Numerische Mathematik 8, 161–177 (1966)

Convergence and Stability of Step-by-step Methods for the Numerical Solution of Initial-value Problems

M. N. SPIJKER

Received July 2, 1965

I. Introduction

Let a and b be real numbers, $a<b$, and let V be a normed real vector space, not necessarily of finite dimension. In the following we shall consider step-by-step methods for the approximation of a certain unknown function X with domain $[a, b]$ and range in V. In chapter II a stability criterion for step-by-step methods will be derived.

In chapter III we shall use the results of chapter II in the case where V is a real Banach space and the unknown function X is a solution of the initial-value problem:

$$\begin{gathered} X^{(p)}=F[X], \\ X(a)=c_0,\ X^{(1)}(a)=c_1,\ \ldots,\ X^{(p-1)}(a)=c_{p-1}, \end{gathered} \tag{1}$$

p being an integer $\geqq 0$ and $X^{(q)}$ being the q-th derivative $\frac{d^q}{dt^q} X$ of X ($q=0, 1, \ldots, p$, for the definition of derivatives and integrals of vector-valued functions, see [1], chapter VI). F is a functional that transforms each vector-valued function Y which is p times continuously differentiable on the interval $[a, b]$, in a new continuous vector-valued function $F[Y]$, also with domain $[a, b]$. F need not be linear. Throughout chapter III we shall assume that for each $c_0, c_1, \ldots, c_{p-1} \in V$ there is at least one solution X to problem (1). If $p=0$ problem (1) is simply $X=F[X]$, and no initial values are given.

Certain step-by-step methods for the approximation of a solution to special problems of type (1) have been studied at length, among others by DAHLQUIST and HENRICI for the case that $p \geqq 1$ and $(F[Y])(t) \equiv f(t, Y(t))$, and $V=R_m=$ the m-dimensional real vector space. In [4], p. 124, a convergence criterion is derived for methods of the form

$$x_{n+1}-x_n=h \cdot \Phi(t_n, x_n, h) \tag{2}$$

where the function Φ has to satisfy certain requirements ($p=1$, the definition of the stepsize h and of the approximation x_n to $X(t_n)$ follows in chapter II). In [6], p. 12, a theorem is stated according to which certain purely algebraic conditions on the polynomials $\varrho(\zeta)=\sum_{i=0}^{k} \alpha_i \zeta^i$ and $\sigma(\zeta)=\sum_{i=0}^{k} \beta_i \zeta^i$ are sufficient for the convergence of the "general linear multistep methods":

$$\sum_{i=0}^{k} \alpha_i x_{n+i}=h \cdot \sum_{i=0}^{k} \beta_i f(t_{n+i}, x_{n+i}) \qquad \text{—here also } p=1\text{—}.$$

HENRICI remarks that generalization of this theorem to arbitrary Banach spaces is possible. In [2] DAHLQUIST discusses methods of the form

$$\sum_{i=0}^{k} \alpha_i x_{n+i} = h^p \cdot \sum_{i=0}^{k} \beta_i f(t_{n+i}, x_{n+i}) \tag{3}$$

for $p = 1, 2, 3, \ldots$.

In chapter III we shall consider step-by-step methods of the form

$$\sum_{i=0}^{k} \alpha_i x_{n+i} = h^p \cdot \Psi_n(x_0, x_1, \ldots, x_n, x_{n+1}, \ldots, x_{n+k}, h) \tag{4}$$

where the functions Ψ_n have to satisfy certain requirements. In III.2 a theorem concerning these methods — theorem 3 — will be derived (for $p \geqq 0$) which can be considered as a generalization of the above-mentioned theorems. This theorem makes it possible to prove or disprove the convergence of methods of type (2) or (3) as well as the relation between accumulated and local errors (stability). Many other methods, e.g. those presented in [7] are also concluded in our theory. As a consequence of the generality of problem (1), it is also possible, with the aid of this theorem, to prove the convergence of methods of type (4) for approximating the solution of certain integral equations, integro-differential equations and partial differential equations. For instance the convergence of step-by-step methods for solving integro-differential equations of the type

$$X^{(p)}(t) = \int_a^t f\left(t, s, X(t), \ldots, X^{(p-1)}(t), X(s), \ldots, X^{(p-1)}(s), X^{(p)}(s)\right) ds$$

with initial conditions $X(a) = c_0, \ldots, X^{(p-1)}(a) = c_{p-1}$ can be proved by this theorem ($p \geqq 0$, and $V = R_m$). Integro-differential equations of the form

$$\frac{\partial}{\partial t} u(s, t) = g\left(s, t, u(s, t)\right) + \int_0^1 f\left(s, t, u(s, t), \sigma, u(\sigma, t)\right) d\sigma$$

with initial condition $u(s, a) = c(s)$ for $0 \leqq s \leqq 1$, can also be solved by methods of type (4) and their speed of convergence can again be determined by theorem 3 (in this case $p = 1$ and $V = C_0[0, 1]$).

In section III.3 we shall establish a second theorem — theorem 4 — concerning methods of type (4). This theorem which holds for $p \geqq 1$ imposes slightly weaker requirements upon the functions Ψ_n. As an application of this theorem we shall present in III.4 some step-by-step methods for the numerical solution of second order initial-value problems of type (1) with $\left(F[Y]\right)(t) \equiv f\left(t, Y(t), Y'(t)\right)$ and $V = R_1$.

The theorems of chapter II and III also enable us to determine the speed of convergence of step-by-step methods which are not of the form (4). E.g., sufficient conditions can be established for the convergence of the methods considered in [5], p. 261, and also their speed of convergence can be determined by these theorems.

Remarks. 1. In the following chapters we have adopted the convention that $\sum_{i=q}^{r} \cdots = 0$ whenever $q > r$.

2. If $g(n, h)$ is an arbitrary real function defined for some integer values n and real values h, then we shall denote by $\max_{n\geqq 0} g(n, h)$ the maximum of the numbers $g(n, h)$, where h is fixed and n may take on all integer values $\geqq 0$ for which $g(n, h)$ is defined (of course it is assumed that there actually exist an $n\geqq 0$ for which $g(n, h)$ makes sense).

II. p-Stability, p-Convergence and p-Consistency

II.1. A General Step-by-step Method

Let V be a normed real vector space; a and b real numbers with $a<b$. Let h_0 be a positive real number, and $0<h\leqq h_0$. We define $t_n^h=a+nh$ for $n=0, 1, 2, \ldots$. If there is no danger of confusion we shall write t_n instead of t_n^h. Let k be a fixed integer $\geqq 0$. Let Ω_n be a sequence of vector-valued functions $(n=0, 1, 2, \ldots)$ that satisfy the following two conditions:

(5) The function $\Omega_n(v_0, \ldots, v_{n+k}, h)$ is defined for $v_i\in V$ $(i=0, 1, \ldots, n+k)$ and $h\cdot(n+k)\leqq b-a$, where $0<h\leqq h_0$.

(6) For any $v_i\in V$ $(i=0, 1, \ldots, n+k)$ and h $(0<h\leqq h_0$ and $h\cdot(n+k)\leqq b-a)$ the equation $x=\Omega_n(v_0, \ldots, v_{n+k-1}, x, h)$ has exactly one solution for x.

The sequence Ω_n is related to a step-by-step method in the following way: we take $x_{n+k}^h=\Omega_n(x_0^h, \ldots, x_n^h, x_{n+1}^h, \ldots, x_{n+k}^h, h)$ for $n=0, 1, 2, \ldots$ and $(n+k)\cdot h\leqq b-a$, where h is some fixed number $\in(0, h_0]$. Hence if $x_0^h, \ldots, x_{k-1}^h$ are given vectors, then vectors $x_k^h, x_{k+1}^h, \ldots$ can be determined successively. In most applications x_n^h is an approximation of $X(t_n^h)$, and we want that $x_n^h\to X(t_n^h)$ if $h\to 0$ and $nh=t-a$ is fixed. $X(t)$ is some unknown function mapping $[a, b]$ into V. Frequently we shall write x_n in stead of x_n^h.

II.2. A p-Stability Criterion

Definition. Let Ω_n be a sequence of functions $(n=0, 1, \ldots)$ satisfying condition (5), and let p be an integer $\geqq 0$.

The sequence Ω_n is *p-stable*, if there is a fixed number $\alpha>0$, and a fixed number h_1, $0<h_1\leqq h_0$, such that whenever

$$y_0=w_0+u_0, \ldots, y_{k-1}=w_{k-1}+u_{k-1}, y_{n+k}=\Omega_n(y_0, \ldots, y_{n+k}, h)+u_{n+k},$$

and

$$\tilde{y}_0=w_0+\tilde{u}_0, \ldots, \tilde{y}_{k-1}=w_{k-1}+\tilde{u}_{k-1}, \tilde{y}_{n+k}=\Omega_n(\tilde{y}_0, \ldots, \tilde{y}_{n+k}, h)+\tilde{u}_{n+k}$$

for $n=0, 1, 2, \ldots$, and $h\cdot(n+k)\leqq b-a$ where $0<h\leqq h_1$,

then $\|y_N-\tilde{y}_N\|\leqq\alpha\cdot\max_{0\leqq i\leqq N}\|u_i-\tilde{u}_i\|$, if $p=0$,

or $\|y_N-\tilde{y}_N\|\leqq\alpha\cdot(N+1)^{p-1}\cdot\sum_{i=0}^{N}\|u_i-\tilde{u}_i\|$, if $p\geqq 1$,

for each integer N, with $0\leqq N\leqq(b-a)/h$. ($\|\ldots\|$ refers to the norm of V). It is clear, that if $k=0$, no vectors w_i will occur in the above definition. If the sequence Ω_n also satisfies condition (6), we call the step-by-step method de-

termined by Ω_n p-stable if and only if the sequence Ω_n is p-stable. For a slightly different definition of (1-)stability see [3].

When a sequence of functions Φ_n, satisfying condition (5) is p-stable, it may be wondered if this stability is preserved when the function values of each Φ_n are changed a little. It will be shown that the p-stability is preserved if

$$\Omega_n(v_0, \ldots, v_{n+k}, h) \equiv \Phi_n(v_0, \ldots, v_{n+k}, h) + h^p \cdot \Psi_n(v_0, \ldots, v_{n+k}, h)$$

represents the new sequence of functions, and each Ψ_n satisfies condition (5) and some other requirements. The following requirement is sufficient:

(7a) There exists a fixed number $\lambda \geqq 0$ such that whenever $v_0, \ldots, v_{n+k}$ and $\tilde{v}_0, \ldots, \tilde{v}_{n+k}$ are vectors $\in V$, then

$$\|\Psi_n(v_0, \ldots, v_{n+k}, h) - \Psi_n(\tilde{v}_0, \ldots, \tilde{v}_{n+k}, h)\| \leqq \lambda h \cdot \sum_{i=0}^{n+k} \|v_i - \tilde{v}_i\|$$

uniformly for $0 < h \leqq h_0$ and $n = 0, 1, \ldots$, where $h(n+k) \leqq b - a$.

If $p \geqq 1$ the following weaker requirement is sufficient: There exist fixed numbers $\lambda \geqq 0$ and $\mu \geqq 0$ and a fixed integer $l \geqq 0$ such that whenever $v_0, \ldots, v_{n+k}$ and $\tilde{v}_0, \ldots, \tilde{v}_{n+k}$ are vectors $\in V$ then

$$\|\Psi_n(v_0, \ldots, v_{n+k}, h) - \Psi_n(\tilde{v}_0, \ldots, \tilde{v}_{n+k}, h)\| \leqq \lambda h \cdot \sum_{i=0}^{n+k} \|v_i - \tilde{v}_i\| + \mu \cdot \sum_{i=0}^{l} \|v_{j(i)} - \tilde{v}_{j(i)}\|,$$

where the $j(i)$ are integers (which may depend on $v_0, \ldots, v_{n+k}$ and $\tilde{v}_0, \ldots, \tilde{v}_{n+k}$ as well as on h and n) with $0 \leqq j(i) \leqq n+k$, uniformly for $0 < h \leqq h_0$ and $n = 0, 1, 2, \ldots$, where $h(n+k) \leqq b - a$.

We note that

$$\begin{aligned} &\lambda h \cdot \sum_{i=0}^{n+k} \|v_i - \tilde{v}_i\| + \mu \cdot \sum_{i=0}^{l} \|v_{j(i)} - \tilde{v}_{j(i)}\| \\ &\quad \leqq \{\lambda h \cdot (n+k+1) + \mu \cdot (l+1)\} \cdot \operatorname*{maximum}_{0 \leqq i \leqq n+k} \|v_i - \tilde{v}_i\| \\ &\quad \leqq \{\lambda \cdot (b - a + h_0) + \mu \cdot (l+1)\} \cdot \operatorname*{maximum}_{0 \leqq i \leqq n+k} \|v_i - \tilde{v}_i\| = \mu_1 \cdot \|v - \tilde{v}\|, \end{aligned}$$

with $\mu_1 = \lambda(b-a+h_0) + \mu(l+1)$, $v = v_j$, $\tilde{v} = \tilde{v}_j$, where j is an integer $(0 \leqq j \leqq n+k)$. The above requirement (for the case $p \geqq 1$) is therefore equivalent to:

(7b) There exists a fixed number $\lambda \geqq 0$ such that whenever $v_0, \ldots, v_{n+k}$ and $\tilde{v}_0, \ldots, \tilde{v}_{n+k}$ are vectors $\in V$ then

$$\|\Psi_n(v_0, \ldots, v_{n+k}, h) - \Psi_n(\tilde{v}_0, \ldots, \tilde{v}_{n+k}, h)\| \leqq \lambda \cdot \operatorname*{maximum}_{0 \leqq i \leqq n+k} \|v_i - \tilde{v}_i\|,$$

uniformly for $0 < h \leqq h_0$ and $n = 0, 1, \ldots$, where $h \cdot (n+k) \leqq b - a$.

The following theorem expresses the above-mentioned preservation of stability:

Theorem 1. Let $p = 0$ and let the sequence Ψ_n satisfy requirement (7a), or let $p \geqq 1$ and the sequence Ψ_n satisfy requirement (7b). Then the sequence $\Omega_n = \Phi_n + h^p \cdot \Psi_n$ is p-stable if and only if the sequence Φ_n is p-stable.

Proof. Let the sequence Φ_n be p-stable. We shall prove that the sequence Ω_n has also this property. Let $w_i, u_i, \tilde{u}_i, y_i$ and $\tilde{y}_i$ be vectors $\in V$ such that

$$y_0 = w_0 + u_0, \ldots, y_{k-1} = w_{k-1} + u_{k-1}, y_{n+k} = \Omega_n(y_0, \ldots, y_{n+k}, h) + u_{n+k}$$

and

$$\tilde{y}_0 = w_0 + \tilde{u}_0, \ldots, \tilde{y}_{k-1} = w_{k-1} + \tilde{u}_{k-1}, \tilde{y}_{n+k} = \Omega_n(\tilde{y}_0, \ldots, \tilde{y}_{n+k}, h) + \tilde{u}_{n+k},$$

for $n = 0, 1, \ldots$ and $h \cdot (n+k) \leqq b - a$, where h is a fixed number $\in (0, h_0]$. We define $d_n = y_n - \tilde{y}_n$ and $e_n = u_n - \tilde{u}_n$. We note that

$$y_{n+k} = \Phi_n(y_0, \ldots, y_{n+k}, h) + \{h^p \cdot \Psi_n(y_0, \ldots, y_{n+k}, h) + u_{n+k}\},$$

and

$$\tilde{y}_{n+k} = \Phi_n(\tilde{y}_0, \ldots, \tilde{y}_{n+k}, h) + \{h^p \cdot \Psi_n(\tilde{y}_0, \ldots, \tilde{y}_{n+k}, h) + \tilde{u}_{n+k}\} \quad \text{for} \quad n \geqq 0.$$

1. $p \geqq 1$. From the p-stability of the sequence Φ_n it follows that there exist constants $\alpha > 0$ and $h_1 > 0$ with

$$\|y_n - \tilde{y}_n\| \leqq \alpha \cdot (n+1)^{p-1} \cdot \Big\{ \sum_{i=0}^{k-1} \|u_i - \tilde{u}_i\| + \sum_{i=k}^{n} \|(h^p \cdot \Psi_{i-k}(y_0, \ldots, y_i, h) + u_i) - (h^p \cdot \Psi_{i-k}(\tilde{y}_0, \ldots, \tilde{y}_i, h) + \tilde{u}_i)\| \Big\},$$

for $0 < h \leqq h_1 \leqq h_0$, and $n \geqq 0$, $hn \leqq b - a$.
In view of condition (7b) there follows

$$\|d_n\| = \|y_n - \tilde{y}_n\| \leqq \alpha (n+1)^{p-1} \cdot \Big\{ \sum_{i=0}^{n} \|e_i\| + \sum_{i=k}^{n} h^p \lambda \cdot \operatorname{maximum}_{0 \leqq j \leqq i} \|y_j - \tilde{y}_j\| \Big\}$$

$$= \alpha \lambda \cdot \{(n+1)h\}^{p-1} \cdot h \cdot \sum_{i=k}^{n} \operatorname{maximum}_{0 \leqq j \leqq i} \|d_j\| + \alpha (n+1)^{p-1} \sum_{i=0}^{n} \|e_i\|.$$

Let $\sum_{i=k}^{n} \max_{0 \leqq j \leqq i} \|d_j\| = \sum_{i=0}^{n} \eta_{n,i} \|d_i\|$, where $\eta_{n,i}$ are integers with $0 \leqq \eta_{n,i} \leqq \eta_{n+1,i}$, $0 \leqq \eta_{n,n} \leqq 1$, and $0 \leqq \sum_{i=0}^{n+1} \eta_{n+1,i} - \sum_{i=0}^{n} \eta_{n,i} \leqq 1$. Let $n \leqq$ some integer N. Then

$$\|d_n\| \leqq \alpha \lambda (b - a + h)^{p-1} h \cdot \Big\{ \eta_{n,n} \|d_n\| + \sum_{i=0}^{n-1} \eta_{n,i} \|d_i\| \Big\} + \alpha (N+1)^{p-1} \cdot \sum_{i=0}^{N} \|e_i\|.$$

Let us now keep $0 < h \leqq h_2$ where h_2 is so small that $h_2 \leqq h_1$ and $\alpha \lambda (b - a + h_2)^{p-1} h_2 < 1$. Then

$$\|d_n\| \leqq \{1 - \alpha \lambda (b - a + h_2)^{p-1} h_2\}^{-1} \times \times \Big\{ \alpha \lambda (b - a + h_2)^{p-1} h \cdot \sum_{i=0}^{n-1} \eta_{n,i} \|d_i\| + \alpha (N+1)^{p-1} \cdot \sum_{i=0}^{N} \|e_i\| \Big\}.$$

Or $\|d_n\| \leqq \beta h \cdot \sum_{i=0}^{n-1} \eta_{n,i} \|d_i\| + S$, where $S = \gamma \cdot (N+1)^{p-1} \cdot \sum_{i=0}^{N} \|e_i\|$, and β and γ are constants, not depending on h and N. Thus, if the numbers D_n are defined by: $D_0 = S$, $D_n = \beta h \cdot \sum_{i=0}^{n-1} \eta_{n,i} D_i + S$, then $\|d_n\| \leqq D_n$ $(n = 0, 1, \ldots, N)$. From

$$D_{n+1} - D_n = \beta h \Big(\sum_{i=0}^{n} \eta_{n+1,i} D_i - \sum_{i=0}^{n-1} \eta_{n,i} D_i \Big) = \beta h \cdot \sum_{i=0}^{n} (\eta_{n+1,i} - \eta_{n,i}) D_i + \beta h \eta_{n,n} D_n$$

it follows that $D_{n+1} \geqq D_n$ for $n = 0, 1, \ldots, N-1$. Thus

$$D_{n+1} - D_n \leqq \beta h \cdot \sum_{i=0}^{n} (\eta_{n+1,i} - \eta_{n,i}) D_n + \beta h \eta_{n,n} D_n$$

$$\leqq \beta h \cdot \left\{ \sum_{i=0}^{n+1} \eta_{n+1,i} - \sum_{i=0}^{n} \eta_{n,i} + \eta_{n,n} \right\} \cdot D_n \leqq 2\beta h \cdot D_n.$$

Consequently $D_{n+1} \leqq (1 + 2\beta h) D_n$ for $n = 0, 1, \ldots, N-1$. From this we get

$$\|y_N - \tilde{y}_N\| = \|d_N\| \leqq D_N \leqq (1 + 2\beta h)^N D_0 \leqq \exp\{2\beta(b-a)\} \cdot S$$

$$= \exp\{2\beta(b-a)\} \cdot \gamma \cdot (N+1)^{p-1} \cdot \sum_{i=0}^{N} \|u_i - \tilde{u}_i\|.$$

From this the p-stability of the sequence Ω_n follows.

2. $p = 0$. From the 0-stability of Φ_n it follows that

$$\|d_n\| = \|y_n - \tilde{y}_n\| \leqq \alpha \cdot \max_{0 \leqq i \leqq n} \|u_i - \tilde{u}_i\| +$$

$$+ \alpha \cdot \max_{k \leqq i \leqq n} \|\Psi_{i-k}(y_0, \ldots, y_i, h) - \Psi_{i-k}(\tilde{y}_0, \ldots, \tilde{y}_i, h)\|,$$

for $0 < h \leqq h_1 \leqq h_0$, and $n \geqq 0$, $nh \leqq b - a$. Let $n \leqq N$ where $Nh \leqq b - a$. In view of condition (7a) we get

$$\|d_n\| \leqq \alpha \cdot \max_{0 \leqq i \leqq N} \|e_i\| + \alpha \lambda h \cdot \max_{k \leqq i \leqq n} \sum_{j=0}^{i} \|y_j - \tilde{y}_j\|$$

$$= \alpha \cdot \max_{0 \leqq i \leqq N} \|e_i\| + \alpha \lambda h \cdot \sum_{i=0}^{n} \|d_i\| = \alpha \lambda h \cdot \|d_n\| + \alpha \lambda h \cdot \sum_{i=0}^{n-1} \|d_i\| + \alpha \cdot \max_{0 \leqq i \leqq N} \|e_i\|.$$

Thus

$$\|d_n\| \leqq (1 - \alpha \lambda h)^{-1} \cdot \left\{ \alpha \lambda h \cdot \sum_{i=0}^{n-1} \|d_i\| + \alpha \cdot \max_{0 \leqq i \leqq N} \|e_i\| \right\}$$

$$\leqq (1 - \alpha \lambda h_2)^{-1} \cdot \left\{ \alpha \lambda h \cdot \sum_{i=0}^{n-1} \|d_i\| + \alpha \cdot \max_{0 \leqq i \leqq N} \|e_i\| \right\} = \beta h \cdot \sum_{i=0}^{n-1} \|d_i\| + M,$$

provided $0 < h \leqq h_2$ where h_2 is so small that $h_2 \leqq h_1$ and $\alpha \lambda h_2 < 1$. $M = \gamma \cdot \max_{0 \leqq i \leqq N} \|e_i\|$, and β and γ are constants, not depending on h and N. Let $D_0 = M$ and $D_n = \beta h \cdot \sum_{i=0}^{n-1} D_i + M$ for $n = 0, 1, \ldots, N$, then $\|d_N\| \leqq D_N$. From this it follows that

$$\|y_N - \tilde{y}_N\| = \|d_N\| \leqq D_N = (1 + \beta h)^N D_0 \leqq \exp\{\beta(b-a)\} \cdot \gamma \cdot \max_{0 \leqq i \leqq N} \|u_i - \tilde{u}_i\|.$$

This proves the p-stability of the sequence Ω_n.

It remains to be shown that if Ω_n is p-stable then also the sequence Φ_n is p-stable. In order to prove this we only have to write $\Phi_n = \Omega_n + h^p \cdot (-\Psi_n)$ and to apply the results just proved. This completes the proof of theorem 1.

Remark. Let V be a space with an element $v \neq 0$. If $p = 0$, condition (7a) in theorem 1 cannot be replaced by (7b) as is shown by the following example: $k = 1$, $\Phi_n(v_0, \ldots, v_{n+1}, h) \equiv 0$. The sequence Φ_n is 0-stable. If we take $\Psi_n(v_0, \ldots, v_{n+1}, h) \equiv 2 \cdot v_n$ then condition (7b) is satisfied, but the sequence $\Omega_n = \Phi_n + \Psi_n$ is not 0-stable.

II.3. A Sufficient Condition for p-Convergence

Let Ω_n be a sequence of functions satisfying (5) and (6). Let X be a function from $[a, b]$ to V, and let p be an integer $\geqq 0$.

Definition. The vectors $e_{n+k}=e_{n+k}^h=\Omega_n\left(X(t_0^h), \ldots, X(t_{n+k}^h), h\right)-X(t_{n+k}^h)$ are called *local discretization errors.*

Definition. The method determined by the sequence Ω_n — or briefly: the sequence Ω_n — is *p-consistent* with the function X if $\max\limits_{n\geqq 0}\|e_{n+k}^h\|=o(h^p)$ when $h\to 0$.

Definition. The vectors $d_m=d_m^h=x_m^h-X(t_m^h)$, where the x_m^h are determined successively by the sequence Ω_n starting with predetermined vectors $x_0^h, \ldots, x_{k-1}^h$, are called *accumulated discretization errors.*

Definition. The method determined by Ω_n — or briefly: the sequence Ω_n — is *p-convergent* with respect to the function X if $\lim\limits_{h\to 0}\max\limits_{n\geqq 0}\|d_n^h\|=0$, whenever the starting vectors $x_0^h, \ldots, x_{k-1}^h$ satisfy $\|d_i^h\|=o(h^{q-1})$ $(i=0, 1, \ldots, k-1)$ where $q=\max(1, p)$.

Definition. Let the accumulated discretization errors satisfy $\max\limits_{n\geqq 0}\|d_n^h\|=o(h^{r-1})$ for any set of starting values $x_0, \ldots, x_{k-1}$ with $\|d_i^h\|=o(h^s)$ $(i=0, 1, \ldots, k-1)$, where s is some fixed number. Then the method determined by Ω_n is called of *order r.*

Let Ω_n be a p-stable sequence that is q-consistent with some function X, where $q\geqq p\geqq 0$. We apply the p-stability on the sequences $y_0, y_1, \ldots$ and $\tilde{y}_0, \tilde{y}_1, \ldots$ where $y_n=x_n$, $u_n=0$ $(n\geqq 0)$ and $\tilde{y}_n=X(t_n), \tilde{u}_i=X(t_i)-x_i, \tilde{u}_{n+k}=-e_{n+k}$ (hence equals minus the local discretization error) $(n\geqq 0,\ i=0, 1, \ldots, k-1)$. If $p=0$ we get

$$\|y_N-\tilde{y}_N\|\leqq\alpha\cdot\max_{0\leqq i\leqq k-1}\|\tilde{u}_i\|+o(h^q).$$

If $p\geqq 1$ we get

$$\|y_N-\tilde{y}_N\|\leqq\alpha\cdot(N+1)^{p-1}\cdot\sum_{i=0}^{N}\|\tilde{u}_i\|=\alpha\cdot(N+1)^{p-1}\cdot\sum_{i=0}^{k-1}\|\tilde{u}_i\|+\alpha\cdot(N+1)^{p-1}\cdot\sum_{i=k}^{N}\|\tilde{u}_i\|$$

$$\leqq\alpha\cdot(N+1)^{p-1}\cdot\sum_{i=0}^{k-1}\|\tilde{u}_i\|+\alpha\cdot(N+1)^{p-1}\cdot(N-k+1)\cdot\max_{k\leqq i\leqq N}\|\tilde{u}_i\|$$

$$\leqq\alpha\cdot(N+1)^{p-1}\cdot\sum_{i=0}^{k-1}\|\tilde{u}_i\|+o(h^{q-p}).$$

Thus

$$\|d_n\|=\|x_n-X(t_n)\|=\|y_n-\tilde{y}_n\|=o(h^{q-p})\qquad(n\geqq 0),$$

for any set of starting values $x_0, \ldots, x_{k-1}$ with $\|d_i^h\|=o(h^s)$ $(i=0, 1, \ldots, k-1)$ where $s=q-\min(1, p)$. Consequently the method Ω_n is of order $q-p+1$. Let $q=p\geqq 0$. We then have the following:

Theorem 2. If the sequence Ω_n is p-stable and p-consistent with the function X then the sequence is p-convergent with respect to X.

Remarks. 1. If $k=0$, no starting values $x_0, \ldots, x_{k-1}$ occur in the above definitions.

2. If there is a vector $v\neq 0$ in V, then theorem 2 cannot be reversed: for each $p\geqq 0$ it is possible to construct a sequence Ω_n that satisfies (5) and (6)

and that is p-convergent, p-consistent but not p-stable. Also a sequence Ω_n can be constructed that is p-convergent, p-stable but not p-consistent with some function X.

III. Step-by-step Methods of the Form (4)

III.1. The Characteristic Polynomial

Let V be a normed real vector space with a vector $v \neq 0$, and let $a < b$. We shall derive a criterion for p-convergence for methods of the type (4) with respect to the solutions to problem (1). In view of theorem 1 and 2 we first determine the condition for p-stability of the sequence Φ_n where

$$\Phi_n(v_0, \ldots, v_{n+k}, h) \equiv -\sum_{i=0}^{k-1} \alpha_i v_{n+i}, \qquad (p \geqq 0),$$

k being a fixed integer $\geqq 0$ and $\alpha_0, \alpha_1, \ldots, \alpha_k$ real numbers, $\alpha_k = 1$. To the sequence Φ_n we adjoin the characteristic polynomial $\varrho(\zeta) = \sum_{i=0}^{k} \alpha_i \zeta^i$, where ζ is a complex variable.

Lemma. A necessary and sufficient condition for the p-stability of the sequence Φ_n is that all roots of the polynomial $\varrho(\zeta)$ should have a modulus $\leqq 1$, while the multiplicity of the roots with modulus 1 should not exceed p.

Proof. Let

$$y_0 = w_0 + u_0, \ldots, y_{k-1} = w_{k-1} + u_{k-1}, \qquad \sum_{i=0}^{k} \alpha_i y_{n+i} = u_{n+k}$$

and

$$\tilde{y}_0 = w_0 + \tilde{u}_0, \ldots, \tilde{y}_{k-1} = w_{k-1} + \tilde{u}_{k-1}, \qquad \sum_{i=0}^{k} \alpha_i \tilde{y}_{n+i} = \tilde{u}_{n+k}$$

for $n \geqq 0$, $(n+k)h \leqq b-a$. Writing $d_n = y_n - \tilde{y}_n$ and $e_n = u_n - \tilde{u}_n$ we get $d_i = e_i$ $(0 \leqq i \leqq k-1)$ and $\sum_{i=0}^{k} \alpha_i d_{n+i} = e_{n+k}$ $(n \geqq 0,\ (n+k)h \leqq b-a)$. We shall assume, with no loss of generality, that $k \geqq 1$ and $\alpha_0 \neq 0$.

It can be shown that $d_n = \sum_{m=0}^{n} \eta_n^{(m)} e_m$ where $\eta_n^{(m)}$ are real numbers $(n, m = 0, 1, 2, \ldots)$ satisfying the following conditions: $\eta_n^{(m)} = 0$ for $m > n$, $\eta_n^{(m)} = 0$ for $m < n < k$, $\eta_n^{(m)} = 1$ for $m = n$, $\sum_{i=0}^{k} \alpha_i \eta_{n+i}^{(m)} = 0$ for $m < n+k$ (cf. [4], p. 212).

1. Assume first that the roots of the characteristic polynomial satisfy the conditions of the lemma. Let $\vartheta_n^{(0)}, \ldots, \vartheta_n^{(k-1)}$ be the usual fundamental system of the homogeneous difference equation with characteristic polynomial $\varrho(\zeta)$:

$\vartheta_n^{(i)} = n^{r_i} \omega_i^n$ with $|\omega_i| = 1$ and $0 \leqq r_i \leqq p-1$, or $|\omega_i| < 1$ and $0 \leqq r_i$.

$$\eta_n^{(j)} = \sum_{i=0}^{k-1} \mu_i^{(j)} \vartheta_n^{(i)} \quad \text{and} \quad \eta_n^{(m)} = \eta_{n+k-m-1}^{(k-1)} = \sum_{i=0}^{k-1} \mu_i \vartheta_{n+k-m-1}^{(i)}$$

for $0 \leqq j \leqq k-1$ and $n \geqq m \geqq k-1$ and some constants $\mu_i^{(j)}$, $\mu_i = \mu_i^{(k-1)}$.

Let $p \geqq 1$. Then $|\vartheta_n^{(i)}| = n^{r_i}|\omega_i|^n \leqq \gamma(n+1)^{p-1}$ for some constant γ and for $n \geqq 0$. Thus $|\eta_n^{(m)}| \leqq \alpha(n+1)^{p-1}$ for some constant α. Consequently

$$\|d_n\| \leqq \alpha(n+1)^{p-1} \cdot \sum_{m=0}^{n} \|e_m\|.$$

Let $p=0$. Then $|\vartheta_n^{(i)}| = n^{r_i}|\omega_i|^n \leqq \delta \cdot \beta^n$ for some $\delta > 0$, $1 > \beta > 0$ $(i = 0, 1, 2, \ldots, k-1)$. $|\eta_n^{(j)}| \leqq \mu \beta^n \leqq \mu \beta$ and $|\eta_n^{(m)}| \leqq \mu \beta^{n+k-m-1}$ for some constant $\mu > 0$ $(0 \leqq j \leqq k-1,\ n \geqq m \geqq k-1)$. Consequently

$$\|d_n\| \leqq [(k-1)\mu\beta + \mu(\beta^n + \beta^{n-1} + \cdots + \beta^{k-1})] \cdot \max_{0 \leqq i \leqq n} \|e_i\| \leqq \alpha \cdot \max_{0 \leqq i \leqq n} \|e_i\|$$

with $\alpha = \mu \cdot \left[(k-1)\beta + \frac{\beta^{k-1}}{1-\beta}\right]$. This completes the first part of the proof.

2. Assume now that the sequence Φ_n is p-stable. We shall prove that the roots of $\varrho(\zeta)$ satisfy the above conditions.

Let $p \geqq 1$. If we take $e_n = 0$ for all $n \neq m$ where m is one of the integers $0, 1, \ldots, k-1$ and $e_m = v \neq 0$, then $d_n = \eta_n^{(m)} v$. Thus

$|\eta_n^{(m)}| \cdot \|v\| = \|d_n\| \leqq \alpha \cdot (n+1)^{p-1} \cdot \|v\|$ and $\frac{|\eta_n^{(m)}|}{(n+1)^{p-1}} \leqq \alpha \qquad (n = 0, 1, \ldots)$.

From this it follows that for any set of coefficients γ_m

$\frac{\left|\sum_{m=0}^{k-1} \gamma_m \eta_n^{(m)}\right|}{n^{p-1}}$ is bounded for $n \to \infty$. As $\eta_n^{(0)}, \ldots, \eta_n^{(k-1)}$ is a fundamental system it follows that $\frac{|\vartheta_n^{(i)}|}{n^{p-1}}$ is bounded for $n \to \infty$ $(i = 0, 1, \ldots, k-1$ and $\vartheta_n^{(0)}, \ldots, \vartheta_n^{(k-1)}$ is the fundamental system defined above). Consequently the roots of $\varrho(\zeta)$ satisfy the requirements.

Let $p = 0$. From the 0-stability of Φ_n it follows that this sequence is also 1-stable. Hence it follows from the above that no root of the polynomial $\varrho(\zeta)$ has a modulus exceeding 1 and that the multiplicity of the roots with modulus 1 is at most 1. We shall prove that the assumption that $\varrho(\zeta)$ has roots of modulus 1 leads to a contradiction.

Let $\omega_0, \ldots, \omega_r$ be the roots of modulus 1. Let $e_0 = e_1 = \cdots = e_{k-1} = 0$ and $\vartheta_n^{(j)} = \omega_j^n$ for $j = 0, 1, \ldots, r$. In view of

$$\eta_n^{(m)} = \eta_{n+k-1-m}^{(k-1)} = \sum_{j=0}^{k-1} \mu_j \vartheta_{n+k-m-1}^{(j)}$$

we get

$$d_n = \sum_{m=k}^{n} \left\{ \sum_{j=0}^{r} \mu_j \omega_j^{n-m+k-1} + \sum_{j=r+1}^{k-1} \mu_j \vartheta_{n-m+k-1}^{(j)} \right\} e_m$$

where $|\vartheta_n^{(j)}| \leqq \delta \beta^n$ $(j = r+1, \ldots, k-1)$ with $\delta > 0$, $0 < \beta < 1$. The numbers μ_j satisfy:

$$0 = \eta_0^{(k-1)} = \sum_{j=0}^{k-1} \mu_j \vartheta_0^{(j)}, \ldots, \qquad 0 = \eta_{k-2}^{(k-1)} = \sum_{j=0}^{k-1} \mu_j \vartheta_{k-2}^{(j)},$$

$$1 = \eta_{k-1}^{(k-1)} = \sum_{j=0}^{k-1} \mu_j \vartheta_{k-1}^{(j)}.$$

Let D be the matrix of the coefficients of this set of linear equations for the μ_j, and D_j the matrix D with the $(j+1)$-th column replaced by the column $\begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$.

It can be proved that the determinant $|D|$ of D don't vanishes (cf. [4], p. 214). Hence by Cramer's rule: $\mu_j = |D_j|/|D|$ $(j=0, 1, \ldots, k-1)$. The numbers μ_j satisfy: a) $\mu_j \neq 0$; b) if $\omega_j = \pm 1$ then $\mu_j =$ real; c) if $\omega_i = \bar{\omega}_j$ $(i \neq j)$ then $\mu_i = \bar{\mu}_j$ $(i, j = 0, 1, \ldots, r$; $\bar{\alpha}$ denoting the complex conjugate of the number $\alpha)$.

a) follows from $\mu_j = |D_j|/|D|$, and $|D_j| \neq 0$ for $j=0, 1, \ldots, r$ (cf. [4], p. 214). For $\omega_j = \pm 1$ we have: $|\bar{D}| = (-1)^s |D|$, $|\bar{D}_j| = (-1)^s |D_j|$, $2s$ being the number of columns with at least one non real element, appearing in D. Hence $\bar{\mu}_j = |\bar{D}_j|/|\bar{D}| = |D_j|/|D| = \mu_j$. This proves b). If $\omega_i = \bar{\omega}_j$ $(i \neq j)$ then $|\bar{D}_j| = (-1)^s |D_i|$. Therefore c) holds.

The assumption $\omega_0 = 1$ leads to

$$d_n = \sum_{m=k}^{n} \left\{ \mu_0 + \sum_{j=1}^{r} \mu_j \omega_j^{n-m+k-1} + \sum_{j=r+1}^{k-1} \mu_j \vartheta_{n-m+k-1}^{(j)} \right\} v$$

if we take $e_m = v \neq 0$ for $m \geqq k$. In view of b) we have:

$$d_n = (n-k+1)\mu_0 v + \left\{ \sum_{j=1}^{r} \mu_j \sum_{m=k}^{n} \omega_j^{n-m+k-1} + \sum_{j=r+1}^{k-1} \mu_j \sum_{m=k}^{n} \vartheta_{n-m+k-1}^{(j)} \right\} v.$$

As $\omega_j \neq 1$ $(1 \leqq j \leqq r)$ the second term of the right-hand member of this equality is bounded for $n \to \infty$. Hence it follows from a) that $\|d_n\| \to \infty$ $(n \to \infty)$, which is a contradiction in view of the 0-stability.

In a similar way the assumption $\omega_0 = -1$ leads to a contradiction if we take $e_m = (-1)^m v$. Thus $\omega_j \neq \pm 1$ $(j=0, 1, \ldots, r)$. The assumption $\omega_0 = \exp\{i\varphi\}$, $\omega_1 = \exp\{-i\varphi\}$ with $0 < \varphi < \pi$ leads to a contradiction if we put $e_m = (2\cos m\varphi) \cdot v$. This completes the proof.

III.2. A Convergence Criterion for the Case $p \geqq 0$

Let V be a Banach space with a vector $v \neq 0$. Let $\varrho(\zeta)$ denote the same polynomial as in III.1, and let

$$\Omega_n(v_0, \ldots, v_{n+k}, h) \equiv -\sum_{i=0}^{k-1} \alpha_i v_{n+i} + h^p \cdot \Psi_n(v_0, \ldots, v_{n+k}, h)$$

where the sequence Ψ_n satisfies (7a) if $p=0$, or (7b) if $p \geqq 1$. Let h_2 be so small that $0 < h_2 \leqq h_0$ and $h_2 \lambda < 1$ $(p=0)$ or $h_2^p \lambda < 1$ $(p \geqq 1)$, where λ is the constant appearing in (7a) or (7b). For $0 < h \leqq h_2$ the function

$$y = \psi(x) = -\sum_{i=0}^{k-1} \alpha_i v_{n+i} + h^p \cdot \Psi_n(v_0, \ldots, v_{n+k-1}, x, h)$$

will then be a contracting mapping from V in V. This follows from

$$\|\psi(x) - \psi(\tilde{x})\| = h^p \cdot \|\Psi_n(v_0, \ldots, v_{n+k-1}, x, h) - \Psi_n(v_0, \ldots, v_{n+k-1}, \tilde{x}, h)\| \leqq h^q \lambda \cdot \|x - \tilde{x}\|,$$

with $q = \max(1, p)$. Consequently the equation $x = \psi(x)$ has exactly one solution (see [1], chapter I, § 7). Condition (6) is therefore satisfied if we take $h_0 = h_2$, and the sequence Ω_n defines a step-by-step method.

For $k \geqq 0$ and $p \geqq 0$ we shall prove the following

Theorem 3. *A.* The sequence Ω_n is p-stable if and only if the roots of the polynomial $\varrho(\zeta)$ have modulus $\leqq 1$, and the multiplicity of the roots of modulus 1 is at most p.

B. If the sequence Ω_n is p-stable, the following three propositions will be equivalent:

(P1) $\zeta=1$ is a root of multiplicity p of the polynomial $\varrho(\zeta)$.

$$\lim_{h\to 0}\max_{n\geqq 0}\left\|\varrho^{(p)}(1)\cdot F[X](t_n^h)-p!\,\Psi_n(X(t_0^h),\ldots,X(t_{n+k}^h),h)\right\|=0$$

for any solution $X(t)$ of the equation $X^{(p)}=F[X]$.

(P2) The sequence Ω_n is p-consistent with each solution of the equation $X^{(p)}=F[X]$.

(P3) The sequence Ω_n is p-convergent with respect to each solution of the equation $X^{(p)}=F[X]$.

For each solution $X(t)$ there exists a continuous function f from $[a,b]$ into V that satisfies:

$$\lim_{h\to 0}\max_{n\geqq 0}\left\|\Psi_n(X(t_0^h),\ldots,X(t_{n+k}^h),h)-f(t_n^h)\right\|=$$

$$\lim_{h\to 0}\max_{n\geqq 0}\left\|\Psi_n(x_0^h,\ldots,x_{n+k}^h,h)-f(t_n^h)\right\|=0$$

for any set of vectors x_m^h determined by the sequence Ω_n $(m\geqq k)$ with $x_i^h-X(t_i^h)=o(h^{q-1})$ $(i=0,1,\ldots,k-1,\ q=\max(1,p))$.

Proof. A. From theorem 1 and the lemma of section III.1 it follows that the above condition for the roots of $\varrho(\zeta)$ is necessary and sufficient for the p-stability of the sequence Ω_n.

B. If the function X from $[a,b]$ into V is p times continuously differentiable, we have

$$X(t+h)=X(t)+h\cdot X^{(1)}(t)+\cdots+\frac{h^p}{p!}\cdot X^{(p)}(t)+\frac{h^p}{p!}\cdot S(t,h)$$

for $a\leqq t\leqq b$, $a\leqq t+h\leqq b$ with $\|S(t,h)\|\leqq$ supremum $\|X^{(p)}(s_1)-X^{(p)}(s_2)\|$ where s_1 and s_2 may take on all values in the interval limited by t and $t+h$.

1. We assume that (P1) holds; we shall prove (P2). From $\varrho(1)=\varrho^{(1)}(1)=\cdots=\varrho^{(p-1)}(1)=0$ it follows that for $p\geqq 1$

$\sum_{i=0}^{k} i^m\alpha_i=0$ $(m=0,1,\ldots,p-1)$. For $p\geqq 0$

$\sum_{i=0}^{k} i^p\alpha_i=\varrho^{(p)}(1)$. Let $X(t)$ be a solution to (1), and e_{n+k}^h the corresponding local discretization errors:

$$\begin{aligned}
e_{n+k}^h&=\Omega_n(X(t_0^h),\ldots,X(t_{n+k}^h),h)-X(t_{n+k}^h)\\
&=-\sum_{i=0}^{k-1}\alpha_i X(t_{n+i}^h)+h^p\cdot\Psi_n(X(t_0^h),\ldots,X(t_{n+k}^h),h)-X(t_{n+k}^h)\\
&=-\sum_{i=0}^{k}\alpha_i\left\{X(t_n^h)+ih\cdot X^{(1)}(t_n^h)+\cdots+\frac{i^p h^p}{p!}\cdot X^{(p)}(t_n^h)+\frac{i^p h^p}{p!}\cdot S(t_n^h,ih)\right\}+\\
&\quad+h^p\cdot\Psi_n=h^p\cdot\Psi_n\left(X(t_0^h),\ldots,X(t_{n+k}^h),h\right)-\frac{h^p}{p!}\varrho^{(p)}(1)\cdot X^{(p)}(t_n^h)-\\
&\quad-\frac{h^p}{p!}\sum_{i=0}^{k}\alpha_i i^p S(t_n^h,ih).
\end{aligned}$$

Consequently

$$p!\,\|e_{n+k}^h\| \leqq h^p \cdot \|\varrho^{(p)}(1)\cdot F[X](t_n^h) - p!\,\Psi_n(X(t_0^h),\ldots,X(t_{n+k}^h),h)\| + \\ + h^p \cdot \sum_{i=0}^{k} i^p|\alpha_i| \cdot \|S(t_n^h, ih)\|.$$

Since the function $X^{(p)}(t)$ is uniformly continuous on the interval $[a, b]$, it follows from (P1) that the sequence Ω_n is p-consistent with $X(t)$.

2. Let (P2) hold. We shall prove (P3). Let X be a solution to (1). In view of theorem 2 the sequence Ω_n is p-convergent with respect to $X(t)$. From $\|e_{n+k}^h\| = o(h^p)$ it follows that

$$\left\|\Psi_n(X(t_0^h),\ldots,X(t_{n+k}^h),h) - \frac{1}{h^p}\sum_{j=0}^{p-1}\frac{h^j}{j!}\left(\sum_{i=0}^{k} i^j\alpha_i\right)\cdot X^{(j)}(t_n^h) - \frac{1}{p!}\left(\sum_{i=0}^{k} i^p\alpha_i\right)\cdot X^{(p)}(t_n^h)\right\|$$

$$= \frac{1}{h^p}\cdot\left\|h^p\cdot\Psi_n(X(t_0^h),\ldots,X(t_{n+k}^h),h) - \sum_{i=0}^{k}\alpha_i\left\{X(t_n^h) + ih\cdot X^{(1)}(t_n^h) + \cdots + \right.\right.$$

$$\left.\left. + \frac{i^p h^p}{p!}X^{(p)}(t_n^h)\right\}\right\| = \frac{1}{h^p}\cdot\left\|h^p\cdot\Psi_n - \sum_{i=0}^{k-1}\alpha_i X(t_{n+i}^h) - X(t_{n+k}^h) + \right.$$

$$\left. + \sum_{i=0}^{k}\alpha_i\frac{i^p h^p}{p!}S(t_n^h, ih)\right\| \leqq \frac{\|e_{n+k}^h\|}{h^p} + \frac{1}{p!}\sum_{i=0}^{k} i^p|\alpha_i|\cdot\|S(t_n^h, ih)\| \to 0$$

for $h\to 0$, uniformly for $n\geqq 0$.

Let $p\geqq 1$. a) Let $n=0$ and $X^{(j)}(a)=c_j=0$ $(j=0, 1, \ldots, p-1)$. Thus

$$\left\|\Psi_0(X(t_0^h),\ldots,X(t_k^h),h) - \frac{1}{p!}\left(\sum_{i=0}^{k} i^p\alpha_i\right)X^{(p)}(a)\right\| \to 0 \qquad (h\to 0)$$

where $X(t)$ is a solution of (1) with $c_j=0$ $(j=0, 1, \ldots, p-1)$. From

$$\|\Psi_0(0,\ldots,0,h)\| \leqq \|\Psi_0(0,\ldots,0,h) - \Psi_0(X(t_0^h),\ldots,X(t_k^h),h)\| + \\ + \left\|\Psi_0(X(t_0^h),\ldots,X(t_k^h),h) - \frac{1}{p!}\left(\sum_{i=0}^{k} i^p\alpha_i\right)X^{(p)}(a)\right\| + \frac{1}{p!}\left|\sum_{i=0}^{k} i^p\alpha_i\right|\cdot\|X^{(p)}(a)\|$$

and property (7b) it follows that $\Psi_0(0, \ldots, 0, h)$ is bounded for $h\to 0$.

b) Let $n=0$, $X^{(j)}(a)=c_j=0$ $(j\neq j_0,\ j\leqq p-1)$ and $X^{(j_0)}(a)=c_{j_0}=v\neq 0$ $(j_0\leqq p-1)$. Thus

$$\frac{1}{h^{p-j_0}}\cdot\left\|\frac{1}{j_0!}\left(\sum_{i=0}^{k} i^{j_0}\alpha_i\right)v\right\| \leqq \left\|\Psi_0(X(t_0^h),\ldots,X(t_k^h),h) - \frac{1}{h^p}\sum_{j=0}^{p-1}\frac{h^j}{j!}\left(\sum_{i=0}^{k} i^j\alpha_i\right)c_j - \right.$$

$$\left. - \frac{1}{p!}\left(\sum_{i=0}^{k} i^p\alpha_i\right)X^{(p)}(a)\right\| + \|\Psi_0(X(t_0^h),\ldots,X(t_k^h),h)\| + \frac{1}{p!}\left|\sum_{i=0}^{k} i^p\alpha_i\right|\cdot\|X^{(p)}(a)\|$$

where $X(t)$ is a solution of (1) with c_j as defined above. Since

$$\|\Psi_0(X(t_0^h),\ldots,X(t_k^h),h)\| \leqq \|\Psi_0(X(t_0^h),\ldots,X(t_k^h),h) - \Psi_0(0,\ldots,0,h)\| + \|\Psi_0(0,\ldots,0,h)\| \\ \leqq \lambda\cdot\max_{a\leqq t\leqq b}\|X(t)\| + \|\Psi_0(0,\ldots,0,h)\|$$

it follows that $\sum_{i=0}^{k} i^{j_0}\alpha_i$ must be $=0$. Thus $\sum_{i=0}^{k} i^j\alpha_i=0$ $(j=0, 1, \ldots, p-1)$.

For $p \geqq 0$ we now define $f(t) = \frac{1}{p!}\left(\sum_{i=0}^{k} i^p \alpha_i\right) X^{(p)}(t)$. We note that

$$\|\Psi_n(X(t_0^h), \ldots, X(t_{n+k}^h), h) - f(t_n^h)\| \to 0$$

($h \to 0$, uniformly for $n \geqq 0$). From this and property (7a) or (7b) we obtain (P3).

3. Let (P3) hold. We shall prove (P1). Suppose first that $p = 0$. Let $X(t)$ be a solution to (1). If $k \geqq 1$ we take $x_i = X(t_i)$ $(i = 0, 1, \ldots, k-1)$. We define x_m for $m \geqq k$ by $\sum_{i=0}^{k} \alpha_i x_{n+i} = \Psi_n(x_0, \ldots, x_{n+k}, h)$ $(n \geqq 0,\ h(n+k) \leqq b-a)$. From (P3) it follows that the right-hand members of the following inequality tend to zero as $h \to 0$, uniformly for $n \geqq 0$:

$$\|\varrho(1) \cdot F[X](t_n) - \Psi_n(X(t_0), \ldots, X(t_{n+k}), h)\| \leqq \left\|\sum_{i=0}^{k} \alpha_i \{X(t_n) - X(t_{n+i})\}\right\| +$$
$$+ \left\|\sum_{i=0}^{k} \alpha_i \{X(t_{n+i}) - x_{n+i}\}\right\| + \|\Psi_n(x_0, \ldots, x_{n+k}, h) - \Psi_n(X(t_0), \ldots, X(t_{n+k}), h)\|.$$

Thus the left-hand side of this inequality also tends to zero, which is the content of (P1) for $p = 0$.

Let $p \geqq 1$. It is clear that also $k \geqq 1$. If $y_0, y_1, \ldots$ is any sequence of vectors we define $\Delta y_n = y_{n+1} - y_n$, $E y_n = y_{n+1}$, $\nabla y_m = y_m - y_{m-1}$ for $n \geqq 0$, $m \geqq 1$. It can be proved by induction that

$$(8) \qquad \sum_{n_q=0}^{n} \sum_{n_{q-1}=0}^{n_q} \cdots \sum_{n_2=0}^{n_3} \sum_{n_1=0}^{n_2} \Delta^q y_{n_1} = y_{n+q} - \sum_{j=0}^{q-1} \binom{n+j}{j} \nabla^j y_{q-1}$$

for $q = 1, 2, 3, \ldots$, $n = 0, 1, 2, \ldots$.

Let $X(t)$ be a solution to the equation $X^{(p)} = F[X]$. We define

$$x_i^h = X(a) + (ih) X^{(1)}(a) + \cdots + \frac{(ih)^{p-1}}{(p-1)!} X^{(p-1)}(a)$$

for $i = 0, 1, \ldots, k-1$. It can be proved that

$$(9) \quad \nabla^j x_i^h = h^j X^{(j)}(a) + O(h^{j+1}) \quad \text{for} \quad i = 0, 1, \ldots, k-1;\ j = 0, 1, \ldots, \min(i, p-1).$$

We define x_m^h for $m \geqq k$ by $\varrho(E) x_n^h = h^p \cdot \Psi_n(x_0^h, \ldots, x_{n+k}^h, h)$. We shall prove that $\varrho(1) = 0$. From (P3) it follows that the right-hand members of the inequality

$$|\varrho(1)| \cdot \|X(a)\| \leqq \left\|\sum_{i=0}^{k} \alpha_i \{X(a) - x_i^h\}\right\| + h^p \cdot \|\Psi_0(x_0^h, \ldots, x_k^h, h) - f(a)\| + h^p \cdot \|f(a)\|$$

tend to zero as $h \to 0$. Considering the case $X(a) = c_0 = v \neq 0$ we see that $\varrho(1) = 0$. Next we shall prove that $\zeta = 1$ is a root of multiplicity p. Let $\varrho(\zeta) = \sigma(\zeta) \cdot (\zeta - 1)^q$ where $\sigma(\zeta)$ is a polynomial and $1 \leqq q \leqq p$. With the definition $y_n = \sigma(E) x_n$ it follows from (8) and (9) that

$$\sum_{n_q=0}^{n} \cdots \sum_{n_1=0}^{n_2} \varrho(E) x_{n_1} = \sum_{n_q=0}^{n} \cdots \sum_{n_1=0}^{n_2} \Delta^q y_{n_1} = y_{n+q} - \sum_{j=0}^{q-1} \binom{n+j}{j} \sigma(E) \nabla^j x_{q-1}$$
$$= \sigma(E) x_{n+q} - \sum_{j=0}^{q-1} \binom{n+j}{j} \sigma(E) \{h^j X^{(j)}(a) + O(h^{j+1})\}.$$

Thus

$$h^p \sum_{n_q=0}^{n} \cdots \sum_{n_1=0}^{n_2} \Psi_{n_1}(x_0, \ldots, x_{n_1+k}, h) = \sigma(E)\, x_{n+q} - \sigma(1) \sum_{j=0}^{q-1} \binom{n+j}{j} \{h^j X^{(j)}(a) + O(h^{j+1})\}. \tag{10}$$

The assumption $q<p$ leads to $\sigma(1) \cdot \left\{X(t) - \sum_{j=0}^{q-1} \frac{(t-a)^j}{j!} X^{(j)}(a)\right\} = 0$ if $h \to 0$ in (10) ($nh = t-a < b-a$). If $\sigma(1) \neq 0$, $X(t)$ would only depend on $X(a), \ldots, X^{(q-1)}(a)$ and not on $X^{(i)}(a)$ for $q \leqq i \leqq p$. Thus $\sigma(1)=0$ must hold if $q<p$; $\zeta=1$ is a root of multiplicity p of $\varrho(\zeta)$.

Finally we shall prove the last part of (P1). We write $\varrho(\zeta) = \sigma(\zeta) \cdot (\zeta-1)^p$. From (10) and (P3) it follows that

$$\lim_{h\to 0} h^p \sum_{n_p=0}^{n} \cdots \sum_{n_1=0}^{n_2} f(t_{n_1}^h) = \lim_{h\to 0} h^p \sum_{n_p=0}^{n} \cdots \sum_{n_1=0}^{n_2} \Psi_{n_1}(x_0^h, \ldots, x_{n+k}^h, h)$$
$$= \lim_{h\to 0} \left\{\sigma(E)\, x_{n+p} - \sigma(1) \sum_{j=0}^{p-1} \binom{n+j}{j} \{h^j X^{(j)}(a) + O(h^{j+1})\}\right\}.$$

$f(t)$ being continuous we may write

$$\int_a^t \cdots \int_a^{s_2} f(s_1)\, ds_1\, ds_2 \ldots ds_p = \sigma(1) \cdot \left\{X(t) - \sum_{j=0}^{p-1} \frac{(t-a)^j}{j!} X^{(j)}(a)\right\}$$

for $a<t<b$. By p times differentiation of both members of this equality we obtain $f(t) \equiv \sigma(1) \cdot X^{(p)}(t)$. In view of $\sigma(1) = \frac{\varrho^{(p)}(1)}{p!}$ and $X^{(p)} = F[X](t)$ this completes the proof of (P1). The theorem is thus proved.

III.3. A Convergence Criterion for the Case $p \geqq 1$

The methods (2) and (3), mentioned in chapter I, can be considered as methods determined by a sequence Ω_n as occurs in theorem 3. There are, however, very simple step-by-step methods for the numerical integration of differential equations of the type

$$X''(t) = f(t, X(t), X'(t))$$

which cannot be considered as methods of type (4) with a Ψ_n satisfying (7b). For instance, it is not possible to determine, by virtue of theorem 3, whether the method

$$x_{n+2} - 2x_{n+1} + x_n = h^2 \cdot f\left(t_{n+1}, x_{n+1} \frac{x_{n+2} - x_n}{2h}\right),$$

is 2-stable or not. We shall therefore state a theorem in this chapter which covers this and other cases.

Let $\varrho(\zeta)$ denote the same polynomial as in III.1, and let $k \geqq p \geqq 1$. Let

$$\Omega_n(v_0, \ldots, v_{n+k}, h) \equiv -\sum_{i=0}^{k-1} \alpha_i v_{n+i} + h^p \cdot \Psi_n(v_0, \ldots, v_{n+k}, h),$$

where the functions Ψ_n have to satisfy (5) as well as:

(7c) $$\Psi_n(v_0, \ldots, v_{n+k}, h) \equiv \Lambda_n\left(v_0, \ldots, v_{n+k}, \frac{\Delta v_0}{h}, \ldots, \frac{\Delta v_{n+k-1}}{h}, \ldots, \frac{\Delta^{p-1}}{h^{p-1}} v_0, \ldots, \frac{\Delta^{p-1}}{h^{p-1}} v_{n+r}, h\right),$$

$r = k - p + 1$ and the functions $\Lambda_n(u_0, \ldots, u_{q(n)}, h)$ satisfy

$$\|\Lambda_n(u_0, \ldots, u_{q(n)}, h) - \Lambda_n(\tilde{u}_0, \ldots, \tilde{u}_{q(n)}, h)\| \leqq \lambda \cdot \max_{0 \leqq i \leqq q(n)} \|u_i - \tilde{u}_i\|$$

for some constant λ, uniformly for $n \geqq 0$, $0 < h \leqq h_0$, $(n+k)h \leqq b - a$ $\left(q(n) = p\left(n + r + \frac{p+1}{2}\right) - 1\right)$.

It is clear that for $p \geqq 2$ (7c) is a weaker requirement than (7b).

For h_0 sufficiently small condition (6) is satisfied, the mapping $y = \psi(x) = \Omega_n(v_0, \ldots, v_{n+k-1}, x, h)$ then being a contracting mapping from V in V. This follows from

$$\begin{aligned}
&\|\psi(x) - \psi(\tilde{x})\| = \|\Omega_n(v_0, \ldots, v_{n+k-1}, x, h) - \Omega_n(v_0, \ldots, v_{n+k-1}, \tilde{x}, h)\| \\
&= h^p \cdot \|\Psi_n(v_0, \ldots, v_{n+k-1}, x, h) - \Psi_n(v_0, \ldots, v_{n+k-1}, \tilde{x}, h)\| \\
&\leqq h^p \lambda \cdot \max_{0 \leqq j \leqq p-1} \left\| \frac{\Delta^j}{h^j} v_{n+k-j} - \frac{\Delta^j}{h^j} \tilde{v}_{n+k-j} \right\| = h^p \lambda \cdot \max_{0 \leqq j \leqq p-1} \frac{1}{h^j} \|\Delta^j (v_{n+k-j} - \tilde{v}_{n+k-j})\| \\
&= h^p \lambda \cdot \max_{0 \leqq j \leqq p-1} \frac{1}{h^j} \|x - \tilde{x}\| \leqq h \lambda \cdot \|x - \tilde{x}\| \leqq h_0 \lambda \cdot \|x - \tilde{x}\|,
\end{aligned}$$

where $v_{n+k} = x$ and $\tilde{v}_{n+k} = \tilde{x}$, $\tilde{v}_i = v_i$ $(i = 0, 1, \ldots, n+k-1)$, and $h \leqq h_0$ with $0 < h_0 < \min(1, 1/\lambda)$ — x and $\tilde{x}$ are arbitrary vectors $\in V$—.

We then have

Theorem 4[1]. Let the roots of the polynomial $\varrho(\zeta)$ have modulus $\leqq 1$ and let the multiplicity of the roots $\zeta \neq 1$ of modulus 1 be at most 1 and let $\zeta = 1$ be a root of multiplicity $\leqq p$. Then

A. The sequence Ω_n is p-stable.

B. The propositions (P1), (P2) and (P3) are equivalent.

III.4. Step-by-step Methods for the Solution of Second Order Ordinary Differential Equations

We consider the second order initial-value problem

$$X''(t) = f(t, X(t), X'(t)),$$
$$X(a) = c_0, \qquad X'(a) = c_1$$

where $X(t)$ denotes an unknown function which maps $[a, b]$ into R_1. f is a given function from $[a, b] \times R_1 \times R_1$ to R_1, continuous in t and satisfying a Lipschitz condition $|f(t, x, y) - f(t, \tilde{x}, \tilde{y})| \leqq \lambda \cdot |x - \tilde{x}| + \lambda \cdot |y - \tilde{y}|$, uniformly for $a \leqq t \leqq b$, and all real numbers $x, \tilde{x}, y, \tilde{y}$. In the following we assume that the function f possesses continuous derivatives of a sufficiently high order.

[1] The proof of this theorem is rather lengthy and will not be published here. It is available on request at the Centraal Reken-Instituut of Leyden University.

Method 1.

$$y_{n+1}=y_n+h\cdot f\left(t_{n+1},\, x_{n+1},\, y_n+\frac{h}{2}\cdot f(t_{n+1},\, x_{n+1},\, y_n)\right),$$

$$x_{n+2}=x_{n+1}+h\cdot y_{n+1} \qquad (n=0,\,1,\,\ldots;\; h(n+2)\leqq b-a).$$

If we put

$$y_0=c_1+\frac{h}{2}\cdot f(t_0,\, c_0,\, c_1) \quad \text{and} \quad x_1=c_0+h\cdot y_0,$$

then the accumulated discretization errors satisfy $d_n=x_n-X(t_n)=O(h^2)$.

Method 2.

$$y_{n+1}=y_n+\frac{h}{3}\cdot\{K_0+K_1+K_2\},$$

where

$$K_0=f\left(t_{n+1}-\frac{h}{2},\, x_{n+1}-\frac{h}{2}\cdot y_n,\, y_n\right),$$

$$K_1=f\left(t_{n+1},\, x_{n+1}+\frac{h^2}{4}\cdot K_0,\, y_n+\frac{h}{2}\cdot K_0\right),$$

$$K_2=f\left(t_{n+1}+\frac{h}{2},\, x_{n+1}+\frac{h}{2}\cdot y_n,\, y_n+h\cdot K_1\right),$$

$$x_{n+2}=x_{n+1}+h\cdot y_{n+1} \qquad (n=0,\,1,\,\ldots;\; h(n+2)\leqq b-a).$$

If we put

$$y_0=c_1+\frac{h}{6}\cdot\{2f(t_0,\, c_0,\, c_1)+f(t_1,\, c_0+hc_1,\, c_1+h\cdot f(t_0,\, c_0,\, c_1))\}$$

and $x_1=c_0+h\cdot y_0$ the accumulated discretization errors satisfy $d_n=x_n-X(t_n)=O(h^3)$.

These two methods have been derived by a procedure analogous to the one developed by RUNGE and KUTTA. If we eliminate y_n from the formulas (by setting $y_n=\frac{x_{n+1}-x_n}{h}$) they take on the form (4) with a Ψ_n satisfying (7c). The above accumulated discretization error estimates are obtainable from the local error estimates by means of theorem 4.

Method 3.

$$y_{n+1}=y_n+\frac{h}{2}\cdot\left\{f\left(t_{n+1}+\frac{h}{\sqrt{6}},\, x\left(t_{n+1}+\frac{h}{\sqrt{6}}\right),\, x'\left(t_{n+1}+\frac{h}{\sqrt{6}}\right)\right)+\right.$$

$$\left.+f\left(t_{n+1}-\frac{h}{\sqrt{6}},\, x\left(t_{n+1}-\frac{h}{\sqrt{6}}\right),\, x'\left(t_{n+1}-\frac{h}{\sqrt{6}}\right)\right)\right\},$$

$$x_{n+2}=x_{n+1}+h\cdot y_{n+1}.$$

This formula is based on the relation

$$X(t_{n+2})-2X(t_{n+1})+X(t_n)=\frac{h^2}{2}\cdot\left\{X''\left(t_{n+1}+\frac{h}{\sqrt{6}}\right)+X''\left(t_{n+1}-\frac{h}{\sqrt{6}}\right)\right\}+O(h^6),$$

which holds for any sufficiently differentiable function $X(t)$. We interpolate and extrapolate by

$$x\left(t_{n+1}\pm\frac{h}{\sqrt{6}}\right)=x_{n+1}\pm\frac{1}{\sqrt{6}}\nabla x_{n+1}+\cdots,$$

and

$$x'\left(t_{n+1} \pm \frac{h}{\sqrt{6}}\right) = y_n + \left(\pm \frac{1}{\sqrt{6}} + \frac{1}{2}\right) \nabla y_n + \cdots,$$

neglecting backward differences of an order exceeding some integer q. With the substitution $y_n = \frac{x_{n+1} - x_n}{h}$ we again obtain a formula of type (4). From theorem 4 it follows that $d_n = x_n - X(t_n) = O(h^4)$ if $q \geqq 3$ and the starting values x_i, y_i satisfy

$$y_i - \frac{X(t_{i+1}) - X(t_i)}{h} = O(h^4), \quad x_{i+1} = x_i + h \cdot y_i, \quad x_0 = c_0 \quad (i = 0, 1, \ldots, q).$$

Remarks. 1. The above methods 1, 2 and 3 seem to be more appropriate for practical use than the (theoretically) equivalent algorithms of type (4). If round-off errors are present it may be of importance that the methods (4) with $p = 2$ are 2-stable, whereas the above algorithms are 1-stable.

2. Similar formulas as 1, 2 and 3 — if desired having a higher order — can be derived for the case that $V = R_m$ and $p = 2, 3, 4, \ldots$.

References

[1] LJUSTERNIK, L. A., u. W. I. SOBOLEW: Elemente der Funktionanalysis. Berlin: Akademie-Verlag 1955.

[2] DAHLQUIST, G.: Stability and error bounds in the numerical integration of ordinary differential equations. Uppsala: Almquist & Wiksells Boktryckeri 1959.

[3] HULL, T. E., and W. A. J. LUXEMBURG: Numerical methods and existence theorems for ordinary differential equations. Numerische Mathematik **2**, 30—41 (1960).

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

[5] CESCHINO, F., et J. KUNTZMANN: Problèmes différentiels de conditions initiales. Paris: Dunod 1963.

[6] HENRICI, P.: Error propagation for difference methods. New York: J. Wiley & Sons 1963.

[7] GRAGG, W. B., and H. J. STETTER: Generalized multistep predictor-corrector methods. Journal of the A.C.M. **11**, 188—209 (1964).

Centraal Reken-Instituut van de
Rijksuniversiteit te Leiden
Stationsweg 46