MATHEMATICS OF COMPUTATION
VOLUME 45, NUMBER 172
OCTOBER 1985, PAGES 377–392

Stepsize Restrictions for Stability of One-Step Methods in the Numerical Solution of Initial Value Problems

By M. N. Spijker

Abstract. This paper deals with the analysis of general one-step methods for the numerical solution of initial (-boundary) value problems for stiff ordinary and partial differential equations. Restrictions on the stepsize are derived that are necessary and sufficient for the rate of error growth in these methods to be of moderate size. These restrictions are related to disks contained in the stability region of the method, and the errors are measured with arbitrary norms (not necessarily generated by an inner product).

The theory is illustrated in the numerical solution of a diffusion-convection problem where the error growth is measured with the maximum norm.

1. Introduction.

1.1. *The Relevance of $A[D]$-Stability.* In 1963 Dahlquist [4] introduced the concept of A-stability in the analysis of numerical methods for solving initial value problems for ordinary differential equations. This concept is based on the error propagation in a numerical method when it is applied to the simple scalar test equation $dU(t)/dt = \lambda U(t)$ with $\lambda \in \mathbf{C}$. Between 1976–1979 the criterion of A-stability proved to be of great relevance in a rigorous analysis of error propagation in methods for solving problems that are essentially more general than the scalar test problem (cf. [2], [8], [1], [13], [5]). A priori estimates were obtained of error propagation in general one-step methods applied to arbitrarily stiff linear systems of ordinary differential equations and linear partial differential equations (cf. [2], [8], [1]).

After 1963, the concept of a stability region S in the complex plane (cf. Subsection 1.2) was studied, which led to several weaker versions of A-stability (cf., e.g., [14]). One of these versions, $A(\alpha)$-stability, can also be applied successfully in a rigorous analysis analogous to the above (see [2], [20]).

In the present paper, we focus on some other weaker version of A-stability namely $A[D]$-stability (cf. [14], [9], [10] and Subsection 1.2). Here D denotes a disk in the complex plane (bounded by a circle passing through the origin) and it is required that $D \subset S$. In contrast to the requirements of A-stability and $A(\alpha)$-stability, the requirement $D \subset S$ can be fulfilled by explicit methods.

Under the assumption of $A[D]$-stability, we give in this paper an analysis of error propagation in a framework that has similarity to those mentioned above (cf. [2], [8],

Received March 26, 1984; revised February 18, 1985.
1980 *Mathematics Subject Classification.* Primary 65L20, 65M10.

0025-5718/85 $1.00 + $.25 per page

[1], [20]). We shall arrive at conditions on the stepsize that are (necessary and) sufficient for the rate of error growth, in a given method, to be of moderate size. These conditions will be illustrated in the solution of a diffusion-convection problem.

Further, we shall relate the concept of $A[D]$-stability to some other known stability concepts (spectral stability condition, von Neumann stability condition, weak and strong stability, contractivity).

We note that, under the assumption of $A[D]$-stability, already rigorous results on error propagation were obtained for linear k-step methods by Nevanlinna [13] and Dahlquist [5]. These results are more general than ours in that the differential equations considered were allowed to be nonlinear, and k was allowed to be > 1. On the other hand, the norms we shall deal with need not be generated by an inner product, and the number of stages in the numerical schemes that we shall consider is not restricted to 1—as was the case in [13], [5].

1.2. *Notations and Definitions.* With ϕ we denote a given rational function, $\phi(\zeta) = P_1(\zeta)/P_0(\zeta)$, where P_1, P_0 are polynomials with real coefficients, no common zero and $P_1(0) = P_0(0) = 1$. For any $s \times s$ matrix T we say that $\phi(T)$ *exists*, and we write $\phi(T) = P_1(T)P_0(T)^{-1}$ whenever the matrix $P_0(T)$ is regular.

We shall be concerned with the rate of growth of vectors $u_n \in \mathbf{R}^s$ that are computed from the recurrence relation

$$u_n = \phi(hA)u_{n-1} \qquad (n = 1,2,3,\ldots). \tag{1.1}$$

Here $h > 0$ is the *stepsize*, and A denotes a given real $s \times s$ matrix.

In many applications u_n stands for an approximation to $U(nh)$. Here $U(t) \in \mathbf{R}^s$ denotes the solution to a given initial value problem for a system of ordinary differential equations

$$\frac{d}{dt}U(t) = AU(t) \qquad (t \geqslant 0),\ U(0) = u_0. \tag{1.2}$$

Many known numerical methods for solving ordinary differential equations, such as *Runge-Kutta* methods and *Rosenbrock* methods, result, when applied to (1.2), in a procedure of type (1.1).

Further, many numerical schemes for solving *initial-boundary value problems* in *partial differential equations* can be written in the form (1.1) (see, e.g., Section 7, and [17], [8], [19], [23], [26]). Here s will often stand for the (large) number of gridpoints involved at any fixed time level $t = nh$.

In this paper, the rate of growth of the u_n will be related to the size of h and of quantities r and R that we are now going to define.

The *stability region* S of ϕ is defined by

$$S = \{\zeta \mid \zeta \in \mathbf{C} \text{ and } \phi \text{ is regular at } \zeta \text{ with } |\phi(\zeta)| \leqslant 1\}.$$

The disk $D = \mathscr{D}(\xi, \rho)$ in the complex plain is defined by

$$D = \{\zeta \mid \zeta \in \mathbf{C} \text{ and } |\zeta - \xi| \leqslant \rho\}$$

for $\xi \in \mathbf{C}$, $0 \leqslant \rho < \infty$. Further, the procedure (1.1) is called $A[D]$-*stable* if $D \subset S$.

We define the *stability radius* $r \in [0, \infty]$ of ϕ by

$$r = \sup\{\rho \mid 0 \leqslant \rho < \infty \text{ and } \mathscr{D}(-\rho, \rho) \subset S\}.$$

Finally, $R \in [0, \infty]$ is defined by

$$R = \sup\{\rho \mid \rho = 0, \text{ or } 0 < \rho < \infty \text{ and } \phi \text{ absolutely monotonic on } [-\rho, 0]\}$$

(a function is called absolutely monotonic on an interval if the values of the function and of all its derivatives are finite and $\geqslant 0$ on that interval). For reasons becoming evident in the subsequent we call R the *contractivity radius* of ϕ.

In this paper, we deal mainly with matrices A such that (1.2) is *stable* in the sense that the solution to (1.2) remains bounded as $t \to \infty$ (for each starting vector $u_0 \in \mathbf{R}^s$). We aim at transparent conditions on the stepsize $h > 0$ that guarantee an analogous stability-behavior for the vectors u_n satisfying (1.1). Similarly to the frameworks in [2], [8], [1], [20], and motivated by the applications to partial differential equations mentioned above, we shall focus on stability results for (1.1) which hold uniformly with respect to the dimension s.

In most applications, it is the growth of the difference between two solutions, say u_n and $\tilde{u}_n$, to (1.1) which is significant. For instance, $\tilde{u}_n$ may stand for the numerical approximation obtained in the presence of a rounding error $v_0 = \tilde{u}_0 - u_0$. Since the resulting error $v_n = \tilde{u}_n - u_n$ then also satisfies (1.1), our stability results on the growth of u_n will also be relevant to the *growth of errors* v_n in the application of method (1.1).

1.3. *Organization of the Paper.* In Section 2 we introduce the classes of matrices A for which we shall analyze the stability of process (1.1). We prove a theorem on these classes that was already applied in [22].

In Section 3 we deal with a recurrence relation for u_n that can be viewed as originating from an application of method (1.1) to a simple (but nonscalar) testproblem of type (1.2). This recurrence relation enables us to relate in a natural way $A[D]$-stability to a number of well-known other stability concepts.

In Theorem 4.1 of Section 4, we present a condition (in terms of the radius r) on the stepsize $h > 0$ which is necessary and sufficient for (a weak version of) stability of the general process (1.1). At the end of Section 4, we also present stability results for the case where h in (1.1) is replaced by a variable stepsize $h_n > 0$ ($n = 1, 2, 3, \ldots$).

Theorem 5.1 in Section 5 is concerned with an analogous condition (in terms of the radius R) on the stepsize $h > 0$ which is necessary and sufficient for the stronger version of stability called contractivity.

In Section 6, we discuss modifications of Theorems 4.1, 5.1.

Section 7 contains an illustration of the material of Sections 2, 4, 5 in the numerical solution of a diffusion-convection problem.

We note that the tools used in the proofs in Sections 4, 5 mainly consist in power series expansions for matrix-valued functions and Parseval's formula for complex functions. Our arguments are therefore more elementary than those used in the proofs of the important paper by Brenner and Thomée [1] referred to in Subsection 1.1.

2. Two Classes of Matrices A. Let $\mathbf{R}^s$ denote the s-dimensional real vector space equipped with an arbitrary norm $|\cdot|$. Let $\omega, \tau, \alpha \in \mathbf{R}$ with $\tau > 0$.

Following [23], we denote by $\mathscr{L}(\mathbf{R}^s, \omega, \tau)$, the collection of all real $s \times s$ matrices A such that

$$\|A + \tau^{-1}\| \leqslant \tau^{-1} + \omega. \tag{2.1}$$

Here $\|\cdot\|$ denotes the (lub-) matrix norm induced by $|\cdot|$, i.e.,

$$\|T\| = \sup\{|Tx|: x \in \mathbf{R}^s \text{ with } |x| = 1\}$$

for any $s \times s$ matrix T.

Further, we denote by $\mathscr{B}(\mathbf{R}^s, \omega, \alpha)$, the collection of all real $s \times s$ matrices A such that

$$\|A\| \leqslant \alpha, \qquad \mu[A] \leqslant \omega. \tag{2.2}$$

Here $\mu[\cdot]$ denotes the logarithmic norm induced by $|\cdot|$, i.e.,

$$\mu[A] = \lim_{t \to 0+} t^{-1}(\|I + tA\| - 1)$$

(cf. [3]; I denotes the identity matrix).

The inequality (2.1) implies that the spectrum of the matrix A is contained in the disk $\mathscr{D}(-\tau^{-1}, \tau^{-1} + \omega)$. For normal A and the Euclidean norm this spectral property is even equivalent to (2.1). In [10], [23] property (2.1) was called a *circle condition*. For further interpretations of the class $\mathscr{L}(\mathbf{R}^s, \omega, \tau)$, see [10], [23].

The inequalities in (2.2) only involve the (logarithmic) norm of A itself so that the intuitive meaning of the class $\mathscr{B}(\mathbf{R}^s, \omega, \alpha)$ is even more easily grasped than that of $\mathscr{L}(\mathbf{R}^s, \omega, \tau)$. Therefore, the following Theorem 2.1, relating $\mathscr{L}(\mathbf{R}^s, \omega, \tau)$ to $\mathscr{B}(\mathbf{R}^s, \omega, \alpha)$ is of importance.

We refer to Section 7 for an example involving the classes $\mathscr{L}(\mathbf{R}^s, \omega, \tau)$, $\mathscr{B}(\mathbf{R}^s, \omega, \alpha)$.

With $\mathbf{R}^s_p$, we shall denote $\mathbf{R}^s$ when equipped with the pth Hölder norm $|x|_p = (\sum_j |\xi_j|^p)^{1/p}$ (when $1 \leqslant p < \infty$), $|x|_p = \max_j |\xi_j|$ (when $p = \infty$) for $x = (\xi_1, \xi_2, \ldots, \xi_s)^T \in \mathbf{R}^s$. The corresponding matrix norm is denoted by $\|\cdot\|_p$.

THEOREM 2.1. (Relations between $\mathscr{L}(\mathbf{R}^s, \omega, \tau)$ and $\mathscr{B}(\mathbf{R}^s, \omega, \alpha)$.) *Let $\omega, \alpha \in \mathbf{R}$ with $\alpha > \omega$. Then*

(i) $\mathscr{B}(\mathbf{R}^s, \omega, \alpha) \supset \mathscr{L}(\mathbf{R}^s, \omega, 2(\alpha - \omega)^{-1})$ *for each norm $|\cdot|$ in $\mathbf{R}^s$;*

(ii) $\mathscr{B}(\mathbf{R}^s, \omega, \alpha) = \mathscr{L}(\mathbf{R}^s, \omega, 2(\alpha - \omega)^{-1})$ *when the norm in $\mathbf{R}^s$ is given by $|x| = |Qx|_p$ where Q is a regular $s \times s$ matrix and $p = 1$ or $p = \infty$.*

Proof. 1. Defining $\tau = 2(\alpha - \omega)^{-1}$, we have for any $A \in \mathscr{L}(\mathbf{R}^s, \omega, \tau)$, the inequalities

$$\|A\| - \tau^{-1} \leqslant \|A + \tau^{-1}\| \leqslant \omega + \tau^{-1},$$

and therefore,

$$\|A\| \leqslant \omega + 2\tau^{-1} = \alpha.$$

This proves (i) since $\mu[A] \leqslant \tau^{-1}(\|I + \tau A\| - 1) \leqslant \tau^{-1}(\tau(\omega + \tau^{-1}) - 1) = \omega$ (cf. [3]).

2. In order to prove (ii) we first assume $p = \infty$, $Q = I$ (the identity). Let $A \in \mathscr{B}(\mathbf{R}^s_\infty, \omega, \alpha)$, $A = (\alpha_{ij})$.

For $x = (\xi_1, \ldots, \xi_s)^T$, $y = (\eta_1, \ldots, \eta_s)^T \in \mathbf{R}^s_\infty$ with $|x|_\infty = 1$, $y = (I + \tau A)x$ we have

$$|\eta_i| \leqslant |1 + \tau\alpha_{ii}| \cdot |\xi_i| + \tau \sum_{j \neq i} |\alpha_{ij}| \cdot |\xi_j| \qquad (1 \leqslant i \leqslant s).$$

Hence,

$$|\eta_i| \leqslant \theta_i(1 + \tau\alpha_{ii}) + (1 - \theta_i)(-1 - \tau\alpha_{ii}) + \tau \sum_{j \neq i} |\alpha_{ij}| \qquad (1 \leqslant i \leqslant s)$$

with $\theta_i = 0$ (when $1 + \tau\alpha_{ii} < 0$) or $\theta_i = 1$ (when $1 + \tau\alpha_{ii} \geqslant 0$). Consequently (cf. [3]),

$$\begin{aligned} |\eta_i| &\leqslant \theta_i\left\{1 + \tau\alpha_{ii} + \tau \sum_{j \neq i} |\alpha_{ij}|\right\} + (1 - \theta_i)\left\{-1 - \tau\alpha_{ii} + \tau \sum_{j \neq i} |\alpha_{ij}|\right\} \\ &\leqslant \theta_i(1 + \tau\omega) + (1 - \theta_i)(-1 + \tau\|A\|) \\ &\leqslant \theta_i(1 + \tau\omega) + (1 - \theta_i)\left(-1 + \tau \cdot (\omega + 2\tau^{-1})\right) = 1 + \tau\omega. \end{aligned}$$

Therefore, $|(I + \tau A)x|_\infty = |y|_\infty \leqslant 1 + \tau\omega$, which proves $A \in \mathscr{L}(\mathbf{R}^s_\infty, \omega, \tau)$. Hence, $\mathscr{B}(\mathbf{R}^s_\infty, \omega, \alpha) = \mathscr{L}(\mathbf{R}^s_\infty, \omega, \tau)$.

3. We note that for any $s \times s$ matrix $A = (\alpha_{ij})$ we have $\|A\|_1 = \|A^T\|_\infty$, $\mu_1[A] = \mu_\infty[A^T]$. From $\mathscr{B}(\mathbf{R}^s_\infty, \omega, \alpha) = \mathscr{L}(\mathbf{R}^s_\infty, \omega, \tau)$, we thus conclude that also $\mathscr{B}(\mathbf{R}^s_1, \omega, \alpha) = \mathscr{L}(\mathbf{R}^s_1, \omega, \tau)$.

4. Let $|\cdot|$ be any norm in $\mathbf{R}^s$ for which we have $\mathscr{B}(\mathbf{R}^s, \omega, \alpha) = \mathscr{L}(\mathbf{R}^s, \omega, \tau)$. Let Q be a regular $s \times s$ matrix and put $|x|^* = |Qx|$ (for $x \in \mathbf{R}^s$). The matrix norm and logarithmic norm induced by $|\cdot|^*$ are denoted by $\|\cdot\|^*$, $\mu^*[\cdot]$.

For any $s \times s$ matrix A we have

$$\mu^*[A] = \mu\left[QAQ^{-1}\right], \qquad \|A\|^* = \left\|QAQ^{-1}\right\| \quad \text{and} \quad \|A + \tau^{-1}\|^* = \left\|QAQ^{-1} + \tau^{-1}\right\|.$$

From these relations it easily follows that also $\mathscr{B}(\mathbf{R}^s, \omega, \alpha) = \mathscr{L}(\mathbf{R}^s, \omega, \tau)$ with respect to the norm $|\cdot|^*$.

In view of the parts 2, 3 the proof is now complete. □

Remarks. 1. By part (i) of the above theorem, we have for each $A \in \mathscr{L}(\mathbf{R}^s, \omega, \tau)$, the inequality $\mu[A] \leqslant \omega$, and therefore (cf. [3]),

$$|U(t)| \leqslant \exp(\omega t)|u_0| \qquad (\text{for } t \geqslant 0)$$

for the solution to (1.2). If $\omega \leqslant 0$, the solution to (1.2) is thus bounded uniformly for $0 \leqslant t < \infty$.

2. One easily verifies that statement (ii) in the above theorem is not valid in case $Q = I$ and $p = 2$.

3. Statement (ii) with $Q = I$, $p = \infty$, is a basic means for proving the expression for the stepsize threshold that was presented, without full proof, in [22].

3. Relating $A[D]$-Stability to Other Stability Concepts. In order to obtain insight into the possible rate of growth of vectors u_n computed from (1.1) when A belongs to one of the classes defined in Section 2, it is useful to consider the case

$$u_n = \phi(A_s)u_{n-1} \qquad (n = 1, 2, 3, \ldots). \tag{3.1}$$

Here A_s denotes a bi-diagonal $s \times s$ matrix of the form

$$A_s = \begin{pmatrix} \xi & \rho & 0 & & \cdots & 0 \\ 0 & \xi & \rho & \ddots & & \\ & & & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \ddots & 0 \\ & & & & & \rho \\ 0 & & \cdots & & 0 & \xi \end{pmatrix} \tag{3.2}$$

with ξ, $\rho \in \mathbf{R}$, $\rho \geqslant 0$. Moreover, the subsequent stability analysis will relate $A[D]$-stability in a natural fashion to a number of other stability concepts.

THEOREM 3.1. (Necessary condition for weak stability of (3.1).) *Let $\phi(\zeta)$ be regular at $\zeta = \xi$, and let $1 \leqslant p \leqslant \infty$. Assume there exist γ, $q < \infty$, such that*

$$|u_n|_p \leqslant \gamma n^q \cdot |u_0|_p \quad (\textit{for all } n \geqslant 1, s \geqslant 1, u_0 \in \mathbf{R}^s) \tag{3.3}$$

whenever u_n satisfies (3.1). *Then,*

$$\mathscr{D}(\xi, \rho) \subset S. \tag{3.4}$$

Proof. Apply Lemma 3.2, to be given at the end of this section. □

The above theorem, constituting the main result of this section, shows that there is *weak stability* (in the sense of (3.3)) only if there is $A[D]$-*stability* with disk $D = \mathscr{D}(\xi, \rho)$. This contrasts with the familiar *spectral condition for stability*, which for (1.1) requires that the spectrum of the matrix hA is contained in (the interior of) S. The latter condition reduces for the case of (3.1) to the requirement $|\phi(\xi)| < 1$. This is, generally, a much weaker requirement than (3.4). Similar shortcomings of the spectral condition for stability (uniformly with respect to s) were stated, e.g. in [16], [17, p. 152], [11, pp. 258–261], [12], [19], [26].

We note that Theorem 3.1 can also be proved, slightly differently, by using the *Godunov-Ryabenkii criterion for stability* (cf. [17, p. 153]). We have preferred the proof via Lemma 3.2 since it is more direct and shorter.

It is interesting, and not surprising in view of the Godunov-Ryabenkii criterion, that condition (3.4) can be arrived at in a heuristic fashion by putting $s = \infty$ in (3.1). The components $u_{n,j}$ of the vector u_n would then satisfy

$$u_{n,j} = \sum_{k=0}^{\infty} \frac{\phi^{(k)}(\xi)}{k!} \rho^k u_{n-1,j+k} \qquad (j \in \mathbf{Z}, n \geqslant 1),$$

and the *von Neumann condition for stability* (cf. [17]) applied to this difference scheme, reads

$$\left| \sum_{k=0}^{\infty} \frac{\phi^{(k)}(\xi)}{k!} \rho^k e^{ikt} \right| \leqslant 1 \qquad (\text{for } -\infty < t < \infty). \tag{3.5}$$

Clearly (3.5) is equivalent to (3.4).

We conclude this section with

LEMMA 3.2. *Let the rational function $\phi(\zeta)$ be regular at $\zeta = \xi$, and let $1 \leqslant p \leqslant \infty$. Then*

(i) $\lim_{s \to \infty} \|[\phi(A_s)]^n\|_p = \infty$ (*for each* $n \geqslant 1$) *when* $\phi(\zeta)$ *has a singularity in* $\mathscr{D}(\xi, \rho)$;

(ii) $2 \cdot \|[\phi(A_s)]^n\|_p \geqslant M^n - g(s, n)$ (*for each* $n \geqslant 1$) *with* $\lim_{s \to \infty} g(s, n) = 0$, $M = \max\{|\phi(\zeta)|: \zeta \in \mathscr{D}(\xi, \rho)\}$ *when* $\phi(\zeta)$ *is regular on* $\mathscr{D}(\xi, \rho)$.

Proof. With E we denote the $s \times s$ matrix all of whose entries ε_{ij} vanish with the exception of $\varepsilon_{i,i+1} = 1$ $(1 \leqslant i \leqslant s-1)$, and we define $\gamma_j = (j!)^{-1}\phi^{(j)}(\xi) \cdot \rho^j$ (for $j = 0, 1, 2, \ldots$).

For $n \geqslant 1$, we have $\phi(A_s)^n = \Sigma \gamma_{j_1}\gamma_{j_2} \cdot \cdots \cdot \gamma_{j_n} \cdot E^{j_1 + j_2 + \cdots + j_n}$ where the summation is for all integers $j_1 \geqslant 0, j_2 \geqslant 0, \ldots, j_n \geqslant 0$ with $j_1 + j_2 + \cdots + j_n \leqslant s - 1$ (cf. [6]).

Denoting also for complex vectors $y \in \mathbf{C}^s$ the pth Hölder norm by $|y|_p$, we have for $1 \leqslant p < \infty$ and for any $y = (\eta_1, \eta_2, \ldots, \eta_s)^T \in \mathbf{C}^s$ with $|y|_p \leqslant 1$, the inequality

$$2 \cdot \|\phi(A_s)^n\|_p \geqslant |\phi(A_s)^n y|_p = \left\{ \sum_{j=1}^{s} \left| \sum_{k=0}^{s-j} \beta_k \eta_{j+k} \right|^p \right\}^{1/p},$$

where

$$\beta_k = \sum_{j_1 + \cdots + j_n = k} \gamma_{j_1} \cdot \gamma_{j_2} \cdot \cdots \cdot \gamma_{j_n} \qquad (k \geqslant 0).$$

Let $\theta \in (0, 1]$ be such that $\phi(\zeta)$ is regular on $\mathscr{D}(\xi, \theta\rho)$. Choosing $y = (\eta_1, \eta_2, \ldots, \eta_s)^T \in \mathbf{C}^s$ with $\eta_k = \eta_0 \theta^k \exp(ikt)$ (for $1 \leqslant k \leqslant s$), $\eta_0 = s^{-1/p}$ (when $\theta = 1$), $\eta_0 = \theta^{-1}(1 - \theta^p)^{1/p}$ (when $0 < \theta < 1$), $t \in \mathbf{R}$, we thus obtain

$$2 \cdot \|\phi(A_s)^n\|_p \geqslant \eta_0 \left\{ \sum_{m=0}^{s-1} \theta^{p(s-m)} \left| \sum_{k=0}^{m} \beta_k \theta^k \exp(ikt) \right|^p \right\}^{1/p}.$$

Writing $\zeta = \xi + \theta\rho \exp(it)$, we have

$$\lim_{m \to \infty} \sum_{k=0}^{m} \beta_k \theta^k \exp(ikt) = \phi(\zeta)^n.$$

From this relation, we can conclude that, for $1 \leqslant p < \infty$,

$$2 \cdot \|\phi(A_s)^n\|_p \geqslant |\phi(\zeta)^n| - g(s, n, \theta)$$

with $\lim_{s \to \infty} g(s, n, \theta) = 0$. Since $\|\phi(A_s)^n\|_\infty = \|\phi(A_s)^n\|_1$, this result is also valid for $p = \infty$.

Statement (i) of the lemma follows easily by varying t and θ appropriately, and statement (ii) follows by choosing $\theta = 1$ and t such that $|\phi(\xi + \rho \cdot \exp(it))| = M$. □

Remark. Although it is not essential in the stability considerations concerning process (3.1), it is worth mentioning that the factor 2 occurring in statement (ii) (Lemma 3.2) can be omitted. This follows from a proof communicated to us recently by M. Crouzeix.

4. On the Relevance of the Radius r to the Stability of (1.1).

4.1. *A Necessary and Sufficient Condition for Stability of* (1.1). In this section, we give a stability analysis of (1.1) based on $A[D]$-stability with $D = \mathscr{D}(-r, r)$, where r is the stability radius defined in Subsection 1.2.

In the present subsection, we state the main result of this analysis in Theorem 4.1. The proof of the theorem will be based on Theorem 3.1 and on two corollaries to the technical Lemma 4.2. This lemma will be presented in Subsection 4.2 while its corollaries are given in Subsection 4.3. The corollaries in the latter subsection also contain some conclusions that have not been incorporated into the main Theorem 4.1, among other things a conclusion on the variable stepsize version (4.1) of (1.1).

The subsequent Theorem 4.1 provides a restriction on the stepsize $h > 0$ (see (s1)) which is sufficient for weak stability of the process (1.1) (uniformly with respect to s; see (s3)). Further, the theorem implies that restriction (s1) is also a necessary condition, already for a weaker version (namely (s2)) of the weak stability property (s3). Finally, the theorem shows that, under some additional condition on A, the

stepsize restriction (s1) is even sufficient for (strong) stability of (1.1) (uniformly with respect to s; see (s4)).

An illustration of Theorem 4.1 is presented in Section 7.

THEOREM 4.1. *Let h, τ, p be given with $0 < h < \infty$, $0 < \tau < \infty$, $1 \leqslant p \leqslant \infty$. Then the following four statements* (s1)–(s4) *are equivalent to each other.*

(s1) $0 < h \leqslant r\tau$;

(s2) $\phi(hA)$ *exists and* (1.1) *implies* $|u_n|_p \leqslant \gamma n^q |u_0|_p$ (*for each* $n \geqslant 1$, $s \geqslant 1$, $u_0 \in \mathbf{R}^s$, $A = A_s$ (*see* (3.2)) *with* $-\xi = \rho = \tau^{-1}$). *Here* γ, q *are constants independent of* n, s, u_0;

(s3) $\phi(hA)$ *exists and* (1.1) *implies* $|u_n| \leqslant \gamma n^{1/2} |u_0|$ (*for each* $n \geqslant 1$, $s \geqslant 1$, $u_0 \in \mathbf{R}^s$, $A \in \mathscr{L}(\mathbf{R}^s, 0, \tau)$ *and each norm* $|\cdot|$ *in* $\mathbf{R}^s$). *Here* γ *is a constant independent of* n, s, u_0, A, $|\cdot|$;

(s4) $\phi(hA)$ *exists and* (1.1) *implies* $|u_n| \leqslant [(-\omega\tau)(2 + \omega\tau)]^{-1/2} |u_0|$ (*for each* $n \geqslant 1$, $s \geqslant 1$, $u_0 \in \mathbf{R}^s$, $\omega < 0$, $A \in \mathscr{L}(\mathbf{R}^s, \omega, \tau)$ *and each norm* $|\cdot|$ *in* $\mathbf{R}^s$).

Proof. 1. Assume (s1). Then (s3) follows easily from Corollary 4.3 stated in the next Subsection 4.3. We only have to apply this corollary with $h_n \equiv h$, $\omega = 0$ and $\rho = r$ (when $r < \infty$) or $\rho \in [h\tau^{-1}, \infty)$ (when $r = \infty$).

Further, (s4) follows immediately by applying Corollary 4.4.

2. Since $A = A_s$ with $-\xi = \rho = \tau^{-1}$ (see (3.2)) belongs to the class $\mathscr{L}(\mathbf{R}^s_p, 0, \tau)$, one easily sees that (s3) implies (s2) with $q = \frac{1}{2}$.

Applying Theorem 3.1 (with ξ, ρ replaced by $h\xi$, $h\rho$), we see that statement (s2) implies

$$\mathscr{D}(-h\tau^{-1}, h\tau^{-1}) = \mathscr{D}(h\xi, h\rho) \subset S.$$

Hence, (s2) implies (s1).

3. It remains to show that (s4) implies (s1).

Let $\omega < 0$, $1 + \omega\tau \geqslant 0$ and $A = A_s$ with $\xi = -\tau^{-1}$, $\rho = \tau^{-1} + \omega$. Since $A \in \mathscr{L}(\mathbf{R}^s_p, \omega, \tau)$, statement (s4) implies that $\phi(\zeta)$ is regular at $\zeta = -h\tau^{-1}$, and that

$$\left\| \left[\phi(hA_s) \right]^n \right\|_p \leqslant \gamma \quad (\text{for all } s \geqslant 1, n \geqslant 1)$$

with $\gamma = [(-\omega\tau)(2 + \omega\tau)]^{-1/2}$. An application of Theorem 3.1 (with $q = 0$, $\xi = -h\tau^{-1}$, $\rho = h(\tau^{-1} + \omega)$) shows that

$$\mathscr{D}(-h\tau^{-1}, h\tau^{-1} + h\omega) \subset S.$$

Since S is closed, it follows, by letting $\omega \to 0-$, that

$$\mathscr{D}(-h\tau^{-1}, h\tau^{-1}) \subset S.$$

Hence, $h\tau^{-1} \leqslant r$ which implies (s1), and completes the proof of the theorem. □

Remarks. 1. In the majorization (s4) the factor $n^{1/2}$, which is present in (s3), has disappeared. This is compensated by the factor $\tau^{-1/2}$ in (s4), which satisfies

$$\tau^{-1/2} \leqslant r^{1/2} h^{-1/2} = (r/t)^{1/2} \cdot n^{1/2} \quad \text{with } t = nh.$$

2. The constant γ in (s3) can be chosen, to some degree, independently of the parameters $h \in (0, \infty)$ and $\tau \in (0, \infty)$.

Assume (s1) and $r < \infty$. Then (s3) holds with γ only depending on ϕ.

Assume (s1) and $r = \infty$. Let $h\tau^{-1} \leqslant \rho < \infty$. Then (s3) holds with γ only depending on ϕ and on ρ.

These two conclusions easily follow from Corollary 4.3.

4.2. *Formulation and Proof of Lemma* 4.2. In this subsection we deal with a slightly generalized version of procedure (1.1). We consider the recurrence relation

$$u_n = \phi(h_n A)u_{n-1} \qquad (n = 1, 2, 3, \ldots) \tag{4.1}$$

with arbitrary stepsizes

$$h_n > 0 \qquad (n = 1, 2, \ldots).$$

The subsequent lemma is a convenient means for obtaining upper bounds for $|u_n|$ under various conditions on A and h_n.

We introduce some notations needed in the formulation of the lemma.

For $\xi \in \mathbf{C}$, $0 \leqslant \rho < \infty$, we define $\mathscr{M}(\xi, \rho) = \infty$ when ϕ has a singularity in $\mathscr{D}(\xi, \rho)$, and $\mathscr{M}(\xi, \rho) = \max\{|\phi(\zeta)|: \zeta \in \mathscr{D}(\xi, \rho)\}$ when ϕ is regular on $\mathscr{D}(\xi, \rho)$.

We assume $\tau, \omega, \theta \in \mathbf{R}$ with

$$\tau > 0, \quad 1 + \omega\tau \geqslant 0, \quad \theta > 0, \tag{4.2.a}$$

and we define λ_n, μ_n (for $n = 1, 2, 3, \ldots$) by

$$\lambda_n = \mathscr{M}\left(-h_n\tau^{-1}, (1 + \omega\tau)h_n\tau^{-1}\right), \tag{4.2.b}$$

$$\mu_n = \mathscr{M}\left(-h_n\tau^{-1}, \theta^{-1}h_n\tau^{-1}\right). \tag{4.2.c}$$

LEMMA 4.2. *Assume* (4.2) *and* $\lambda_n < \infty$, $\mu_n < \infty$ $(n = 1, 2, 3, \ldots)$. *Let* m *be an arbitrary integer* $\geqslant 0$. *Then, for each* $n \geqslant 1$, $s \geqslant 1$, $A \in \mathscr{L}(\mathbf{R}^s, \omega, \tau)$ *and each norm* $|\cdot|$ *in* $\mathbf{R}^s$, *the following statements* (4.3), (4.4) *are valid.*

$$\phi(h_n A) \quad \textit{exists}, \tag{4.3}$$

$$\begin{aligned} &\|\phi(h_n A) \cdot \cdots \cdot \phi(h_2 A) \cdot \phi(h_1 A)\| \\ &\leqslant \left(\prod_{j=1}^{n} \lambda_j\right)\left\{\sum_{k=0}^{m-1} (1 + \omega\tau)^{-2k}\|(I + \tau A)^k\|^2\right\}^{1/2} \\ &\quad + \left(\prod_{j=1}^{n} \mu_j\right)\left\{\sum_{k=m}^{\infty} \theta^{2k}\|(I + \tau A)^k\|^2\right\}^{1/2}. \end{aligned} \tag{4.4}$$

The first sum in the right-hand member of (4.4) *stands for zero when* $m = 0$, *and it stands for* 1 *when* $m > 0$, $(1 + \omega\tau) = 0$.

Proof. Defining $\xi_n = -h_n\tau^{-1}$ we have, in view of (2.1), for any $A \in \mathscr{L}(\mathbf{R}^s, \omega, \tau)$, the representation $h_n A = \xi_n - \xi_n B$ where the $s \times s$ matrix B satisfies

$$B = I + \tau A, \qquad \|B\| \leqslant 1 + \omega\tau.$$

From (4.2.b), $\lambda_n < \infty$, we see that ϕ is regular on $\mathscr{D}(\xi_n, |\xi_n(1 + \omega\tau)|)$. Therefore, using the above representation for $h_n A$, it follows (cf. [6]) that (4.3) holds.

From (4.2.b), (4.2.c), $\lambda_j < \infty$, $\mu_j < \infty$ (for $1 \leqslant j \leqslant n$), we see that the rational function f defined by

$$f(\zeta) = \phi(\xi_n(1 - \zeta)) \cdot \cdots \phi(\xi_2(1 - \zeta)) \cdot \phi(\xi_1(1 - \zeta))$$

is regular for all $\zeta \in \mathbf{C}$ with $|\zeta| \leqslant \max[(1+\omega\tau), \theta^{-1}]$. Denoting the coefficients of the Maclaurin expansion of $f(\zeta)$ by γ_k, we thus have

$$f(\zeta) = \gamma_0 + \gamma_1\zeta + \gamma_2\zeta^2 + \cdots \quad \big(\text{for } |\zeta| \leqslant \max\big[(1+\omega\tau), \theta^{-1}\big]\big). \tag{4.5}$$

Consequently (cf. [6]),

$$\phi(h_nA)\cdot \cdots \cdot\phi(h_2A)\cdot\phi(h_1A) = f(B) = \gamma_0 + \gamma_1 B + \gamma_2 B^2 + \cdots. \tag{4.6}$$

From (4.6) we obtain, by two applications of Schwarz' inequality, the upper bound

$$\|f(B)\| \leqslant \left\{\sum_{k=0}^{m-1}\big(\gamma_k(1+\omega\tau)^k\big)^2\right\}^{1/2}\left\{\sum_{k=0}^{m-1}(1+\omega\tau)^{-2k}\|B^k\|^2\right\}^{1/2}$$
$$+\left\{\sum_{k=m}^{\infty}\big(\gamma_k\theta^{-k}\big)^2\right\}^{1/2}\left\{\sum_{k=m}^{\infty}\theta^{2k}\|B^k\|^2\right\}^{1/2}.$$

For any σ with $0 < \sigma \leqslant \max[1+\omega\tau, \theta^{-1}]$, we obtain from (4.5), by applying Parseval's formula, the relation

$$\sum_{k=0}^{\infty}\big(\gamma_k\sigma^k\big)^2 = (2\pi)^{-1}\int_0^{2\pi}\big|f(\sigma e^{it})\big|^2\,dt.$$

In view of the definition of f, it thus follows that

$$\sum_{k=0}^{\infty}\big(\gamma_k\sigma^k\big)^2 \leqslant \prod_{j=1}^{n}\big[\mathscr{M}\big(\xi_j, |\xi_j\sigma|\big)\big]^2.$$

Applying the last inequality successively with $\sigma = 1+\omega\tau$ and with $\sigma = \theta^{-1}$, we easily obtain from the above upper bound for $\|f(B)\|$ the inequality (4.4). □

4.3. *Upper bounds for* $|u_n|$. We now state three interesting corollaries to the above lemma, the first two corollaries of which were essential in our proof of Theorem 4.1.

COROLLARY 4.3. *Let* $0 < \tau < \infty$, $0 < \rho < \infty$, $\rho \leqslant r$ *and* $0 < h_n \leqslant \rho\tau$. *Assume* $\omega \geqslant 0$ *and* ϕ *regular on* $\mathscr{D}(-\rho, (1+\omega\tau)\rho)$. *Then* $\phi(h_nA)$ *exists and* (4.1) *implies* $|u_n| \leqslant \gamma n^{1/2}L^n|u_0|$ (*for each* $n \geqslant 1$, $s \geqslant 1$, $u_0 \in \mathbf{R}^s$, $A \in \mathscr{L}(\mathbf{R}^s, \omega, \tau)$ *and each norm* $|\cdot|$ *in* $\mathbf{R}^s$). *Here* γ *only depends on* ϕ, ρ *and on the product* $\omega\tau$. *Further*, $L = \mathscr{M}(-\rho, (1+\omega\tau)\rho)$, *and* $L = 1$ *when* $\omega = 0$.

Proof. We shall apply Lemma 4.2 with $\theta > 0$ such that

$$1 + \omega\tau < \theta^{-1}, \qquad \mathscr{M}\big(-\rho, \theta^{-1}\rho\big) < \infty.$$

Clearly, the relations (4.2) are fulfilled with

$$\lambda_n \leqslant L = \mathscr{M}\big(-\rho, (1+\omega\tau)\rho\big) < \infty, \qquad \mu_n \leqslant M = \mathscr{M}\big(-\rho, \theta^{-1}\rho\big) < \infty.$$

Note that $L = 1$ when $\omega = 0$, since $\phi(0) = 1$, $\rho \leqslant r$.

Let $n \geqslant 1$, $s \geqslant 1$, $u_0 \in \mathbf{R}^s$, $u_0 \neq 0$, $A \in \mathscr{L}(\mathbf{R}^s, \omega, \tau)$ and $|\cdot|$ be given. By Lemma 4.2 we then have (4.3), and (4.1) then implies

$$|u_n|/|u_0| \leqslant L^n m^{1/2} + M^n\left\{\sum_{k=m}^{\infty}\theta^{2k}(1+\omega\tau)^{2k}\right\}^{1/2}.$$

Choosing $m = nj$, we thus have

$$|u_n|/|u_0| \leqslant L^n(jn)^{1/2} + \big\{M\big[\theta(1+\omega\tau)\big]^j\big\}^n\cdot\big\{1-\theta^2(1+\omega\tau)^2\big\}^{-1/2}.$$

Taking the integer j so large that $M[\theta(1+\omega\tau)]^j \leqslant L\{1-\theta^2(1+\omega\tau)^2\}^{1/2}$ there follows $|u_n|/|u_0| \leqslant (j^{1/2}+1)L^n n^{1/2}$, which proves the corollary since $\gamma = (j^{1/2}+1)$ only depends on ϕ, ρ, $\omega\tau$. □

COROLLARY 4.4. *Let* $0 < \tau < \infty$, $0 < h < \infty$ *and* $0 < h \leqslant r\tau$. *Assume* $\omega < 0$. *Then* $\phi(hA)$ *exists and* (1.1) *implies*

$$|u_n| \leqslant [(-\omega\tau)(2+\omega\tau)]^{-1/2}|u_0|$$

(*for each* $n \geqslant 1$, $s \geqslant 1$, $u_0 \in \mathbf{R}^s$, $A \in \mathscr{L}(\mathbf{R}^s, \omega, \tau)$ *and each norm* $|\cdot|$ *in* $\mathbf{R}^s$).

Proof. When $A \in \mathscr{L}(\mathbf{R}^s, \omega, \tau)$, we can apply Lemma 4.2 with $h_n \equiv h$, $\theta = 1$, $m = 0$. By (4.3) the matrix $\phi(hA)$ thus exists, and by (4.4) we easily arrive at

$$\|\phi(hA)^n\| \leqslant \left\{ \sum_{k=0}^{\infty} \|(I+\tau A)^k\|^2 \right\}^{1/2}.$$

In view of (2.1), we have $\|I+\tau A\| \leqslant 1+\omega\tau$, and, consequently,

$$\|\phi(hA)^n\|^2 \leqslant \sum_{k=0}^{\infty} (1+\omega\tau)^{2k} = [(-\omega\tau)(2+\omega\tau)]^{-1}. \quad \square$$

The next corollary is obtained by applying Lemma 4.2 similarly as in the proof of Corollary 4.4 with $h_n \equiv h$, $m = 0$, but with $\theta = (1+\omega\tau/2)^{-1}$. Since the proof is analogous to the above, we omit it.

COROLLARY 4.5. *Let* $0 < \tau < \infty$, $0 < h < \infty$ *and* $0 < h \leqslant r\tau$. *Assume* $\omega < 0$. *Then* $\phi(hA)$ *exists and* (1.1) *implies* $|u_n| \leqslant \gamma M^n |u_0|$ (*for each* $n \geqslant 1$, $s \geqslant 1$, $u_0 \in \mathbf{R}^s$, $A \in \mathscr{L}(\mathbf{R}^s, \omega, \tau)$ *and each norm* $|\cdot|$ *in* $\mathbf{R}^s$). *Here* $\gamma = (2+\omega\tau)[-\omega\tau(4+3\omega\tau)]^{-1/2}$, *and* $M = \mathscr{M}(-h\tau^{-1}, (1+\omega\tau/2)h\tau^{-1})$ *satisfies* $M < 1$ (*provided the rational function* ϕ *is no constant*).

This corollary implies that when $\omega < 0$, $A \in \mathscr{L}(\mathbf{R}^s, \omega, \tau)$ and h is restricted as in Theorem 4.1, any u_n satisfying (1.1) damps out exponentially when $n \to \infty$. This asymptotic behavior of the approximations $u_n \simeq U(nh)$ is desirable since an analogous behavior is shown by the true solution $U(t)$ to (1.2) (cf. Remark 1 of Section 2).

5. On the Relevance of the Radius R to the Stability of (1.1). In this section the parameter R, defined in Subsection 1.2, will be compared with the radius r and related to the stability of the process (1.1).

The definition of R implies that, for any $\rho \in (0, R)$, the Taylor coefficients $\gamma_k = (k!)^{-1}\phi^{(k)}(-\rho)$ satisfy $\gamma_k \geqslant 0$ and

$$\gamma_0 + \gamma_1\rho + \gamma_2\rho^2 + \cdots = \phi(0) = 1.$$

Consequently, procedure (1.1) is $A[D]$-stable for any $D = \mathscr{D}(-\rho, \rho)$ with $\rho \in (0, R)$, so that

$$R \leqslant r. \tag{5.1}$$

The subsequent Theorem 5.1 has a structure similar to the one of Theorem 4.1 in the above section. Comparing both theorems we see that, in view of (5.1), the stepsize restriction (S1) occurring in Theorem 5.1 is, generally, more severe than the analogous restriction (s1) in Theorem 4.1. This is in agreement with the fact that the corresponding stability result (S3) in the subsequent theorem is much stronger than

the analogous result (s3) in Theorem 4.1. The property $|u_n| \leqslant |u_0|$ occurring in (S3) is often called *contractivity* (cf., e.g., [23], [26]) and is related to what sometimes is called *practical stability* (cf., e.g., [12]).

An illustration of Theorem 5.1 is presented in Section 7.

THEOREM 5.1. *Let h, τ, p be given with $0 < h < \infty$, $0 < \tau < \infty$, and $p = 1$ or $p = \infty$. Then, the following three statements are equivalent to each other.*

(S1) $0 < h \leqslant R\tau$;

(S2) $\phi(hA)$ *exists and* (1.1) *implies* $|u_n|_p \leqslant |u_0|_p$ (*for each* $n \geqslant 1$, $s \geqslant 1$, $u_0 \in \mathbf{R}^s$, $A = A_s$ (*see* (3.2)) *with* $-\xi = \rho = \tau^{-1}$);

(S3) $\phi(hA)$ *exists and* (1.1) *implies* $|u_n| \leqslant |u_0|$ (*for each* $n \geqslant 1$, $s \geqslant 1$, $u_0 \in \mathbf{R}^s$, $A \in \mathscr{L}(\mathbf{R}^s, 0, \tau)$ *and each norm* $|\cdot|$ *in* $\mathbf{R}^s$).

Proof. 1. (S3) implies (S2) since $A = A_s$ with $-\xi = \rho = \tau^{-1}$ belongs to $\mathscr{L}(\mathbf{R}^s_p, 0, \tau)$.

2. Assuming (S2) we prove (S1). With A as in (S2) we have $\phi(hA) = \gamma_0 + \gamma_1(\beta E) + \cdots + \gamma_{s-1}(\beta E)^{s-1}$, where $\beta = h\rho$, $\gamma_k = (k!)^{-1}\phi^{(k)}(-\beta)$ and E is the matrix defined in the proof of Lemma 3.2. Consequently,

$$1 \geqslant \|\phi(hA)\|_p = |\gamma_0| + |\gamma_1|\beta + \cdots + |\gamma_{s-1}|\beta^{s-1}$$

for each $s \geqslant 1$. We thus obtain

$$1 \geqslant \sum_{k=0}^{\infty} |\gamma_k|\beta^k \geqslant \sum_{k=0}^{\infty} \gamma_k \beta^k = \phi(0) = 1.$$

It follows that $\phi^{(k)}(-\beta) \geqslant 0$ and $\phi(\zeta)$ is regular on $\mathscr{D}(-\beta, \beta)$.

For any $t \in [-\beta, 0]$ we thus have

$$\phi^{(k)}(t) = \phi^{(k)}(-\beta) + (t+\beta)\phi^{(k+1)}(-\beta) + (2!)^{-1}(t+\beta)^2\phi^{(k+2)}(-\beta) + \cdots \geqslant 0$$

(for all $k \geqslant 0$). Hence $\beta \leqslant R$, which proves (S1).

3. Assume (S1) and let $A \in \mathscr{L}(\mathbf{R}^s, 0, \tau)$. In view of (5.1) we have, similarly as in the beginning of the proof of Lemma 4.2 (with $n = 1$),

$$\phi(hA) = \gamma_0 + \gamma_1 B + \gamma_2 B^2 + \cdots$$

with $\gamma_k = (k!)^{-1}(h\tau^{-1})^k\phi^{(k)}(-h\tau^{-1})$, $\|B\| \leqslant 1$. Hence,

$$\|\phi(hA)\| \leqslant |\gamma_0| + |\gamma_1| + |\gamma_2| + \cdots = \gamma_0 + \gamma_1 + \gamma_2 + \cdots = \phi(0) = 1,$$

which completes the proof of (S3). □

Remark. The implication (S1) ⇒ (S3) follows immediately from [23, Theorem 3.3]. We have included the above proof of this implication since it is very short and keeps the paper self-contained.

6. Modifications of Theorems 4.1, 5.1. Suppose the norm in $\mathbf{R}^s$ is *generated by an inner product*, i.e., $|x| = \langle x, x\rangle^{1/2}$ (for all $x \in \mathbf{R}^s$). Then, an interesting modification of Theorems 4.1, 5.1 is possible.

One arrives at this modification by a straightforward application of Theorem 3.1 and a theorem of J. von Neumann (see [18, p. 432], [8], [2]). The latter theorem, applied to the situation at hand, says that for any $s \times s$ matrix T with lub-norm $\|T\| \leqslant 1$ and rational f mapping $\mathscr{D}(0,1)$ into $\mathscr{D}(0,1)$, one has the bound $\|f(T)\| \leqslant 1$.

By choosing $f(\zeta) = \phi(-h\tau^{-1}(1 - \zeta))$, $T = I + \tau A$, one thus obtains

THEOREM 6.1. *Let $h > 0$, $\tau > 0$ be given. Then, the following three statements are equivalent to each other.*

(σ1) $0 < h \leqslant r\tau$;

(σ2) $\phi(hA)$ *exists and* (1.1) *implies* $|u_n|_2 \leqslant |u_0|_2$ (*for each* $n \geqslant 1$, $s \geqslant 1$, $u_0 \in \mathbf{R}^s$, $A = A_s$ (*see* (3.2)) *with* $-\xi = \rho = \tau^{-1}$);

(σ3) $\phi(hA)$ *exists and* (1.1) *implies* $|u_n| \leqslant |u_0|$ (*for each* $n \geqslant 1$, $s \geqslant 1$, $u_0 \in \mathbf{R}^s$, $A \in \mathscr{L}(\mathbf{R}^s, 0, \tau)$ *and each norm* $|\cdot|$ *in* $\mathbf{R}^s$ *generated by an inner product*).

By this theorem we thus have contractivity (see (σ3)) for stepsizes h subject to (σ1). This stepsize restriction is, generally, weaker than restriction (S1) in Theorem 5.1 (cf. (5.1)). On the other hand, (σ1) equals restriction (s1) of Theorem 4.1, but (σ3) cannot be deduced from the stability statement (s3) of that theorem.

From Theorem 6.1 we conclude that in Theorem 5.1 one cannot allow $p = 2$. Only for ϕ with $R = r$, Theorem 5.1 remains valid with $p = 2$, but this is exceptional (cf. (5.1)).

We conclude this section by briefly discussing the possibility of replacing the factor $(\gamma n^{1/2})$ in statement (s3) of Theorem 4.1 by a factor, say γ_0, only depending on ϕ. For $A \in \mathscr{L}(\mathbf{R}^s, 0, \tau)$ the true solution $U(t)$ to (1.2) is, by Remark 1 of Section 2, bounded (uniformly for $0 \leqslant t < \infty$). Therefore, one might hope for an analogous boundedness behavior in (s3) for the approximations $u_n \simeq U(nh)$ (uniformly for $n \geqslant 0$).

Unfortunately, without additional conditions on ϕ, Theorem 4.1 does not allow such a modification. In fact, $\phi(\zeta) = (1 - \zeta/2)^{-1}(1 + \zeta/2)$ provides a counterexample with $r = \infty$. Using arguments taken from [25, pp. 280–287], it can be proved that for $A = E - I \in \mathscr{L}(\mathbf{R}^s_\infty, 0, 1)$ (with E as in the proof of Lemma 3.2) one has $\sup\|[\phi(hA)]^n\|_\infty = \infty$, the supremum being for $s \geqslant 1$, $n \geqslant 1$, $h > 0$.

7. Illustration in a Diffusion-Convection Problem. We turn to the application of the above to the stability analysis of difference methods for solving partial differential equations. Due to the framework we have been using, applications seem possible with arbitrary norms and any number of space variables, any boundary conditions, variable coefficients and variable space discretizations. On the other hand, due to the generality of the above theorems, one would not expect that in any actual application sharp, or refined, stability results can be obtained.

As an illustration, we consider the problem

$$\frac{\partial}{\partial t}U(x,t) = \frac{\partial^2}{\partial x^2}U(x,t) + b(x)\frac{\partial}{\partial x}U(x,t) + c(x)U(x,t), \tag{7.1.a}$$

$$U(0,t) = U(1,t) = 0, \tag{7.1.b}$$

$$U(x,0) = U_0(x), \tag{7.1.c}$$

where $0 < x < 1$, $t > 0$ and U_0, b, c are given bounded real functions (which need not be smooth) with

$$c(x) \leqslant \omega \qquad (0 < x < 1).$$

The following finite-difference scheme has been constructed according to well-known principles (cf. [12], [7], [26], [24], [15]).

$$h^{-1}\left(u_j^n - u_j^{n-1}\right) = \delta^{-2}(1 - \beta_j + \varepsilon_j|\beta_j|)\left(\theta u_{j-1}^n + (1-\theta)u_{j-1}^{n-1}\right)$$

$$+ \delta^{-2}\left(-2 - 2\varepsilon_j|\beta_j| + c_j\delta^2\right)\left(\theta u_j^n + (1-\theta)u_j^{n-1}\right) \tag{7.2.a}$$

$$+ \delta^{-2}\left(1 + \beta_j + \varepsilon_j|\beta_j|\right)\left(\theta u_{j+1}^n + (1-\theta)u_{j+1}^{n-1}\right),$$

$$u_0^{n-1} = u_{s+1}^{n-1} = 0, \tag{7.2.b}$$

$$u_j^0 = U_0(j\delta), \tag{7.2.c}$$

where $j = 1, 2, \ldots, s$ and $n = 1, 2, 3, \ldots$. In (7.2) we use the notations $h = \Delta t > 0$, $\delta = \Delta x > 0$, $(s+1)\delta = 1$, $\beta_j = \delta \cdot b(j\delta)/2$, $c_j = c(j\delta)$, $u_j^n \simeq U(j\delta, nh)$, and $\varepsilon_j \in [0,1]$, $\theta \in [0,1]$ are parameters specifying the method. The choices $\varepsilon_j \equiv 0$ and $\varepsilon_j \equiv 1$ yield central finite-difference and fully upwinded finite-difference approximations to $\partial U(x,t)/\partial x$, respectively. The choices $\theta = 0, \frac{1}{2}, 1$ correspond to the explicit, the Crank-Nicolson and the fully implicit schemes, respectively.

With the definitions $u_n = (u_1^n, u_2^n, \ldots, u_s^n)^T$ and

$$\phi(\zeta) = (1 + (1-\theta)\zeta)(1 - \theta\zeta)^{-1},$$

the relations (7.2.a), (7.2.b) become equivalent to the recurrence relation (1.1) provided the matrix $A = (\alpha_{ij})$ is given by $\alpha_{ij} = 0$ $(|i-j| > 1)$,

$$\alpha_{j,j-1} = \delta^{-2}(1 - \beta_j + \varepsilon_j|\beta_j|) \qquad (2 \leqslant j \leqslant s),$$

$$\alpha_{jj} = \delta^{-2}\left(-2 - 2\varepsilon_j|\beta_j| + c_j\delta^2\right) \qquad (1 \leqslant j \leqslant s),$$

$$\alpha_{j,j+1} = \delta^{-2}(1 + \beta_j + \varepsilon_j|\beta_j|) \qquad (1 \leqslant j \leqslant s-1).$$

Using the maximum norm in $\mathbf{R}^s$, we have (cf. [3], [26])

$$\|A\|_\infty = \max_i \sum_j |\alpha_{ij}|, \qquad \mu_\infty[A] = \max_i \left(\alpha_{ii} + \sum_{j \neq i} |\alpha_{ij}|\right).$$

Consequently, choosing ε_j such that

$$\varepsilon_j \geqslant 1 - \frac{2}{\delta|b(j\delta)|} \qquad (1 \leqslant j \leqslant s), \tag{7.3}$$

we have $\mu_\infty[A] \leqslant \omega$, $\|A\|_\infty \leqslant \alpha = \delta^{-2}(4 + 2\delta\lambda + \delta^2|c|)$, where

$$\lambda = \max_j |\varepsilon_j b(j\delta)|, \qquad |c| = \sup_x |c(x)|.$$

In view of (2.2), we have $A \in \mathscr{B}(\mathbf{R}_\infty^s, \omega, \alpha)$, and from Theorem 2.1, we thus obtain

$$A \in \mathscr{L}(\mathbf{R}_\infty^s, \omega, \tau), \qquad \tau = \delta^2 \cdot \left(2 + \delta\lambda + \delta^2(|c| - \omega)/2\right)^{-1}. \tag{7.4}$$

One easily verifies that, with ϕ as defined above,

$$r = (1 - 2\theta)^{-1} \quad \left(0 \leqslant \theta < \tfrac{1}{2}\right), \qquad r = \infty \quad \left(\tfrac{1}{2} \leqslant \theta \leqslant 1\right), \tag{7.5.a}$$

$$R = (1 - \theta)^{-1} \quad (0 \leqslant \theta < 1), \qquad R = \infty \quad (\theta = 1). \tag{7.5.b}$$

From Theorem 4.1 it thus follows that the solution to (7.2) satisfies the stability estimate

$$\max_j |u_j^n| \leqslant \delta^{-1} \cdot |\omega|^{-1/2}\left(2 + \delta\lambda + \delta^2(|c| - \omega)/2\right)^{1/2} \cdot \max_j |u_j^0| \tag{7.6.a}$$

whenever $\omega < 0$, (7.3) and

$$(7.7.a)\qquad \begin{cases} h\delta^{-2} \leqslant (1-2\theta)^{-1}\big(2+\delta\lambda+\delta^2(|c|-\omega)/2\big)^{-1} & (\text{for } 0 \leqslant \theta < \tfrac{1}{2}), \\ h\delta^{-2} < \infty & (\text{for } \tfrac{1}{2} \leqslant \theta \leqslant 1). \end{cases}$$

Similarly, Theorem 5.1 yields the estimate

$$(7.6.b)\qquad \max_j |u_j^n| \leqslant \max_j |u_j^0|$$

whenever $\omega \leqslant 0$, (7.3) and

$$(7.7.b)\qquad \begin{cases} h\delta^{-2} \leqslant (1-\theta)^{-1}\big(2+\delta\lambda+\delta^2(|c|-\omega)/2\big)^{-1} & (\text{for } 0 \leqslant \theta < 1), \\ h\delta^{-2} < \infty & (\text{for } \theta = 1). \end{cases}$$

Further, Corollary 4.3 yields estimates when $\omega \geqslant 0$ and the stepsizes $h_n > 0$ vary under restrictions similar to (7.7.a).

Of course, (7.6) can be used in a standard way (cf. [17], [11]) to obtain maximum-norm bounds for the global error $u_j^n - U(j\delta, nh)$. Here the factor δ^{-1} in the right-hand member of (7.6.a) (as well as the factor $n^{1/2}$ appearing in Corollary 4.3) need not stand in the way to obtain bounds for the global error that are of the same order as the local discretization errors (cf., e.g., [17, pp. 124–130], [21]).

We conclude with comparing the above estimates to some stability estimates obtainable from the literature.

For $h\delta^{-2} < \infty$, $\frac{1}{2} < \theta \leqslant 1$, an estimate that is essentially sharper than (7.6.a) follows from the general theory in [1]. But, for $0 \leqslant \theta \leqslant \frac{1}{2}$, the estimate (7.6.a) is no direct consequence of that theory. Also, the variable stepsize result mentioned above does not follow from [1].

For the case $\theta = 0$ and the pure diffusion-convection equation (7.1.a), with $c(x) \equiv 0$, $b(x) \equiv$ constant, Griffiths et al. [7] presented restrictions on $\varepsilon_j \equiv \varepsilon$ and on $h\delta^{-2}$ that are necessary and sufficient for (7.6.b) to be valid. It is interesting to note that for this case the conditions (7.3), (7.7.b) neatly reduce to the restrictions presented in [7].

Acknowledgment. I wish to thank J. M. Sanz-Serna for a stimulating discussion on the topic of this paper.

Institute of Applied Mathematics and Computer Science
University of Leiden
P. O. Box 9512
2300 RA Leiden, The Netherlands

1. P. BRENNER & V. THOMÉE, "On rational approximations of semigroups," *SIAM J. Numer. Anal.*, v. 16, 1979, pp. 683–694.

2. M. CROUZEIX & P. A. RAVIART, "Approximation d'équations d'évolution linéaires par des méthodes multipas," in *Etude Numérique des Grands Systèmes*, Rencontres IRIA-Novosibirsk, 1976, Dunod, Paris.

3. G. DAHLQUIST, "Stability and error bounds in the numerical integration of ordinary differential equations," *Trans. Roy. Inst. Techn.*, No. 130, Stockholm, 1959.

4. G. DAHLQUIST, "A special stability problem for linear multistep methods," *BIT*, v. 3, 1963, pp. 27–43.

5. G. DAHLQUIST, "*G*-stability is equivalent to *A*-stability," *BIT*, v. 18, 1978, 384–401.

6. N. DUNFORD & J. T. SCHWARTZ, *Linear Operators, Part* I, Interscience, New York, 1958.

7. D. F. GRIFFITHS, I. CHRISTIE & A. R. MITCHELL, "Analysis of error growth for explicit difference schemes in conduction-convection problems," *Internat. J. Numer. Methods Engrg.*, v. 15, 1980, pp. 1075–1081.

8. R. HERSH & T. KATO, "High-accuracy stable difference schemes for well-posed initial-value problems," *SIAM J. Numer. Anal.*, v. 16, 1979, pp. 670–682.

9. R. JELTSCH & O. NEVANLINNA, "Largest disk of stability of explicit Runge-Kutta methods," *BIT*, v. 18, 1978, pp. 500–502.

10. R. JELTSCH & O. NEVANLINNA, "Stability of explicit time discretizations for solving initial value problems," *Numer. Math.*, v. 37, 1981, pp. 61–91.

11. A. R. MITCHELL & D. F. GRIFFITHS, *The Finite Difference Method in Partial Differential Equations*, Wiley, Chichester, 1980.

12. K. W. MORTON, "Stability of finite difference approximations to a diffusion-convection equation," *Internat. J. Numer. Methods Engrg.*, v. 15, 1980, pp. 677–683.

13. O. NEVANLINNA, "On the numerical integration of nonlinear initial value problems by linear multistep methods," *BIT*, v. 17, 1977, pp. 58–71.

14. F. ODEH & W. LINIGER, "A note on unconditional fixed $-h$ stability of linear multistep formulae," *Computing*, v. 7, 1971, pp. 240–253.

15. S. PAOLUCCI & D. R. CHENOWETH, "Stability of the explicit finite differenced transport equation," *J. Comput. Phys.*, v. 47, 1982, pp. 489–496.

16. S. V. PARTER, "Stability, convergence, and pseudo-stability of finite-difference equations for an over-determined problem," *Numer. Math.*, v. 4, 1962, pp. 277–292.

17. R. D. RICHTMYER & K. W. MORTON, *Difference Methods for Initial-Value Problems*, 2nd ed., Wiley, New York, 1967.

18. F. RIESZ & B. SZ-NAGY, *Leçons d'Analyse Fonctionelle*, 2nd ed., Akademiai Kiado, Budapest, 1953.

19. J. M. SANZ-SERNA, *Convergent Approximations to Partial Differential Equations and Stability Concepts of Methods for Stiff Systems of Ordinary Differential Equations*, Report, University of Valladolid, 1983.

20. B. SCHMITT, "Norm bounds for rational matrix functions," *Numer. Math.*, v. 42, 1983, pp. 379–389.

21. M. N. SPIJKER, *Equivalence Theorems for Nonlinear Finite-Difference Methods*, Lecture Notes in Math., Vol. 267, Springer-Verlag, Berlin and New York, 1972, pp. 233–266.

22. M. N. SPIJKER, "Numerical contractivity in the solution of initial value problems," in *Proc. Zweites Seminar über Numerische Behandlung von Differentialgleichungen*, Halle (DDR), 1983.

23. M. N. SPIJKER, "Contractivity in the numerical solution of initial value problems," *Numer. Math.*, v. 42, 1983, pp. 271–290.

24. G. STOYAN, "Monotone difference schemes for diffusion-convection problems," *Z. Angew. Math. Mech.*, v. 59, 1979, pp. 361–372.

25. V. THOMÉE, "Stability of difference schemes in the maximum-norm," *J. Differential Equations*, v. 1, 1965, pp. 273–292.

26. J. G. VERWER & K. DEKKER, *Step-by-Step Stability in the Numerical Solution of Partial Differential Equations*, Report NW 161/83, Mathematical Centre, Amsterdam, 1983.