $h \in H$ and q is an integer $\leq r$. Throughout this section we write

$$V(h) = \sum_{i=0}^{r-1} h^i V_i$$

and we assume

Condition V. There are real numbers $\rho > 0$, $\lambda \geq 0$, $c \geq 0$ such that the open ball with center $\Delta_h V(h)$ and radius ρh^c is contained in D_h and such that for x in this ball

$$||\Phi_h'(x) - \Phi_h'(\Delta_h V(h))|| \leq \lambda h^{-c} ||x - \Delta_h V(h)|| \text{ , for } h \in H \text{ .}$$

Condition VI. The linear operator $\Phi_h' (\Delta_h V(h))$ has a bounded inverse mapping B_h onto A_h (for each $h \in H$) .

THEOREM 4 ***(Equivalence** of convergence and stability)* . Let the conditions IV, V and VI be fulfilled. Assume $q \geq 1$ is an integer and $\alpha \geq 0$ a real number with

$$c + 2\alpha < q + \alpha \leq r \text{ .} \tag{3.16}$$

Let $z_h \in D_h$ be defined by

$$z_h = \Delta_h \{\sum_{i=0}^{q-1} h^i V_i\} \text{ .} \tag{3.17}$$

Then Φ_h is convergent with an accuracy of order q at this z_h under perturbations of order $q + \alpha$ if and only if the linear operator $\Phi_h'(z_h)$ is stable of order α .

*Proof.*1. Assume $\Phi_h'(z_h)$ is stable of order α . By theorem 1 $\Phi_h'(z_h)$ has a bounded inverse and

$$||\Phi_h'(z_h)^{-1}|| \leq \gamma h^{-\alpha} \text{ .} \tag{3.18}$$

From the definition of $V(h)$ we have $||z_h - \Delta_h V(h)|| = \mathcal{O}(h^q)$. In view of

(3.16) we thus have $||z_h - \Delta_h V(h)|| < \rho h^c$ for h sufficiently small and by condition V:

(3.19) $$||\Phi_h'(z_h) - \Phi_h'(\Delta_h V(h))|| \le \lambda h^{-c} \cdot O(h^q) .$$

As $-c + q - \alpha > 0$ it follows from (3.18), (3.19) that

$$||\Phi_h'(z_h) - \Phi_h'(\Delta_h V(h))|| \cdot ||\Phi_h'(z_h)^{-1}|| = O(h^{-c+q-\alpha}) < 1$$

for h sufficiently small. By Banach's lemma $\Phi_h'(\Delta_h V(h))$ thus also has a bounded inverse and

$$||\Phi_h'(\Delta_h V(h))^{-1}|| \le \gamma_o \cdot h^{-\alpha}$$

for some fixed $\gamma_o > 0$ and h sufficiently small in H .

Applying theorem 2 with $z_h = \Delta_h V(h)$ there follows that Φ_h is $(c + 2\alpha)$ restricted stable of order α at $\Delta_h V(h)$. Consequently there are constants $\beta, \gamma_1, h_1 > 0$ such that for $h \le h_1$, $h \in H$ and all $w \in B_h$ with $||w - \Phi_h \Delta_h V(h)|| < \beta h^{c+2\alpha}$ there is a unique $v \in D_h$ with $\Phi_h v = w$, $||v - \Delta_h V(h)|| < \gamma_1 \beta h^{c+\alpha}$; moreover $||v - \Delta_h V(h)|| \le \gamma_1 h^{-\alpha} \cdot ||\Phi_h v - \Phi_h \Delta_h V(h)||$.

Now let $w_h \in B_h$, $||w_h|| = O(h^{q+\alpha})$ (for $h \to 0$) . Then $||w_h - \Phi_h \Delta_h V(h)|| \le ||w_h|| + ||\Phi_h \Delta_h V(h)|| = O(h^{q+\alpha}) + O(h^r)$. In view of (3.16) we thus have $||w_h - \Phi_h \Delta_h V(h)|| = o(h^{c+2\alpha})$. Hence for h sufficiently small $||w_h - \Phi_h \Delta_h V(h)|| < \beta h^{c+2\alpha}$ and there is a unique $v_h \in D_h$ with $\Phi_h v_h = w_h$, $||v_h - \Delta_h V(h)|| < \gamma_1 \cdot \beta h^{c+\alpha}$. Moreover $||v_h - \Delta_h V(h)|| \le \gamma_1 h^{-\alpha} ||\Phi_h v_h - \Phi_h \Delta_h V(h)|| \le \gamma_1 h^{-\alpha} \cdot \{O(h^{q+\alpha}) + O(h^r)\} = O(h^q)$. Consequently $||v_h - z_h|| \le ||v_h - \Delta_h V(h)|| + ||\Delta_h V(h) - z_h|| = O(h^q) + O(h^q)$. It follows that Φ_h is convergent with an accuracy of order q at z_h under perturbations of order $q + \alpha$.

2. Assume Φ_h is convergent with an accuracy of order q at z_h under perturbations of order $q + \alpha$. We shall prove that $\Phi_h'(z_h)$ is stable of order α .

We define $\Gamma_h = \Phi_h'(\Delta_h V(h))^{-1}$ for $h \in H$. We choose an element $y(h) \in B_h$ with $||y(h)|| = 1$ and $||\Gamma_h y(h)|| \ge \frac{1}{2} \cdot ||\Gamma_h||$ and we define

$$w_h = \Phi_h \Delta_h V(h) + h^{q+\alpha} y(h) .$$

By virtue of (3.12) (which holds in view of condition IV) we have

$||w_h|| = O(h^r) + O(h^{q+\alpha}) = O(h^{q+\alpha})$ (for $h \to 0$) . According to definition 1 for h sufficiently small there exist $v_h \in D_h$ with $\Phi_h v_h = w_h$, $||v_h - z_h|| \leq \gamma_o h^q$ where γ_o is independent of h . Hence $||v_h - \Delta_h V(h)|| \leq \gamma_o \cdot h^q + ||z_h - \Delta_h V(h)|| \leq \gamma_1 \cdot h^q$ where γ_1 is some constant independent of h . In view of condition V and lemma 1 we have for h sufficiently small that $h^{q+\alpha}\, y(h) = \Phi_h v_h - \Phi_h \Delta_h V(h) =$ $\Phi_h'(\Delta_h V(h))\,(v_h - \Delta_h V(h)) + e$ with

$$||e|| \leq \frac{1}{2}\lambda h^{-c} \cdot ||v_h - \Delta_h V(h)||^2 .$$

Since $\Gamma_h(h^{q+\alpha}\, y(h)) = (v_h - \Delta_h V(h)) + \Gamma_h(e)$ we get

$$\frac{1}{2} h^{q+\alpha} \cdot ||\Gamma_h|| \leq ||\Gamma_h\,(h^{q+\alpha}\, y(h))|| \leq ||v_h - \Delta_h V(h)|| + ||\Gamma_h|| \cdot ||e|| .$$

Consequently

$$h^{q+\alpha}\; ||\Gamma_h|| \leq 2\gamma_1 \cdot h^q + \lambda h^{-c} \cdot (\gamma_1 h^q)^2 \cdot ||\Gamma_h|| .$$

Dividing both members of this inequality by $h^{q+\alpha}$ and combining the terms in which $||\Gamma_h||$ enters we obtain

$$(1 - \lambda \gamma_1{}^2 \cdot h^{q-c-\alpha})\; ||\Gamma_h|| \leq 2\gamma_1\, h^{-\alpha} .$$

Since $q - c - \alpha > 0$ (see (3.16)) we have

$$||\Gamma_h|| \leq 4\, \gamma_1\, h^{-\alpha}$$

provided $h \in H$ is sufficiently small. By using Banach's lemma in a similar way as in part 1 of the proof it follows that $\Phi_h'(z_h)$ also has a bounded inverse and

$$||\Phi_h'(z_h)^{-1}|| \leq \gamma_2\, h^{-\alpha}$$

for some constant γ_2 and h sufficiently small. By theorem 1 $\Phi_h'(z_h)$ is stable of order α . The theorem has thus been proved.

We now are in a position to derive our main result which is formulated in the following equivalence theorem 5. In addition to the conditions IV, V, VI we assume here:

Condition VII. If $X \in A$ and $\lim\limits_{h\to 0} ||\Delta_h X|| = 0$ then $X = 0$.

Condition VIII. If v and $\tilde{v}$ satisfy $\Phi_h v = \Phi_h \tilde{v}$ for some $h \in H$ then

$v = \tilde{v}$.

THEOREM 5 *(Equivalence of convergence to stability and consistency)* . Let the conditions IV - VIII be fulfilled. Assume p and q are integers with $1 \leq p \leq q$ and $\alpha \geq 0$ is a real number with

(3.16) $\quad c + 2\alpha < q + \alpha \leq r$.

Then Φ_h is convergent with an accuracy of order q under perturbations of order $q + \alpha$ at some z_h of the form

(3.20) $\quad z_h = \Delta_h \{U + \sum_{i=p}^{q-1} h^i U_i\} \qquad (h \in H)$

with $U_p, U_{p+1}, \ldots, U_{q-1} \in A$, if and only if the linear operator $\Phi_h'(\Delta_h V(h))$ is stable of order α and simultaneously Φ_h is consistent of order p .

Proof. 1. Let $\Phi_h'(\Delta_h V(h))$ be stable of order α and let Φ_h be consistent of order p . By virtue of condition IV (see (3.15)) we have $V_o = U$, $V_1 = V_2 = \ldots = V_{p-1} = 0$. We define $U_i = V_i$ (for $p \leq i \leq q-1$) and we define z_h by (3.20). Hence this z_h is equal to the z_h defined by (3.17). Since $||z_h - \Delta_h V(h)|| = O(h^q)$ and $||\Phi_h'(\Delta_h V(h))^{-1}|| \leq \gamma h^{-\alpha}$ (for some constant γ and h sufficiently small) it follows by a similar application of Banach's lemma as in the first part of the proof of theorem 4 that $\Phi_h'(z_h)$ is stable of order α . From theorem 4 we thus obtain the result that Φ_h is convergent with an accuracy of order q at z_h defined by (3.20) under perturbations of order $q + \alpha$.

2. Let Φ_h be convergent with an accuracy of order q under perturbations of order $q + \alpha$ at some z_h of the form (3.20).

We define $w_h = \Phi_h \Delta_h V(h)$. By condition IV (see (3.12)) we have $||w_h|| = O(h^r) = O(h^{q+\alpha})$ $(h \to 0)$. Consequently $\Phi_h v_h = w_h$, $v_h = z_h + O(h^q)$ for some $v_h \in D_h$ and $h \leq$ some h_1, $h \in H$. Since $\Phi_h v_h = \Phi_h \Delta_h V(h)$ we have from condition VIII that $\Delta_h V(h) = v_h = z_h + O(h^q)$ $(h \leq h_1, h \in H)$. Hence

$$\Delta_h \{V_o + \sum_{i=1}^{r-1} h^i V_i\} = \Delta_h \{U + \sum_{i=p}^{q-1} h^i U_i\} + O(h^q) .$$

Combining the terms in which the same powers of h appear and by letting $h \to 0$ we get (cf. condition VII):

$\Delta_h (V_o - U) \to 0$, $\Delta_h V_i \to 0$ $(1 \leq i \leq p-1)$,

$\Delta_h (V_i - U_i) \to 0$ $(p \leq i \leq q-1)$ (for $h \to 0$) and

$V_o = U$, $V_i = 0$ $(1 \leq i \leq p-1)$, $V_i = U_i$ $(p \leq i \leq q-1)$.

Since (3.15) thus holds we have by condition IV that Φ_h is consistent of order p . Furthermore our z_h of the form (3.20) equals the z_h defined by (3.17). Applying theorem 4 it thus follows that the linear operator $\Phi_h'(z_h)$ is stable of order α . Applying Banach's lemma in exactly the same way as in the first part of the proof of theorem 4 it follows that the linear operator $\Phi_h' (\Delta_h V(h))$ is stable of order α . This completes the proof of theorem 5.

In the applications usually c, α and p are given numbers determined by the finite-difference scheme under consideration. In order to prove convergence of the finite-difference method one only has to choose q and r so large that $q > c + \alpha$ and $r \geq q + \alpha$ (which ensures (3.16) to hold) and to check for stability (of order α) of the linear operator $\Phi_h' (\Delta_h V(h))$. An application of theorem 5 then yields (among other things) the result that for h sufficiently small the finite-difference equation $\Phi_h(u) = 0$ has a solution $u = u_h$ satisfying $||u_h - \Delta_h U|| = O(h^p)$. We emphasize the fact that this holds without any restriction on p (like e.g. (3.8)).

4. AN APPLICATION TO PARABOLIC DIFFERENTIAL EQUATIONS

4.1 THE DIFFERENTIAL EQUATION

In this chapter it is assumed that U is a real function which is defined on the infinite strip

$$G = \{(s,t) \mid -\infty < s < \infty,\ 0 \leq t \leq T\}$$

and which solves the initial value problem

(4.1.a) $\quad U(s,0) - f(s) = 0, \ -\infty < s < \infty$,

(4.1.b) $\quad g\,[U_t(s,t), U_{ss}(s,t), U_s(s,t), U(s,t), s, t] = 0, \ (s,t) \in G$

the subscrips in (4.1.b) denoting partial differentiation.

In the following we shall use the notation $\partial_i\, g\,[z_1,z_2,z_3,z_4,s,t]$ to denote the first order partial derivative of g with respect to the variable z_i $(i=1,2,3,4)$. Furthermore we shall denote by $C^\infty(E)$ the class of all bounded real functions defined on E which have the property that each of their partial derivatives exists and is bounded on the set E .

Throughout this chapter it is assumed that f, g and U satisfy the following smoothness conditions:

(A) All partial derivatives of the function $g\,[z_1,z_2,z_3,z_4,s,t]$ exist for $-\infty < z_i < \infty$, $(s,t) \in G$; and for any value of α the function g and each of its partial derivatives remains bounded when z_i, s, t vary in such a way that $|z_i| \leq \alpha$ $(i=1,2,3,4)$, $(s,t) \in G$. Furthermore there are constants $\mu_1 > 0$, $\mu_2 > 0$ such that for $-\infty < z_i < \infty$, $(s,t) \in G$:

$$\partial_1\, g\,[z_1,z_2,z_3,z_4,s,t] \leq -\mu_1, \quad \partial_2\, g\,[z_1,z_2,z_3,z_4,s,t] \geq \mu_2 \ ,$$

(B) The function f belongs to class $C^\infty(-\infty, \infty)$,

(C) U belongs to class $C^\infty(G)$.

In order to use the concepts of chapter 3 we define the vector space A by $A = C^\infty(G)$. For $X \in A$ we define the norm

$$||X|| = \max_{0 \leq i \leq r+1} \sup_G \left|\frac{\partial^i}{\partial t^i} X(s,t)\right| + \max_{0 \leq i \leq 2r+2} \sup_G \left|\frac{\partial^i}{\partial s^i} X(s,t)\right|$$

r denoting an arbitrary but fixed integer ≥ 1 . The vector space B is defined by $B = \{Y \mid Y = (Y_o,Y_1)$ where $Y_o \in C^\infty(-\infty,\infty)$, $Y_1 \in C^\infty(G)\}$. For $Y = (Y_o,Y_1)$ in B we define the norm

$$||Y|| = \sup |Y_o(s)| + \sup |Y_1(s,t)|$$

the supremums being for $-\infty < s < \infty$ and $(s,t) \in G$, respectively. Finally we define $D = A$ and the operator F mapping D into B is defined by $F(X) = (Y_o,Y_1)$ where $Y_o(s) = X(s,0) - f(s)$, $Y_1(s,t) = g\,[X_t(s,t), X_{ss}(s,t), X_s(s,t), X(s,t), s,t]$. With these definitions (3.1) is equivalent to (4.1).

4.2 THE FINITE-DIFFERENCE EQUATION

In order to construct a finite-difference scheme by means of which the solution U to (4.1) can be approximated we choose increments $h = \Delta t > 0$, $\Delta s > 0$ of the variables t and s, respectively in such a way that the quotient $h/(\Delta s)^2$ is a constant independent of h. We define the interval $H = (0, h_o]$ where h_o is a constant $\leq T$. For $h \in H$ we define

$G(h) = \{(s,t) | (s,t) \in G \text{ and } s = m\Delta s,\ t = nh,\ m = 0, \pm 1, \pm 2, \ldots,\ n = 0,1,2,\ldots\}$,

$G_o(h) = \{ s \mid s = m\Delta s,\ m = 0, \pm 1, \pm 2, \ldots\}$,

$G_1(h) = \{(s,t) | (s,t) \in G(h) \text{ and } t+h \leq T\}$.

Using the shifting operator E defined by $E^i\, u(s,t) = u(s+i\Delta s,t)$ the finite-difference equation we shall deal with can be written in the following way:

(4.2.a) $\quad u(s,0) - f(s) = 0, \qquad s \in G_o(h)$,

(4.2.b) $\quad g\left[h^{-1}\cdot\{u(s,t+h) - u(s,t)\},\ (\Delta s)^{-2}\cdot \sum_i \alpha_i E^i u(s,t),\ (\Delta s)^{-1}\cdot \sum_i \beta_i E^i u(s,t),\right.$
$\quad\quad \left. u(s,t),\ s,\ t\right] = 0, \qquad (s,t) \in G_1(h)$.

In (4.2.b) α_i and β_i are constants independent of h satisfying the conditions (4.3) and vanishing for all but a finite number of values $i = 0, \pm 1, \pm 2, \ldots$.

$$(4.3) \quad \begin{cases} \alpha_i = \alpha_{-i}, \beta_i = -\beta_i , \\ \sum_i \alpha_i = 0,\ \sum_i \alpha_i i^2 = 2,\ \sum_i \beta_i i = 1 . \end{cases}$$

We note that since $\partial_1 g \leq -\mu_1 < 0$ (see condition (A)) $u(s,t+h)$ is uniquely determined by (4.2.b) once $u(s,t)(s \in G_o(h))$ is given. Consequently $u(s,t)(t=0,h,2h,\ldots \leq T)$ can be computed uniquely from (4.2) by an application of (4.2.a) and by applying (4.2.b) successively with $t = 0,h,2h,\ldots \leq T-h$.

In order to use the concepts of chapter 3 we define A_h to be the vector space consisting of all real bounded functions defined on $G(h)$ with the norm $||x|| = \sup |x(s,t)|$ for $x \in A_h$, the supremum being for $(s,t) \in G(h)$. B_h is the vector space consisting of all elements $y = (y_o,y_1)$ where y_o and y_1 are real bounded functions defined on $G_o(h)$ and $G_1(h)$, respectively with norm $||y|| = ||(y_o,y_1)|| = \sup |y_o(s)| + \sup |y_1(s,t)|$ the supremums being for $s \in G_o(h)$, $(s,t) \in G_1(h)$, respectively. For $X \in A$ we define $\Delta_h X = x \in A_h$ by setting $x(s,t) = X(s,t)$ (for all $(s,t) \in G(h)$) and for $(Y_o,Y_1) \in B$ we put $L_h(Y_o,Y_1) = (y_o,y_1) \in B_h$ with $y_o(s) = Y_o(s)$, $y_1(s,t) = Y_1(s,t)$ (for all $s \in G_o(h)$, $(s,t) \in G_1(h)$) .

To simplify the notations we introduce operators $C_j = C_j(h)$ defined by

$$(4.4)\quad \begin{cases} C_1x(s,t) = h^{-1}\{x(s,t+h) - x(s,t)\} , \\ C_2x(s,t) = (\Delta s)^{-2} \sum_i \alpha_i E^i x(s,t) , \\ C_3x(s,t) = (\Delta s)^{-1} \sum_i \beta_i E^i x(s,t) , \\ C_4x(s,t) = x(s,t) . \end{cases}$$

We define $D_h = A_h$ and the operator Φ_h mapping D_h into B_h is defined by $\Phi_h(x) = (y_o,y_1)$ where $y_o(s) = x(s,0) - f(s)$ and $y_1(s,t) = g\,[C_1x(s,t),\ C_2x(s,t),\ C_3x(s,t),\ C_4x(s,t),\ s,\ t]$. With these definitions (3.2) is equivalent to our finite-difference scheme (4.2).

Let $z \in D_h$. Then $\Phi_h'(z)$ exists and is given by $\{\Phi_h'(z)\}[x] = (y_o,y_1)$ where

$$(4.5)\quad \begin{aligned} y_o(s) &= x(s,0), \\ y_1(s,t) &= \sum_{i=1}^{4} \partial_i\, g\,[C_1z(s,t),\ C_2z(s,t),\ C_3z(s,t),\ C_4z(s,t),\ s,\ t]\cdot C_ix(s,t) \end{aligned}$$

It follows that all general conditions of section 3.1.1 are satisfied here.

4.3 THE CONDITIONS I - VIII

The purpose of this section is to show that with the definitions of the sections 4.1, 4.2 the conditions I - VIII of chapter 3 are fulfilled.

Condition I.

Let X belong to a given closed ball C contained in A . Then $||X|| \leq \beta$ where β is some constant depending only on C . We have $\Phi_h\Delta_hX = (y_o,y_1)$ with

$$\begin{aligned} y_o(s) &= X(s,0) - f(s) , \\ y_1(s,t) &= g\,[C_1X(s,t),\ C_2X(s,t),\ C_3X(s,t),\ C_4X(s,t),\ s,\ t] . \end{aligned}$$

Using (4.3) and Taylor's formula it follows that

$$(4.6)\qquad \Big| y_1(s,t) - g\,[X_t(s,t),\ X_{ss}(s,t),\ X_s(s,t),\ X(s,t),\ s,\ t] - \sum_{j=1}^{r-1} h^j \{f_j[X](s,t)\}\Big| \leq K \cdot h^r$$

for all $(s,t) \in G_1(h)$. The f_j are operators mapping A into itself determined by the partial derivatives of g . The constant K can be estimated in terms of bounds for the derivatives

$(\partial^i/\partial t^i)X$ $(2 \leq i \leq r+1)$, $(\partial^i/\partial s^i)X$ $(3 \leq i \leq 2r+2)$.

Hence there exists a constant K such that (4.6) holds uniformly for all X with $||X|| \leq \beta$ (see the definition of the norm in A). Defining

$$(4.7.a)\qquad F_o = F \ ,$$

and $F_j(X) = (0, f_j(X))$ $(1 \leq j \leq r-1)$ it follows that (3.6) holds uniformly for $X \in C$.

Conditions II, III.

Using the smoothness condition (A), the definition of the norm in A and the definition of f_j it follows that the operators F_j satisfy the differentiability requirements of condition II.

From (4.7.a) it follows that

$$(4.7.b)\qquad V_o = U$$

satisfies $F_o(V_o) = 0$. The assumption that $F_o(\tilde{V}_o) = 0$ for some $\tilde{V}_o \neq V_o$ can be shown to lead to a contradiction by deriving a linear homogeneous parabolic initial value problem for $V_o - \tilde{V}_o$ by subtracting $F_o(\tilde{V}_o)$ from $F_o(V_o)$. For the initial value problem obtained in this way has a unique solution (see [9, p. 171]), which shows that $V_o - \tilde{V}_o = 0$. Consequently there is a unique V_o with $F_o(V_o) = 0$.

We shall now show that $F_o'(V_o) = F'(U)$ is a one-to-one mapping from A onto B . The equality $F'(U)X = Y$ for $X \in A$, $Y = (Y_o, Y_1) \in B$ is equivalent to

$$(4.8.a)\qquad X(s,0) = Y_o(s)\ (-\infty < s < \infty),$$

$$(4.8.b)\qquad g_1(s,t)X_t(s,t) + g_2(s,t)\,X_{ss}(s,t) + g_3(s,t)\,X_s(s,t) + g_4(s,t)\,X(s,t) = Y_1(s,t)\ ((s,t) \in G)$$

where

(4.9) $\quad g_i(s,t) = \partial_i\, g\, [U_t(s,t),\, U_{ss}(s,t),\, U_s(s,t),\, U(s,t),\, s,\, t]$

(i=1,2,3,4) . Using the result in [9, p. 171], a straightforward extension of [9, p. 175] and the smoothness assumptions (A), (C) it can be proved that the initial value problem (4.8) has a unique solution $X \in A$ whenever $Y = (Y_o, Y_1)$ is a given element in B . This completes the proof of the conditions II, III.

Condition IV.

Since the conditions I, II, III are satisfied there follows from theorem 3 that condition IV can be fulfilled provided $H = (0, h_o]$ where h_o is small enough.

Condition V.

Let $z = \Delta_h V(h) = \Delta_h \sum_{i=0}^{r-1} h^i V_i$ where V_i are the elements in A occuring in condition IV. For $x \in A_h$ we have $||\Phi_h'(x) - \Phi_h'(z)|| = \sup ||\{\Phi_h'(x) - \Phi_h'(z)\}\, v\, ||$ where the supremum is for all $v \in A_h$ with $||v|| = 1$. From (4.5), (4.4) we thus have

(4.10) $\quad ||\Phi_h'(x) - \Phi_h'(z)|| \leq \mu h^{-1} \cdot \sup \sum_{i=1}^{4} |\partial_i \tilde{g} - \partial_i g|$

where the argument of $\partial_i \tilde{g}$ equals

$$[C_1 x(s,t),\, C_2 x(s,t),\, C_3 x(s,t),\, C_4(x,t),\, s,\, t]$$

and the argument of $\partial_i g$ is the same with x replaced by z . In (4.10) the supremum is over all $(s,t) \in G_1(h)$, μ is a constant depending only on α_i, β_i and the factor h^{-1} in (4.10) stems from the inverse powers of the mesh-width occuring in (4.4). By the mean value theorem we obtain from (4.10)

(4.11) $\quad ||\Phi_h'(x) - \Phi_h'(z)|| \leq (\mu h^{-1})^2 \cdot \sup \sum_{i=1}^{4} \sum_{j=1}^{4} |\partial_i \partial_j g|$

the first 4 arguments of $\partial_i \partial_j g$ being values between $C_k\, x(s,t)$ and $C_k\, z(s,t)$ (k=1,2,3,4) .

Let $x \in A_h$, $||x-z|| < h$, $h \in H$. Since z is composed of smooth functions $V_i \in A$ the first 4 arguments of $\partial_i \partial_j g$ in (4.11) will then remain bounded if h and (s,t) vary through H and $G_1(h)$, respectively. In view of condition (A) it thus follows that the supremum appearing in (4.11) is finite, uniformly for $h \in H$, $||x-z|| < h$. Consequently $||\Phi_h'(x) - \Phi_h'(z)|| \leq \lambda h^{-2}\, ||x-z||$ for some constant λ .

Hence condition V is fulfilled with $\rho = 1/h_o$ and

(4.12) $c = 2$.

Conditions VI, VII, VIII.

From (4.4) and condition (A) there follows that for any fixed $h \in H$, $z \in A_h$ and for arbitrary $y = (y_o, y_1) \in B_h$ there exists a unique $x \in A_h$ satisfying (4.5). Moreover $||x|| \leq \gamma \cdot ||y||$, γ being a parameter depending on h and z but independent of y . Consequently VI is fulfilled. It is easily verified that the conditions VII, VIII are also satisfied here.

4.4 A SUFFICIENT CONDITION FOR CONVERGENCE

In this section we prove the following theorem 6 by combining theorem 5 with a well known sufficient condition for stability of linear difference schemes due to John [9].

THEOREM 6. Let the conditions (A), (B), (C) hold. Let $u_h(s,t)$ be the solution of the finite-difference equation (4.2) where α_i, β_i are subject to (4.3). Let $K(s,t) = \{-hg_2(s,t)\}/\{(\Delta s)^2 g_1(s,t)\}$
(see (4.9)) and assume the function

$$\psi(\theta) = 1 + K(s,t) \cdot \{\alpha_o + 2 \sum_{j>0} \alpha_j \cos(j\theta)\}$$

satisfies the inequality

$$|\psi(\theta)| \leq \exp(-M\theta^2)$$

(whenever $-\infty < s < \infty$, $0 \leq t \leq T$, $|\theta| \leq \pi$), M denoting some constant > 0 .

Then there exists an infinite series of functions $U_1, U_2, U_3, \ldots$ each of which is in $C^\infty(G)$ such that for all positive integers q

(4.13) $$u_h(s,t) - U(s,t) = \sum_{i=1}^{q-1} h^i U_i(s,t) + O(h^q)$$

(for $h \to 0$) uniformly for $(s,t) \in G(h)$.

Proof. 1. With the definitions of the sections 4.1, 4.2 theorem 5 may be applied since the conditions IV - VIII have been shown to be fulfilled here. From (4.7.a) and definition 2 (section 3.1) it follows that Φ_h is consistent of order 1.

By choosing $\alpha = 0$, $r = q > 2$ in (3.16) we thus obtain from (4.12) the result that Φ_h is convergent with an accuracy of order q under perturbations of order q at some z_h of the form $z_h = \Delta_h \{U + \sum_{i=1}^{q-1} h^i U_i\}$ provided the linear operator Ψ_h defined by

$$\Psi_h = \Phi_h' (\Delta_h V(h))$$

is stable of order 0. It will be shown that this is the case under the condition on $\psi(\theta)$ stated in theorem 6. Hence (4.13) holds for each integer $q > 2$ (see definition 1, section 3.1). Consequently (4.13) also holds for each integer $q \geq 1$.

2. It remains to be shown that Ψ_h is stable of order 0. It follows from (4.5), (4.7.b), (4.9) that the equality $\Psi_h x = y$ for $x \in A_h$, $y = (y_o, y_1) \in B_h$ is equivalent to

$$x(s,0) = y_o(s), \; s \in G_o(h)$$
$$\sum_{i=1}^{4} \{g_i(s,t) + O(h)\} \cdot C_i x(s,t) = y_1(s,t), \; (s,t) \in G_1(h) .$$

Solving the last equality for $x(s,t+h)$ we obtain by an application of (4.4)

$$x(s,t+h) = x(s,t) + K(s,t) \cdot \sum_i \alpha_i E^i x(s,t) + (\Delta s) L(s,t) \cdot \sum_i \beta_i E^i x(s,t) + O(h||x||) + O(h||y||) \tag{4.14}$$

(for $t = 0, h, 2h, \ldots.$) where $L(s,t)$ is defined by the same formula as $K(s,t)$ with g_2 replaced by g_3. Using the fact that $\alpha_i = \alpha_{-i}$ (see (4.3)) an application of John's stability theorem [9, p. 166] now shows that $x(s,t)$ determined by the initial condition $x(s,0) = y_o(s)$ and by the recurrence relation (4.14) satisfies an inequality of the form $||x|| \leq \gamma \cdot ||y||$ (γ being some constant independent of h and y) provided a constant $M > 0$ exists with the property stated in theorem 6. Consequently Ψ_h is stable of order 0 (see theorem 1) and the theorem has thus been proved.

As an illustration of theorem 6 we consider the finite-difference method (4.2) with α_i, β_i from table 1. The expressions in (4.2.b) which the α_i and β_i

i	≤ -3	-2	-1	0	1	2	≥ 3
$12\ \alpha_i$	0	-1	16	-30	16	-1	0
$12\ \beta_i$	0	1	-8	0	8	-1	0

Table 1. *A choice for the parameters in (4.2.b)*

are involved in, now represent $O((\Delta s)^4)$ approximations to the first and second order partial derivatives with respect to s (see e.g. [3]). An easy calculation shows that the function $\psi(\theta)$ of theorem 6 now takes the form $\psi(\theta) = 1 - \{K(s,t)/3\}\cdot\{7 - 8(\cos\theta) + (\cos\theta)^2\}$ the values of which vary between 1 and $1 - 16\,K(s,t)/3$. By choosing the constant ratio $h/(\Delta s)^2$ in such a way that

(4.15) $$\frac{h}{(\Delta s)^2} < \frac{3}{8}\cdot\inf_G \frac{-g_1(s,t)}{g_2(s,t)}$$

we have $\inf\{1 - 16\,K(s,t)/3\} > -1$. Since $\psi(\theta) = 1 - K(s,t)\cdot\theta^2 + O(\theta^4)$ it follows that for $M > 0$ sufficiently small the condition on $\psi(\theta)$ of theorem 6 is fulfilled. By virtue of theorem 6 we thus arrive at the following

Conclusion. When α_i, β_i are taken from table 1 and (4.15) holds (see (4.9)) then the solution $u_h(s,t)$ of the finite-difference scheme (4.2) satisfies (4.13) for each integer $q \ge 1$.

4.5 NUMERICAL ILLUSTRATION

We consider the following initial value problem of type (4.1):

(4.16.a) $U(s,0) - f(s) = 0, \quad -\infty < s < \infty,$

(4.16.b) $-3\cdot\{1 + [U_s(s,t)]^2\}\cdot U_t(s,t) + 2U_{ss}(s,t) + \cos\,[U_{ss}(s,t)] - R(s,t) = 0,$

$-\infty < s < \infty, \ 0 \le t \le T = 1$

where $f(s) = \exp\,[-s^2]$,

$R(s,t) = \{-7 + 8s^2 - 12s^2\cdot\exp\,[2t - 2s^2]\}\cdot\exp\,[t - s^2] + \cos\,\{(2 - 4s^2)\cdot\exp\,[t - s^2]\}$

It is easily verified that $U(s,t) = \exp\,[t - s^2]$ satisfies (4.16).

The initial value problem (4.16) has been solved numerically by the method (4.2) with α_i, β_i from table 1 and with $h/(\Delta s)^2 = 0.37$.
It can be verified (without any knowledge about the true solution U) that (4.15) is satisfied here. Hence the conclusion at the end of section 4.4 applies to the numerical approximations $u_h(s,t)$ obtained and (4.13) thus holds.

h	u_h	u_h^1	u_h^2	u_h^3
1/16	2.86062			
1/32	2.78461	2.70861		
1/64	2.74953	2.71446	2.71640	
1/128	2.73335	2.71716	2.71807	2.71830

Table 2. *Approximations to* $U(0,1) = 2.71828$

In table 2 we have listed the approximations u_h obtained at the point $(s,t) = (0,1)$ for various values of h . The approximations u_h^j $(j = 1,2,3)$ in table 2 have been obtained by fitting a polynomial $P_j(x)$ of degree j with the data u_x $(x = 2^i h;\ i = 0,1,\ldots,j)$ and setting $u_h^j = P_j(0)$ (by performing Neville extrapolation [8, p. 208]). The expansion (4.13) (with $(s,t) = (0,1)$) can be shown to imply that $u_h^j - U(0,1) = O(h^{j+1})$ $(h\to 0)$ (see e.g. [8, p. 240]). The results in table 2 (and further results not listed here) are in accordance with these considerations and confirm the theory described above.

Remarks 1. By applying (4.13) with q high enough it follows easily that under the conditions of theorem 6 not only $u_h(s,t)$ tends towards $U(s,t)$, but also each finite-difference quotient of $u_h(s,t)$ tends towards the corresponding partial derivative of $U(s,t)$ (for $h\to 0$) .

2. In order to relax the smoothness requirements (A), (B), (C) and still to get a result similar to (4.13) it might be appropriate not to deal with condition III, but instead with III*, or to prove directly that the results 1, 2 of theorem 3 hold. A similar remark applies to the more difficult treatment of initial-boundary value problems for nonlinear parabolic equations.

3. In the example treated above we have $p = 1$, $c = 2$, $\alpha = 0$, $m = c+2\alpha=2$ (see definition 3 and theorem 2). Hence (3.8) is violated here and consequently a combination of theorem 2 with the argument following definition 3 would not even yield a proof of mere convergence of $u_h(s,t)$ towards $U(s,t)$ $(h\to 0)$

Acknowledgements. I am indebted to Professor H. Brunner from Halifax for offering me the opportunity to work a few weeks at the Department of Mathematics of Dalhousie University where part of this paper was written. I also wish to thank Mr. C. den Heyer who performed the computations reported in section 4.5 on the IBM 360 - 65 of Leiden University.

REFERENCES

[1] ANSORGE, R.: Konvergenz von Differenzenverfahren für quasilineare Anfangswertaufgaben. Numer. Math. 13, 217 - 225 (1969).

[2] ANSORGE, R., HASS, R.: Konvergenz von Differenzenverfahren für lineare und nichtlineare Anfangswertaufgaben. Lecture Notes in Mathematics 159. Berlin: Springer-Verlag 1970.

[3] COLLATZ, L.: The numerical treatment of differential equations. Berlin: Springer-Verlag 1960.

[4] COLLATZ, L.: Funktionalanalysis und numerische Mathematik. Berlin: Springer-Verlag 1964.

[5] GODUNOV, S.K., RYABENKI, V.S.: Theory of difference schemes. Amsterdam: North-Holland publishing company 1964.

[6] HASS, R.: Stabilität und Konvergenz von Differenzverfahren für halblineare Probleme. Thesis, Hamburg University 1971.

[7] HENRICI, P.: Discrete variable methods in ordinary differential equations. New York: J. Wiley & Sons 1962.

[8] HENRICI, P.: Elements of numerical analysis. New York: J. Wiley & Sons 1964.

[9] JOHN, F.: On integration of parabolic equations by difference methods. Comm. Pure Appl. Math. 5, 155 - 211 (1952).

[10] KANTOROVICH, L.V., Akilov, G.P.: Functional analysis in normed spaces. Oxford: Pergamon Press 1964.

[11] KINNEBROCK, W.: Stabilität und Konvergenz nichtlinearer Differenzenoperatoren bei Anfangswertaufgaben. Report, Gesellschaft für Kernforshung mbH Karlsruhe 1971.

[12] RICHTMYER, R.D., MORTON, K.W.: Difference methods for initial-value problems. New York: John Wiley & Sons 1967.

[13] SPIJKER, M.N.: Stability and convergence of finite-difference methods. Thesis, Leiden University 1968.

[14] STETTER, H.J.: Asymptotic expansions for the error of discretization algorithms for nonlinear functional equations. Numer. Math. 7, 18 - 31 (1965).

[15] STETTER, H.J.: Stability of nonlinear discretization algorithms. In: Numerical solution of partial differential equations, ed. J.H. Bramble. New York: Academic Press 1966.

[16] STRANG, W.G.: Difference methods for mixed boundary-value problems. Duke Math. J. 27, 221 - 231 (1960).

[17] STRANG, G.: Accurate partial difference methods II, Non-linear problems. Numer. Math. 6, 37 - 46 (1964).