BIT 37:2 (1997), 442–464.

ERROR GROWTH ANALYSIS VIA STABILITY REGIONS FOR DISCRETIZATIONS OF INITIAL VALUE PROBLEMS *

M. N. SPIJKER and F. A. J. STRAETEMANS

Department of Mathematics and Computer Science University of Leiden
Niels Bohrweg 1, 2333 CA Leiden, The Netherlands. email: spijker@wi.leidenuniv.nl

Abstract.

This paper deals with numerical methods for the solution of linear initial value problems. Two main theorems are presented on the stability of these methods.

Both theorems give conditions guaranteeing a mild error growth, for one-step methods characterized by a rational function $\varphi(z)$. The conditions are related to the stability region $S = \{z : z \in \mathbb{C} \text{ with } |\varphi(z)| \leq 1\}$, and can be viewed as variants to the resolvent condition occurring in the reputed Kreiss matrix theorem. Stability estimates are presented in terms of the number of time steps n and the dimension s of the space.

The first theorem gives a stability estimate which implies that errors in the numerical process cannot grow faster than linearly with s or n. It improves previous results in the literature where various restrictions were imposed on S and $\varphi(z)$, including $\varphi'(z) \neq 0$ for $z \in \partial S$ and S be bounded. The new theorem is not subject to any of these restrictions.

The second theorem gives a sharper stability result under additional assumptions regarding the differential equation. This result implies that errors cannot grow faster than n^{β}, with fixed $\beta < 1$.

The theory is illustrated in the numerical solution of an initial-boundary value problem for a partial differential equation, where the error growth is measured in the maximum norm.

AMS subject classification: 65L05, 65L20, 65M12, 65M20.

Key words: Initial value problem, discretization, numerical method, error growth, stability analysis, stability region, resolvent condition.

1 Introduction.

1.1 The purpose of the paper.

This paper is concerned with step-by-step methods for the numerical solution of initial value problems. We shall deal both with methods for partial differential equations and methods for (stiff) ordinary differential equations.

An important issue in the step-by-step solution of initial value problems is the a priori assessment of the *stability behaviour* of a given numerical process. Classical tools to predict whether numerical processes will behave stably or not include the

*Received February 1996. Revised May 1996.

famous *Von Neumann condition* (for partial differential equations) and the so-called *stability regions* in the complex plane (for ordinary differential equations). These two tools are based on the behaviour of numerical methods when applied to very simple linear testproblems. Accordingly, it is not surprising that for linear problems which are more general and realistic than the testproblems, these tools can fail to be relevant. For certain problems a careless application of stability regions can lead to a completely wrong assessment of stability.

Recently stability regions were used successfully in a stability analysis of numerical methods for linear differential equations that are essentially more general than the classical test equations. Under special assumptions about these regions, and by using variants of the so-called *resolvent condition of Kreiss*, rigorous stability estimates were established among others by Crouzeix, Larsson, Piskarev and Thomée [2, 1993], Lenferink and Spijker [10, 1991] and [11, 1991], Lubich and Nevanlinna [12, 1991], Palencia [15, 1995], and Reddy and Trefethen [17, 1992].

However, part of the assumptions made in these references, about the stability regions and the differential equations, are inconvenient in that they are not fulfilled in various cases of practical interest. One of the aims of this paper is to present a new theorem which does not require those assumptions. Moreover, in this paper some of the stability estimates in the above references will be sharpened considerably under conditions fulfilled in certain applications.

1.2 The numerical process.

In this paper we study numerical processes that can be written in the form

$$u_n = \varphi(hA)u_{n-1} + f_n \qquad \text{for } n = 1, 2, 3, \ldots . \tag{1.1}$$

Here u_n denote numerical approximations, belonging to the s-dimensional complex vector space $\mathbb{C}^s$, which are computed successively from (1.1), starting from a given $u_0 \in \mathbb{C}^s$. In the above, $h > 0$ is the so-called *stepsize*, f_n are given vectors in $\mathbb{C}^s$, and A is a complex $s \times s$ matrix. Further, $\varphi(z) = P(z)/Q(z)$ is a given rational function, where $P(z)$, $Q(z)$ are polynomials with no common zero. We define $\varphi(hA) = P(hA)[Q(hA)]^{-1}$ whenever the matrix $Q(hA)$ is invertible.

Processes of the form (1.1) occur in the numerical solution of systems of *ordinary differential equations*. Consider the initial value problem

$$U'(t) = AU(t) + F(t) \qquad \text{for } t \geq 0, \qquad U(0) = u_0. \tag{1.2}$$

Any Runge–Kutta method or Rosenbrock method (cf. Butcher [1, 1987] or Hairer and Wanner [7, 1991] applied to such a problem reduces to a process of the form (1.1). In this case u_n approximates $U(t) \in \mathbb{C}^s$ at $t = nh$, for $n = 1, 2, 3, \ldots$.

We note that initial value problems of the above type can arise when the method of semi-discretization is applied to initial boundary value problems in linear *partial differential equations*. The semi-discretization can be based, e.g., on finite differences, finite volumes, finite elements or spectral approximations.

In all of these cases the dimension s of the above system of ordinary differential equations is related to the accuracy of the semi-discretization, and can attain (arbitrarily) large values.

Processes of the form (1.1) can also occur in cases where a given partial differential equation is discretized directly in full, without the intermediate phase of a system of ordinary differential equations. For instance the Crank-Nicholson scheme can be written in the form (1.1), with $\varphi(z) = (1 + z/2)(1 - z/2)^{-1}$ (see, e.g., Richtmyer and Morton [18, 1967]).

1.3 Stability of the numerical process.

Suppose the numerical calculations, according to (1.1), were performed using a slightly perturbed initial vector $\widetilde{u}_0$, instead of u_0, for instance due to rounding off. We denote the vectors obtained from (1.1) in this situation by $\widetilde{u}_n$. The numerical process is called *stable* if any initial perturbation has only a moderate effect on the corresponding numerical approximations—that is, $v_n = \widetilde{u}_n - u_n$ can be bounded suitably in terms of $v_0 = \widetilde{u}_0 - u_0$. Since

$$v_n = \widetilde{u}_n - u_n = [\varphi(hA)\widetilde{u}_{n-1} + f_n] - [\varphi(hA)u_{n-1} + f_n] = \varphi(hA)v_{n-1} = \varphi(hA)^n v_0,$$

the stability analysis of (1.1) deals with establishing bounds, of moderate size, on the powers of $\varphi(hA)$. As the dimension s can be large we have to pay special attention to the dependence of such bounds on s.

An obvious way to assess the stability of (1.1) consists in using the eigenvalues of $\varphi(hA)$. Denoting the set of all eigenvalues of any matrix B by $\sigma[B]$, we have $\sigma[\varphi(hA)] = \varphi(\sigma[hA])$. In order to guarantee moderate bounds on the powers of $\varphi(hA)$ one might thus require that $|\varphi(z)| < 1$ for all $z \in \sigma[hA]$. We call this the *eigenvalue condition* for stability.

We define the *stability region* of φ by

$$S = \{z : \ z \in \mathbb{C} \text{ with } |\varphi(z)| \leq 1\}.$$

Clearly, the above eigenvalue condition can be cast into the form

$$\sigma[hA] \quad \subset \quad \text{int}(S),$$

where $\text{int}(S)$ denotes the interior of the set S.

This condition is notorious for being very *unreliable*, since in many cases of practical interest it cannot prevent the matrices $\varphi(hA)^n$ from having entries with excessively large magnitudes (see, e.g., Parter [16, 1962], Griffiths, Christie and Mitchell [5, 1980], Morton [13, 1980], Spijker [19, 1985], Reddy and Trefethen [17, 1992] and Section 4.2 of the present paper).

Reliable ways of using stability regions comprise an appropriate condition on the so-called *resolvent* $(zI - hA)^{-1}$ of hA; here I denotes the $s \times s$ identity matrix. In order to outline this condition we assume that $\|\cdot\|$ is a *norm* on the vector space of all complex $s \times s$ matrices (that is $\|A + B\| \leq \|A\| + \|B\|$, $\|\lambda A\| = |\lambda| \cdot \|A\|$, $\|A\| = 0 \Rightarrow A = 0$, for all matrices A, B and $\lambda \in \mathbb{C}$). Further, for any $z \in \mathbb{C}$, $X \subset \mathbb{C}$, we denote the *distance* from z to X by

$$\text{d}(z, X) = \inf\{|z - x| : \ x \in X\}.$$

We assume that K is a real constant, and V a closed subset of $\mathbb{C}$. We consider the situation, where

$$(1.3)\quad (zI-hA) \text{ is invertible, and } \|(zI-hA)^{-1}\| \le \frac{K}{\mathrm{d}(z,V)}, \quad \text{for all } z \in \mathbb{C}\backslash V.$$

Since (1.3) is a variant to one of the equivalent conditions occurring in the famous Kreiss matrix theorem (see, e.g., Richtmyer and Morton [18, 1967]) we refer to it as the *Kreiss resolvent condition on hA with respect to V, with constant K.*

Condition (1.3) can be shown to imply stability estimates of the form

$$(1.4)\qquad \|\varphi(hA)^n\| \le \gamma K s^\alpha n^\beta \quad (\text{for } s \ge 1,\ n \ge 1),$$

under specific assumptions about φ, V and $\|\cdot\|$. Here α, β, γ are nonnegative constants which share the crucial property of depending only on φ (and possibly on V, but *not* on s, n, hA or K).

1.4 A review of stability results from the literature.

We shall review some known estimates of the form (1.4) which are based on (1.3). In our review we need some definitions. For any set $X \subset \mathbb{C}$ we denote by ∂X, $\mathrm{int}(X)$ and $\mathrm{cl}(X)$ the *boundary, interior* and *closure* of X, respectively. For $z \neq 0$ we denote by $\mathrm{Arg}(z)$ the principal part of the *argument* with $-\pi < \mathrm{Arg}(z) \le \pi$, and we define $\mathrm{Arg}(0) = 0$. For $0 \le \theta < \pi/2$ we define the *wedge* $W(\theta) = \{z : z \in \mathbb{C}, \text{ with } |\mathrm{Arg}(-z)| \le \theta\}$.

In our review we assume that $\varphi(0) = \varphi'(0) = 1$, and

$$\|B\| = \max\{|Bx|/|x| :\ x \in \mathbb{C}^s,\ x \neq 0\} \quad \text{for all } s \times s \text{ matrices } B,$$

where $|\cdot|$ denotes an arbitrary norm on $\mathbb{C}^s$. The results from the literature fit into the following two categories.

Category 1. *The exponents in* (1.4) *satisfy:* $\alpha = 1$, $\beta = 0$ *or* $\alpha = 0$, $\beta = 1$.
In Lenferink and Spijker [11, 1991], the estimate (1.4) with $\alpha = 0$, $\beta = 1$ was proved in the situation where (1.3) holds with bounded, convex $V \subset S$. In Lenferink and Spijker [10, 1991], the estimate with $\alpha = 1$, $\beta = 0$ was proved under the additional assumptions that ∂S lies on an algebraic curve and $\varphi'(z) \neq 0$ on $\partial V \cap \partial S$.

In an important paper by Reddy and Trefethen [17, 1992], the estimate (1.4) was proved, both with $\alpha = 0$, $\beta = 1$ and $\alpha = 1$, $\beta = 0$, in the situation (1.3) for bounded $V = S$ with $\varphi'(z) \neq 0$ on whole of ∂S. (These authors assumed the norm on $\mathbb{C}^s$ to be generated by an inner product, but their proof is valid for general norms. On the other hand, their proof is *not* valid for unbounded S or when $\varphi'(z) = 0$ at some $z \in \partial S$).

Lubich and Nevanlinna [12, 1991] perceived that the boundedness of V is not essential—they succeeded in proving (1.4), with $\alpha = 0$, $\beta = 1$ and $\alpha = 1$, $\beta = 0$, for the interesting case where $V = \{z :\ \mathrm{Re}\ z \le 0\} \subset S$.

None of the estimates just mentioned implies any of the other ones, or provides us with a stability result for the general case of arbitrary φ and $V \subset S$.

Category 2. *The exponents in* (1.4) *satisfy*: $\alpha = 0$, $0 \leq \beta < 1$.
Lubich and Nevanlinna [12, 1991] proved (1.4) with $\alpha = 0$, $\beta = 1/2$ in the situation (1.3), where $\{z : \text{Re } z \leq 0\} \subset S$, and V is a closed disk with center on the negative real axis and $0 \in \partial V$. These authors also proved (1.4) with $\alpha = 0$, $0 < \beta < 1$, for the case where $V = \{z : \text{Re } z \leq 0\} \subset S$ and $|\varphi(z)|$ is not identically 1 on the imaginary axis.

In Lenferink and Spijker [11, 1991] the estimate (1.4) was proved with $\alpha = \beta = 0$, for the case of a bounded, convex $V \subset \text{int}(S) \cup \{0\}$ with $V \subset W(\theta)$.

By Crouzeix et al. [2, 1993] and Palencia [15, 1995] interesting stability estimates were derived within the framework of holomorphic semigroups. These estimates, adapted so as to fit within our terminology, are of the form (1.4) with $\alpha = \beta = 0$. They are valid in the situation (1.3), where $V = W(\theta) \subset W(\theta') \subset S$, $0 \leq \theta < \theta' < \pi/2$.

While completing this article the authors learned that estimates with $\alpha = 0$, $0 \leq \beta < 1$ were proved by Giles [4, 1996] for cases where $\varphi(z)$ is a polynomial and $V \subset \text{int}(S) \cup \{0\}$.

1.5 Contents of the rest of the paper.

Section 2 deals with the order of contact between two sets. Theorem 2.3 states that the contact between $\text{cl}[\varphi(V)]$ and the unit circle $\{\zeta : |\zeta| = 1\}$ is of the same order as the contact between ∂V and ∂S. This result will be essential for our proof, in Chapter 3, of general stability estimates belonging to Category 2. Section 2 concludes with Lemma 2.4, which is useful in applications where the actual order of contact between ∂V and ∂S has to be determined.

Section 3.1 presents the main results of the paper, formulated in the Theorems 3.2 and 3.4. The former of these two theorems provides us, for any norm $\|\cdot\|$, with the estimate (1.4), both with $\alpha = 1$, $\beta = 0$ and $\alpha = 0$, $\beta = 1$, in the general case of arbitrary φ and $V \subset S$. This theorem covers, and largely extends, all of the results belonging to Category 1 reviewed in Section 1.4. Theorem 3.4 deals with the case where the contact between ∂V and ∂S is of order $q \in [0, \infty)$. It sharpens Theorem 3.2 in that it provides us with a general error estimate, belonging to Category 2, with $\alpha = 0$, $\beta = 1 - (1+q)^{-1} < 1$. The Lemmas 3.1, 3.3, formulated in Section 3.1, are essential for the proofs of the Theorems 3.2 and 3.4, respectively. These two lemmas are proved in Section 3.2. The proof of Lemma 3.3 makes use of Theorem 2.3.

Chapter 4 deals with examples and applications of the results of Section 3.1. In Section 4.1 we illustrate the Theorems 3.2, 3.4. We establish actual stability estimates in situations which are not covered by the results from the literature reviewed in Section 1.4. Section 4.2 describes numerical experiments pertinent to an initial-boundary value problem for a partial differential equation. These experiments confirm some conclusions of Section 4.1.

2 The order of contact between two sets.

Let q be a given real constant, $0 \leq q < \infty$, and let X, Y denote closed subsets of $\mathbb{C}$. We shall deal with the situation where, for all $x \in X$ which are close to

some $x_0 \in \mathbb{C}$, the distance between x and Y is not smaller than a multiple of $|x - x_0|^{1+q}$. This amounts to the requirement that positive α_0, α_1 exist such that

$$\text{(2.1a)} \qquad \mathrm{d}(x, Y) \geq \alpha_1 |x - x_0|^{1+q} \quad \text{whenever } x \in X, \quad |x - x_0| \leq \alpha_0.$$

In order to be able to deal also with the case where $x_0 = \infty$ it is convenient to introduce the set

$$Y^{-1} = \{y^{-1} : \ y \in Y, \ y \neq 0\},$$

and to consider the requirement that positive α_0, α_1 exist such that

$$\text{(2.1b)} \qquad \mathrm{d}(x^{-1}, Y^{-1}) \geq \alpha_1 |x|^{-1-q} \quad \text{whenever} \quad 0 \neq x \in X, \quad |x^{-1}| \leq \alpha_0.$$

Note that if $x_0 \notin X \cap Y$, condition (2.1a) is always fulfilled for α_0, $\alpha_1 > 0$ sufficiently small. Similarly, if X or Y is bounded, condition (2.1b) is automatically fulfilled for α_0, $\alpha_1 > 0$ sufficiently small.

DEFINITION 2.1.

a. The set X has contact of order q, with Y, at the point $x_0 \in \mathbb{C}$ if there exist positive α_0, α_1 such that (2.1a) *is fulfilled.*

b. The set X has contact of order q, with Y, at the point ∞ if there exist positive α_0, α_1 such that (2.1b) *is fulfilled.*

c. The contact between the sets X and Y is of order q if the intersection $X \cap Y$ contains at most a finite number of points, X has contact of order q with Y at all of these points, and X has contact of order q with Y at the point ∞.

Note that, if X has contact of order p, with Y, at the point $x_0 \in \mathbb{C} \cup \{\infty\}$, then X has also contact of order q, with Y, at the point x_0, for any $q \geq p$. A similar remark applies to the situation where the contact between the sets X and Y is of order p.

Throughout the present chapter we denote by $\varphi(z)$ an arbitrary non-constant rational function, and we define S as in Section 1.3. In the following we shall use Definition 2.1 repeatedly, with Y equal to ∂S or to the unit circle $|\zeta| = 1$.

The following Theorem 2.3 provides conditions under which, for a given set $V \subset S$, the contact between $\mathrm{cl}[\varphi(V)]$ and the unit circle is of order q. This theorem will be essential for the proof of stability estimates in Chapter 3.

The proof of Theorem 2.3 makes use of the subsequent Lemma 2.2. In that lemma we refer to the rational function $f(z)$ defined by

$$\text{(2.2)} \qquad f(z) = \varphi(1/z),$$

and for any set $Q \subset \mathbb{C}$ we use the notation

$$\varphi^{-1}(Q) = \{z : \ z \in \mathbb{C} \text{ with } \varphi(z) \in Q\}.$$

LEMMA 2.2. *Let X and Y be closed subsets of $\mathbb{C}$ which contain no poles of φ. Assume $\varphi^{-1}(\varphi(Y)) = Y$, and let $\xi_0 \in \mathbb{C}$. Denote the elements $x \in \mathbb{C} \cup \{\infty\}$, with $\varphi(x) = \xi_0$, by*

$$x_1, x_2, \ldots, x_m.$$

Assume that X has contact of order p_i, with Y, at the points x_i, where $0 \leq p_i < \infty$, $1 \leq i \leq m$. Denote, for $1 \leq i \leq m$, by ℓ_i the smallest integer with $\ell_i \geq 1$ such that

$$\varphi^{(\ell_i)}(x_i) \neq 0 \quad (\text{if } |x_i| < \infty), \qquad \text{and} \qquad f^{(\ell_i)}(0) \neq 0 \quad (\text{if } |x_i| = \infty).$$

Then the set $\mathrm{cl}[\varphi(X)]$ has contact of order

$$p = \max_{1 \leq i \leq m} p_i/\ell_i,$$

with $\mathrm{cl}[\varphi(Y)]$, at the point ξ_0.

PROOF. This lemma immediately follows from the material in Spijker and Straetemans [22, 1996]. □

THEOREM 2.3. *Let q be given, with $0 \leq q < \infty$. Assume V is a closed subset of S, such that the contact between ∂V and ∂S is of order q. Then the contact between $\mathrm{cl}[\varphi(V)]$ and the unit circle $|\zeta| = 1$ is also of order q.*

PROOF. Since the contact between ∂V and ∂S is of order q,

(2.3a) $\partial V \cap \partial S$ contains at most a finite number of points.

Defining $Q = \{\zeta : \zeta \in \mathbb{C} \text{ with } |\zeta| = 1\}$ we see that also

(2.3b) $\mathrm{cl}[\varphi(\partial V)] \cap Q$ contains at most a finite number of points.

Let $\xi_0 \in \mathrm{cl}[\varphi(\partial V)] \cap Q$.

The sets $X = \partial V$ and $Y = \partial S$ satisfy the assumptions of Lemma 2.2, with $p_i = q$ and $\ell_i \geq 1$. By that lemma the set $\mathrm{cl}[\varphi(\partial V)]$ has contact of order

$$p = \max_{1 \leq i \leq m} q/\ell_i,$$

with $\mathrm{cl}[\varphi(\partial S)] = Q$, at ξ_0.

Since $p \leq q$, the set $\mathrm{cl}[\varphi(\partial V)]$ has also contact of order q, with the unit circle $|\zeta| = 1$, at ξ_0. In view of (2.3b) the contact between $\mathrm{cl}[\varphi(\partial V)]$ and Q is of order q. Since $\partial\{\mathrm{cl}[\varphi(V)]\} \subset \mathrm{cl}[\varphi(\partial V)]$, the contact between $\partial\{\mathrm{cl}[\varphi(V)]\}$ and Q is also of order q. From this it can be seen by simple geometrical arguments, as formulated, e.g., in Theorem 2.2 of Spijker and Straetemans [21, 1996]), that also the contact between $\mathrm{cl}[\varphi(V)]$ and Q is of order q. □

The last result that we state in this section is Lemma 2.4. It is useful in applications, where one has to determine an actual value $q \in [0, \infty)$ for which the assumptions of Theorem 2.3 are satisfied. The lemma will be used in Section 4.2. For its proof we refer to Spijker and Straetemans [22, 1996]).

LEMMA 2.4. *Let q be a nonnegative integer, and $T > 0$. Let $F(t)$, $G(t)$ denote complex valued functions which are $q+1$ times continuously differentiable for $-T \leq t \leq T$. Define the sets X, Y by*

$$X = \{F(t) : -T \leq t \leq T\}, \quad Y = \{G(t) : -T \leq t \leq T\}.$$

Let

$$F(0) = G(0) = x_0.$$

Assume $\theta \in \mathbb{R}$ *to be such that, for all* $t \in [-T, T]$,

$$F(t) = x_0 + e^{i\theta}(t + if(t)), \quad G(t) = x_0 + e^{i\theta}(t + ig(t)),$$

where $f(t)$, $g(t)$ *are real valued functions satisfying*

$$f^{(j)}(0) = g^{(j)}(0) \quad \text{for } 1 \le j \le q, \quad f^{(1+q)}(0) \ne g^{(1+q)}(0).$$

Then X *has contact of order* q, *with* Y, *at* x_0.

3 Stability estimates under the Kreiss resolvent condition.

3.1 Formulation and proof of the main theorems.

Throughout this chapter $\varphi(z)$ stands for a given, non-constant rational function. Further, S is defined as in Section 1.3.

In all of the following, $\|\cdot\|$ denotes an arbitrary norm on the vector space $\mathbb{C}^{s,s}$ of all complex $s \times s$ matrices.

Below we establish the main results of the paper, Theorems 3.2 and 3.4. Both theorems provide us with a stability estimate of the form (1.4) for the numerical process (1.1), under the Kreiss resolvent condition (1.3).

In case V is bounded, we see from (1.3), by letting $z \to \infty$, that

$$K \ge \|I\|.$$

In the following K always denotes an arbitrary constant satisfying this inequality.

Our proof of Theorem 3.2 relies on a combination of the subsequent Lemma 3.1 with a stability result formulated in Dorsselaer, Kraaijevanger and Spijker [3, 1993], whereas the proof of Theorem 3.4 relies on a combination of the subsequent Lemma 3.3 with a stability result from Spijker and Straetemans [21, 1996].

Lemma 3.1 and 3.3 provide conditions under which the matrix $B = \varphi(hA)$ satisfies a resolvent condition of the form

(3.1) $\zeta I - B$ is invertible and $\|(\zeta I - B)^{-1}\| \le L \cdot \mathrm{d}(\zeta, W)^{-1}$ for all $\zeta \in \mathbb{C} \backslash W$.

Here W denotes a closed subset of the unit disk, to be specified below.

LEMMA 3.1. *Let* W *denote the unit disk. Then there is a constant* c, *such that the matrix* $B = \varphi(hA)$ *satisfies* (3.1), *with* $L = cK$, *whenever* hA *satisfies the resolvent condition* (1.3) *with respect to a closed subset* V *of* S. *Here* c *only depends on* φ (*and not on* V, $s \ge 1$, $hA \in \mathbb{C}^{s,s}$, $\|\cdot\|$ *or* $K \ge \|I\|$).

PROOF. See Section 3.2. □

Our first main theorem is as follows:

THEOREM 3.2. *There is a constant* γ *such that*

$$\|\varphi(hA)^n\| \le \gamma K \cdot \min\{s, n\},$$

whenever hA satisfies the resolvent condition (1.3) *with respect to a closed subset V of S. Here γ only depends on φ (and not on V, $n \geq 1$, $s \geq 1$, $hA \in \mathbb{C}^{s,s}$, $\|\cdot\|$ or $K \geq \|I\|$).*

PROOF. Let W be the unit disk, c as in Lemma 3.1, and suppose hA satisfies (1.3). We put $L = cK$ so that, by Lemma 3.1, the matrix $B = \varphi(hA)$ satisfies (3.1). Theorem 2.1 of Dorsselaer et al. [3, 1993] states that (3.1) implies the estimate $\|B^n\| \leq (1 + 1/n)^n L \cdot \min\{s, n+1\}$. That theorem is proved by using the integral representation

$$B^n = \frac{1}{2\pi \mathrm{i}} \int_\Gamma \zeta^n (\zeta I - B)^{-1} \mathrm{d}\zeta,$$

where Γ is the positively oriented circle $|\zeta| = 1 + 1/n$, and by estimating the norm of the integral in an appropriate fashion. For the details of that estimation we refer to the paper just mentioned.

The proof is completed by putting $\gamma = 2ec$. □

In Lemma 3.3 the set W has the following properties:

(3.2a) W is a closed subset of the unit disk $|\zeta| \leq 1$;

(3.2b) The contact between W and the unit circle $|\zeta| = 1$ is of order q.

LEMMA 3.3. *Let q be given, with $0 \leq q < \infty$. Assume V is a closed subset of S, such that the contact between ∂V and ∂S is of order q. Then there is a constant c and a set W, satisfying* (3.2), *such that the matrix $B = \varphi(hA)$ satisfies* (3.1), *with $L = cK$, whenever hA satisfies* (1.3). *Here c and W only depend on q, V, φ (and not on $s \geq 1$, $hA \in \mathbb{C}^{s,s}$, $\|\cdot\|$ or $K \geq \|I\|$).*

PROOF. See Section 3.2. □

Our second main theorem is as follows.

THEOREM 3.4. *Let q be given, with $0 \leq q < \infty$. Assume V is a closed subset of S such that the contact between ∂V and ∂S is of order q.*

Then there is a constant γ such that

$$\|\varphi(hA)^n\| \leq \gamma K n^p, \quad \textit{with} \quad p = 1 - 1/(1+q),$$

whenever hA satisfies (1.3). *Here γ only depends on q, V, φ (and not on $n \geq 1$, $s \geq 1$, $hA \in \mathbb{C}^{s,s}$, $\|\cdot\|$ or $K \geq \|I\|$).*

PROOF. Let c and W be as in Lemma 3.3, and suppose hA satisfies (1.3). We put $L = cK$ so that, by Lemma 3.3, the matrix $B = \varphi(hA)$ satisfies (3.1).

Theorem 3.1 of Spijker and Straetemans [21, 1996] states that there is a constant β, depending only on W and q, such that (3.1), (3.2) imply the estimate $\|B^n\| \leq \beta L n^p$ with $p = 1 - 1/(1+q)$. That theorem is proved by using the integral representation

$$B^n = \frac{1}{2\pi \mathrm{i}} \int_\Gamma \zeta^n (\zeta I - B)^{-1} \mathrm{d}\zeta,$$

where Γ is a positively oriented Jordan curve with W in its interior. The curve Γ depends on n, and is chosen in such a way that the norm of the corresponding integral does not exceed $2\pi\beta Ln^p$. For the details we refer to the paper just mentioned.

The proof is completed by putting $\gamma = \beta c$. □

3.2 The proofs of the above lemmas.

In this section the proofs of Lemma 3.1 and 3.3 will be given.

We shall first prove Lemma 3.3 by making use of Theorem 2.3, of material of Spijker and Straetemans [21, 1996] pertinent to sets W satisfying (3.2), and of a partial fraction decomposition of $\psi_\zeta(z) = (\zeta - \varphi(z))^{-1}$ which has similarity to a neat decomposition introduced by Lubich and Nevanlinna [12, 1991, p. 304].

Next, the proof of Lemma 3.1 will be given. It can be viewed as a modified version of the proof of Lemma 3.3. In proving Lemma 3.1 we shall refer to the proof of Lemma 3.3, and confine ourselves to indicating the necessary modifications.

3.2.1 The proof of Lemma 3.3.

1. We shall use, along with (3.1), the following (apparently weaker) resolvent condition (3.1′).

$$\zeta I - B \text{ is invertible for all } \zeta \in \mathbb{C}\backslash W, \text{ and}$$
$$\|(\zeta I - B)^{-1}\| \leq L_1 \cdot \mathrm{d}(\zeta, W)^{-1} \text{ for all } \zeta \text{ with } 0 < \mathrm{d}(\zeta, W) \leq 1.$$

It will turn out to be sufficient to prove (3.1′), instead of (3.1).

We enlarge the set $\partial V \cap \partial S$ by the adjunction of ∞, if ∂V and ∂S are unbounded. We denote the elements of this enlarged set by $x_1, x_2, \ldots, x_r$, and define $\xi_j = \varphi(x_j)$. In view of Theorem 2.3 the set $X = \mathrm{cl}[\varphi(V)]$ has the following three properties:

$$X \text{ is a closed subset of the unit disk } |\zeta| \leq 1;$$
$$\text{The contact between } X \text{ and the unit circle } |\zeta| = 1 \text{ is of order } q;$$
$$\{\xi : \xi \in X \text{ and } |\xi| = 1\} = \{\xi_1, \xi_2, \ldots, \xi_r\}.$$

According to Spijker and Straetemans [21, 1996] (Theorems 2.2, 2.4) these three properties ensure the existence of a set W and a constant $c_1 > 0$ such that

(3.3a) W is a closed subset of the unit disk $|\zeta| \leq 1$;

(3.3b) The contact between W and the unit circle $|\zeta| = 1$ is of order q;

(3.3c) $X \subset W$ and $X \cap \partial W = \{\xi_1, \xi_2, \ldots, \xi_r\}$;

(3.3d) Condition (3.1) is fulfilled, with $L = c_1 L_1$ whenever (3.1′) holds.

The properties (3.3a,b,c) are proved by simple geometrical arguments, and property (3.3d) is proved by using an integral representation for $(\zeta I - B)^{-1}$ which has some similarity to the representation used below in Section 3.2.2

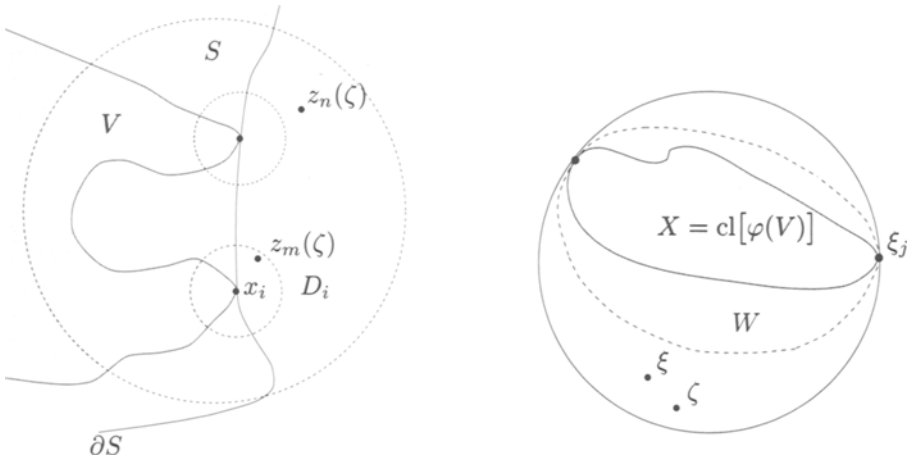

Figure 3.1: The situation in the proof of Lemma 3.3.

(Part 1). For the details of the proof of (3.3) we refer to the above mentioned paper.

We shall prove the lemma with the set W satisfying (3.3).

2. Let hA satisfy (1.3), and put $B = \varphi(hA)$. In view of (3.3d) it is sufficient to show the existence of a factor c_0 (only depending on q, V and φ) such that (3.1′) holds with $L_1 = c_0 K$.

Since the spectrum of B lies in $\varphi(V)$, the matrix $\zeta I - B$ is invertible for all $\zeta \in \mathbb{C}\backslash W$, as required in (3.1′).

We define the set

$$W_1 = \{\zeta : \ 0 < \mathrm{d}(\zeta, W) \leq 1\}.$$

In view of the above we see, by using the compactness of the set $\mathrm{cl}(W_1)$, that it is sufficient to prove

$$\|(\zeta I - B)^{-1}\| \leq \gamma_0 K \mathrm{d}(\zeta, W)^{-1} \tag{3.4}$$

for all $\zeta \in W_1$ belonging to some neighborhood of each individual $\xi \in \mathrm{cl}(W_1)$. Here γ_0 should depend only on q, V, φ and ξ.

In the rest of the proof ξ is fixed, with $\xi \in \mathrm{cl}(W_1)$.

3. In order to prove (3.4) we introduce, for any given $\zeta \in \mathbb{C}$, the rational function

$$\psi_\zeta(z) = (\zeta - \varphi(z))^{-1}.$$

We denote the poles of $\psi_\zeta(z)$ by $z_j(\zeta)$, where j runs through some index set $J(\zeta)$.

Let $\varepsilon > 0$ be given, and define, for $1 \leq i \leq r$,

$$\begin{aligned} D_i &= \{z : \ z \in \mathbb{C} \ \text{ with } \ |z - x_i| < \varepsilon\} \quad \text{if } \ |x_i| < \infty, \\ D_i &= \{z : \ z \in \mathbb{C} \ \text{ with } \ |z| > 1/\varepsilon\} \quad \text{if } \ |x_i| = \infty, \end{aligned}$$

and

$$D = \bigcup_{i=1}^{r} D_i.$$

We decompose $J(\zeta)$ into $J(\zeta) = M(\zeta) \cup N(\zeta)$ such that

$$\begin{aligned} z_m(\zeta) &\in D \qquad \text{for all } m \in M(\zeta), \\ z_n(\zeta) &\in \mathbb{C}\backslash D \qquad \text{for all } n \in N(\zeta). \end{aligned}$$

We denote the sum of the principal parts of $\psi_\zeta(z)$ at the poles $z_m(\zeta)$ (with $m \in M(\zeta)$) by $\mu_\zeta(z)$, and following an idea of Lubich and Nevanlinna [12, 1991, p. 304] we consider the partial fraction decomposition

$$\psi_\zeta(z) = \mu_\zeta(z) + \nu_\zeta(z).$$

Since $(\zeta I - B)^{-1} = \mu_\zeta(hA) + \nu_\zeta(hA)$ (for $\zeta \in \mathbb{C}\backslash W$) we obtain

$$\|(\zeta I - B)^{-1}\| \leq \|\mu_\zeta(hA)\| + \|\nu_\zeta(hA)\| \quad \text{for all } \zeta \in W_1. \tag{3.5}$$

4. We shall bound the first term in the right-hand member of (3.5). Denote, for $1 \leq i \leq r$, by ℓ_i the smallest integer, with $\ell_i \geq 1$ such that

$$\begin{cases} \varphi^{(\ell_i)}(x_i) \neq 0 & (\text{if } |x_i| < \infty), \\ f^{(\ell_i)}(0) \neq 0 & (\text{if } |x_i| = \infty), \end{cases}$$

where f is defined by (2.2). We choose $\varepsilon > 0$ such that the boundary of D contains no poles $z_j(\xi)$ ($j \in J(\xi)$), and such that

$$|\varphi^{(\ell_i)}(z) - \varphi^{(\ell_i)}(x_i)| \leq \tfrac{1}{2}|\varphi^{(\ell_i)}(x_i)| \quad \text{for } |z - x_i| \leq 2\varepsilon \ (\text{if } |x_i| < \infty),$$

and

$$|f^{(\ell_i)}(z^{-1}) - f^{(\ell_i)}(0)| \leq \tfrac{1}{2}|f^{(\ell_i)}(0)| \quad \text{for } |z| \geq 1/(2\varepsilon) \ (\text{if } |x_i| = \infty).$$

Let $\zeta \in W_1$ be given. Below we shall see that $\varphi'(z_m(\zeta)) \neq 0$ ($m \in M(\zeta)$), so that

$$\mu_\zeta(hA) = \sum_{m \in M(\zeta)} [\varphi'(z_m(\zeta))]^{-1} \cdot [z_m(\zeta) I - hA]^{-1}.$$

In view of (1.3) we thus have

$$\|\mu_\zeta(hA)\| \leq K \sum_{m \in M(\zeta)} \sigma_m, \ \text{with } \sigma_m = |\varphi'(z_m(\zeta))|^{-1} \cdot |z_m(\zeta) - x|^{-1}, \tag{3.6a}$$

where

$$x \in V, \ \text{with } |z_m(\zeta) - x| = \mathrm{d}(z_m(\zeta), V).$$

In the following we prove

$$\sigma_m \leq 3 \cdot 2^{\ell} \cdot \mathrm{d}(\zeta, W)^{-1}, \ \text{with } \ell = \max_{1 \leq i \leq r} \ell_i, \tag{3.6b}$$

by distinguishing three cases.

Case 1. Assume $z_m(\zeta) \in D_i$, $|x_i| < \infty$.
The proof of (3.6b) will essentially consist in first bounding σ_m in terms of $|\varphi(z_m(\zeta)) - \varphi(x)|^{-1}$, and next relating the latter quantity to $\mathrm{d}(\zeta, W)^{-1}$. Bounding σ_m suitably is straightforward when $\ell_i = 1$, but requires the following (more involved) arguments in the general case $\ell_i \geq 1$.

Expanding the derivative $\varphi'(z)$ around x_i, by means of Taylor's formula with integral representation for the remainder, one can see that for $|z - x_i| < 2\varepsilon$,

$$\varphi'(z) = \frac{(z - x_i)^{\ell_i - 1}}{(\ell_i - 1)!} \varphi^{(\ell_i)}(x_i)(1 + \delta),$$

where $|\delta| \leq 1/2$. Applying this formula, with $z = z_m(\zeta)$, we obtain

$$|\varphi'(z_m(\zeta))| \geq \tfrac{1}{2} \cdot \frac{|\varphi^{(\ell_i)}(x_i)|}{(\ell_i - 1)!} \cdot |z_m(\zeta) - x_i|^{\ell_i - 1} > 0.$$

Since

$$|z_m(\zeta) - x_i| < \varepsilon \quad \text{and} \quad |x - x_i| \leq 2|z_m(\zeta) - x_i| < 2\varepsilon,$$

we have

$$\varphi(z_m(\zeta)) - \varphi(x) = (z_m(\zeta) - x) \int_0^1 \varphi'(v(\tau))\, \mathrm{d}\tau,$$

where $v(\tau) = x + \tau(z_m(\zeta) - x)$ satisfies $|v(\tau) - x_i| < 2\varepsilon$. Applying the above formula for $\varphi'(z)$, with $z = v(\tau)$, we arrive at

$$\varphi(z_m(\zeta)) - \varphi(x) = \frac{\varphi^{(\ell_i)}(x_i)}{(\ell_i - 1)!} (z_m(\zeta) - x) \int_0^1 [v(\tau) - x_i]^{\ell_i - 1} (1 + \delta(\tau))\, \mathrm{d}\tau,$$

where $|\delta(\tau)| \leq 1/2$. Consequently,

$$|\varphi(z_m(\zeta)) - \varphi(x)| \leq \tfrac{3}{2} \cdot \frac{|\varphi^{(\ell_i)}(x_i)|}{(\ell_i - 1)!} \cdot |z_m(\zeta) - x| \cdot M^{\ell_i - 1},$$

with

$$M = \max\{|z_m(\zeta) - x_i|,\ |x - x_i|\} \ \leq 2|z_m(\zeta) - x_i|.$$

It follows that

$$|\varphi'(z_m(\zeta))| \cdot |z_m(\zeta) - x| \geq 3^{-1} \cdot 2^{-\ell_i + 1} \cdot |\varphi(z_m(\zeta)) - \varphi(x)|.$$

By combining this inequality with the fact that

$$\mathrm{d}(\zeta, W) \leq \mathrm{d}(\zeta, \mathrm{cl}[\varphi(V)]) \leq |\zeta - \varphi(x)| = |\varphi(z_m(\zeta)) - \varphi(x)|,$$

we obtain $\sigma_m \leq 3 \cdot 2^{\ell_i - 1} \cdot \mathrm{d}(\zeta, W)^{-1}$. Relation (3.6b) has thus been proved.

Case 2. Assume $z_m(\zeta) \in D_i$, $|x_i| = \infty$ *and* $|x| \geq |z_m(\zeta)|/2$.

Since

$$|z_m(\zeta)| > 1/\varepsilon \quad \text{and} \quad |x| \geq |z_m(\zeta)|/2 > 1/(2\varepsilon)$$

we see, by arguments similar to those used in Case 1, that

$$\begin{aligned} |\varphi'(z_m(\zeta))| &\geq \frac{1}{2}\cdot\frac{|f^{(\ell_i)}(0)|}{(\ell_i-1)!}\cdot|z_m(\zeta)|^{-1-\ell_i} > 0, \\ |\varphi(z_m(\zeta)) - \varphi(x)| &\leq \frac{3}{2}\cdot\frac{|f^{(\ell_i)}(0)|}{(\ell_i-1)!}\cdot|z_m(\zeta)^{-1} - x^{-1}|\cdot M^{\ell_i-1}, \end{aligned}$$

with

$$M = \max\{|z_m(\zeta)^{-1}|,\ |x^{-1}|\} \leq 2|z_m(\zeta)^{-1}|.$$

It follows that

$$|\varphi'(z_m(\zeta))|\cdot|z_m(\zeta) - x| \geq 3^{-1}\cdot 2^{-\ell_i}\cdot|\varphi(z_m(\zeta)) - \varphi(x)|.$$

Similarly as in Case 1 we obtain (3.6b).

Case 3. Assume $z_m(\zeta) \in D_i$, $|x_i| = \infty$ *and* $|x| < |z_m(\zeta)|/2$.
In this case we use the relations

$$\begin{aligned} |\varphi'(z_m(\zeta))| &\geq \frac{1}{2}\cdot\frac{|f^{(\ell_i)}(0)|}{(\ell_i-1)!}\cdot|z_m(\zeta)|^{-1-\ell_i} > 0, \\ |\varphi(z_m(\zeta)) - \varphi(\infty)| &\leq \frac{3}{2}\cdot\frac{|f^{(\ell_i)}(0)|}{\ell_i!}\cdot|z_m(\zeta)|^{-\ell_i}, \end{aligned}$$

so as to conclude that

$$|\varphi'(z_m(\zeta))|\cdot|z_m(\zeta)| \geq 3^{-1}\cdot\ell_i\cdot|\varphi(z_m(\zeta)) - \varphi(\infty)|.$$

The last inequality can be seen to imply (3.6b) by noting that $|z_m(\zeta)| < 2|z_m(\zeta) - x|$, and that

$$\mathrm{d}(\zeta, W) \leq \mathrm{d}(\zeta, \mathrm{cl}[\varphi(V)]) \leq |\zeta - \varphi(\infty)| = |\varphi(z_m(\zeta)) - \varphi(\infty)|.$$

We note that the number of poles $z_j(\zeta)$ does not exceed the maximum of the degrees of the numerator and the denominator of $\varphi(z)$. Let α_0 denote this maximum multiplied by $3\cdot 2^{\ell}$. In view of (3.6a,b) we have

$$\|\mu_\zeta(hA)\| \leq \alpha_0\cdot K\cdot\mathrm{d}(\zeta, W)^{-1} \quad \text{for all} \quad \zeta \in W_1. \tag{3.6c}$$

5. We shall prove the inequality

$$\|\nu_\zeta(hA)\| \leq \beta_0 K\cdot\mathrm{d}(\zeta, W)^{-1}, \tag{3.7}$$

for all $\zeta \in W_1$ belonging to some neighborhood Δ of ξ, where β_0 only depends on q, V, φ and ξ. This inequality, combined with (3.5), (3.6c), proves (3.4) (with $\gamma_0 = \alpha_0 + \beta_0$) and will thus complete the proof of the lemma.

We have for $\zeta \in \mathbb{C}$, with $\zeta \neq \varphi(\infty)$, the integral representation

$$\nu_\zeta(z) = \frac{1}{2\pi i}\int_C \psi_\zeta(y)(y-z)^{-1}\,\mathrm{d}y + (\zeta - \varphi(\infty))^{-1} \qquad (\text{for all } z \in V),$$

where C is any curve with the following properties:

(3.8a) C is a finite sum of negatively oriented circles with interiors E_j which are mutually disjoint;

(3.8b) The disks $\mathrm{cl}(E_j)$ do not intersect with V or with $\mathrm{cl}(D)$;

(3.8c) $z_n(\zeta) \in \bigcup_j E_j$ (for all $n \in N(\zeta)$).

The above representation for $\nu_\zeta(z)$ is also valid for $\nu_\zeta(hA)$, provided $(y-z)^{-1}$ and $(\zeta-\varphi(\infty))^{-1}$ are replaced by $(yI-hA)^{-1}$ and $(\zeta-\varphi(\infty))^{-1}I$, respectively. Therefore, in view of (1.3) and the inequality $K \geq \|I\|$, we see that (3.8a,b,c) imply

$$\text{(3.8d)} \quad \frac{1}{K}\|\nu_\zeta(hA)\| \leq |\zeta-\varphi(\infty)|^{-1} + \frac{1}{2\pi}\int_C |\mathrm{d}y| \cdot \max_{y\in C} |\zeta-\varphi(y)|^{-1} \cdot \mathrm{d}(y,V)^{-1}.$$

In order to construct a curve C satisfying (3.8a,b,c) it is important to note that

$$\text{(3.9)} \qquad z_n(\xi) \text{ lies outside of } V \text{ and of } \mathrm{cl}(D) \quad (\text{ for all } n \in N(\xi)).$$

This property follows from our choice of $\varepsilon > 0$, the definition of $N(\xi)$ and from (3.3c) (the supposition $z_n(\xi) \in V$ would imply $\xi \in \varphi(V) \cap \mathrm{cl}(W_1) \subset \{\zeta : |\zeta| = 1\}$ and $z_n(\xi) \in V \cap \partial S$, which is impossible).

Below we prove (3.7) by distinguishing three cases.

Case i. Assume $\xi \neq \varphi(\infty)$. In view of (3.9) we can construct negatively oriented circles C_n, with centers $z_n(\xi)$ (for $n \in N(\xi)$) such that $C = \sum_n C_n$ satisfies (3.8a,b). There is a compact neighborhood Δ of ξ such that, for all $\zeta \in \Delta$, we have (3.8c) as well as $|\zeta-\varphi(\infty)|^{-1} \leq 2|\xi-\varphi(\infty)|^{-1}$. For these ζ we thus obtain, by (3.8d), the inequality $K^{-1}\|\nu_\zeta(hA)\| \leq 2|\xi-\varphi(\infty)|^{-1} + \beta$, where

$$\text{(3.10)} \quad \beta = (2\pi)^{-1}\int_C |\mathrm{d}y| \cdot \max\{|\zeta-\varphi(y)|^{-1} \cdot \mathrm{d}(y,V)^{-1} : y \in C,\ \zeta \in \Delta\} < \infty.$$

Hence (3.7) holds, for $\zeta \in \Delta$, with $\beta_0 = 2|\xi-\varphi(\infty)|^{-1} + \beta$.

Case ii. Assume $\xi = \varphi(\infty)$*; V unbounded.* We construct C_n and C as in case (i). Clearly $\xi \in \mathrm{cl}[\varphi(V)] \cap \mathrm{cl}(W_1)$ so that, by (3.3), $|\xi| = 1$. Hence $|\varphi(\infty)| = 1$, and ∂S is unbounded. One of the sets D_i thus is of the form

$$D_i = \{z : z \in \mathbb{C} \text{ with } |z| > 1/\varepsilon\}.$$

For $\zeta \simeq \xi$ all poles $z_j(\zeta)$ lie in this D_i or close to some $z_j(\xi)$, so that all $z_n(\zeta)$ $(n \in N(\zeta))$ must be close to some $z_n(\xi)$ $(n \in N(\xi))$. Hence, there is a compact neighborhood Δ of ξ such that, for all $\zeta \in \Delta$, we have (3.8c). Applying (3.8d) to $\zeta \in W_1 \cap \Delta$ we obtain $K^{-1}\|\nu_\zeta(hA)\| \leq |\zeta-\varphi(\infty)|^{-1} + \beta$, where β satisfies (3.10). Hence (3.7) holds, for $\zeta \in W_1 \cap \Delta$, with $\beta_0 = 1 + \beta$.

Case iii. Assume $\xi = \varphi(\infty)$; V bounded. In this case we use, for any $\zeta \in \mathbb{C}$, the integral representation

$$\nu_\zeta(z) = \frac{1}{2\pi i} \int_C \psi_\zeta(y)(y - z)^{-1} \mathrm{d}y \qquad \text{(for all } z \in V),$$

where now C is any curve such that

(3.11a) C is the sum of one positively oriented circle and a finite number of negatively oriented circles. The interiors of the negatively oriented circles and the exterior of the positively oriented circle are denoted by E_j and are all mutually disjoint;

(3.11b) The (generalized) disks $\mathrm{cl}(E_j)$ do not intersect with V or with $\mathrm{cl}(D)$;

(3.11c) $z_n(\zeta) \in \bigcup_j E_j$ (for all $n \in N(\zeta)$).

The above representation carries over to $\nu_\zeta(hA)$, and (3.11a,b,c) thus imply an upperbound for $K^{-1}\|\nu_\zeta(hA)\|$ which is similar to the right-hand member of (3.8d) (without the term $|\zeta - \varphi(\infty)|^{-1}$).

We construct C satisfying (3.11a,b) such that $z_n(\xi) \in \cup_j E_j$ (for all $n \in N(\xi)$). There is a compact neighborhood Δ of ξ such that, for all $\zeta \in \Delta$, we have (3.11c). Applying to these ζ our upperbound for $K^{-1}\|\nu_\zeta(hA)\|$ we obtain $K^{-1}\|\nu_\zeta(hA)\| \leq \beta$, with β given by (3.10). Hence, (3.7) holds, for $\zeta \in \Delta$, with $\beta_0 = \beta$. This completes the proof of Lemma 3.3. □

3.2.2 The proof of Lemma 3.1.

1. Assume, without loss of generality, that $V = S$, and let W denote the unit disk. We shall use, similarly as in the proof of Lemma 3.3, the resolvent condition (3.1′). By using arguments stated in Reddy and Trefethen [17, 1992, Chapter 7] we shall prove that (3.3d) holds, with $c_1 = 6$.

In order to prove (3.3d) we assume (3.1′) and we consider an arbitrary $\zeta \in \mathbb{C}$ with $|\zeta| > 2$. We have

$$(\zeta I - B)^{-1} = \frac{1}{2\pi \mathrm{i}} \int (\zeta - \xi)^{-1}(\xi I - B)^{-1} \, \mathrm{d}\xi,$$

where the integration is along the positively oriented circle $|\xi| = 3/2$. By taking norms, there follows

$$\|(\zeta I - B)^{-1}\| \leq 6L_1(2|\zeta| - 3)^{-1} \leq 6L_1(|\zeta| - 1)^{-1},$$

which proves (3.3d).

We enlarge the set $\{z : z \in \partial S, \varphi'(z) = 0\}$ with the adjunction of ∞ if ∂S is unbounded, and we denote the elements of this (enlarged) set by x_j ($j = 1, 2, \ldots, r$).

2. This part of the proof is analogous to Part 2 of Section 3.2.1. Now the constant γ_0 in (3.4) should only depend on φ and ξ.

3. We define $\psi_\zeta(z)$, $z_j(\zeta)$ and D_i $(1 \le i \le r)$ as in Part 3 of Section 3.2.1, but we define the set D differently by

$$D = D_0 \cup D_1 \cup \cdots \cup D_r,$$

where

$$D_0 = \{z : \ z \in \mathbb{C} \ \text{ with } \ \mathrm{d}(z, \partial S) < \varepsilon \ \text{ and } \ z \notin D_i \ (1 \le i \le r)\}.$$

Now we can decompose $\psi_\zeta(z)$, using our present set D, in a similar way as in Part 3 of Section 3.2.1 so as to obtain (3.5).

4. We define ℓ_i as in Part 4 of Section 3.2.1. The value $\varepsilon > 0$ is chosen according to the same part, and is now, additionally, required to be so small that

$$|\varphi(z)| < \infty, \quad \varphi'(z) \neq 0 \qquad \text{for } \ z \in \mathrm{cl}(D_0).$$

We again have (3.6a).

In case $z_m(\zeta) \in \bigcup_{i=1}^{r} D_i$ we have (3.6b) by the same arguments as in Section 3.2.1.

Assume $z_m(\zeta) \in D_0$. The quantities

$$\mu_0 = \inf\{|\varphi'(z)| : \ z \in D_0\}, \qquad \mu_1 = \sup\{|\varphi'(z)| : \ z \in D\},$$

satisfy $\mu_0 > 0$, $\mu_1 < \infty$. Further,

$$|\varphi'(z_m(\zeta))| \ge \mu_0, \ \text{ and } \ |\varphi(z_m(\zeta)) - \varphi(x)| \le \mu_1 |z_m(\zeta) - x|,$$

so that

$$\sigma_m \le \frac{\mu_1}{\mu_0} |\varphi(z_m(\zeta)) - \varphi(x)|^{-1} \le \frac{\mu_1}{\mu_0} \cdot \mathrm{d}(\zeta, W)^{-1}.$$

Similarly as in Section 3.2.1 we find a constant α_0, only depending on φ, such that (3.6c) holds.

5. This part of the proof is analogous to Part 5 of Section 3.2.1. Now the constant β_0 in (3.7) should only depend on φ and ξ.

In the present situation property (3.9) follows from our choice of $\varepsilon > 0$ and the definition of $N(\xi)$ (the supposition $z_n(\xi) \in V = S$ would imply $\xi \in \varphi(S) \cap \mathrm{cl}(W_1) \subset \{\zeta : \ |\zeta| = 1\}$ and $z_n(\xi) \in \partial S$, which is impossible).

Further, if $\xi = \varphi(\infty)$ and S is unbounded (Case ii of Section 3.2.1), we have $|\xi| = 1$ since $\xi \in \mathrm{cl}[\varphi(S)] \cap \mathrm{cl}(W_1) \subset \{\zeta : \ |\zeta| = 1\}$.

The proof is completed similarly as in Section 3.2.1. □

4 Examples and applications.

4.1 Illustrations to the main theorems.

In this section we illustrate our Theorems 3.2, 3.4 by applying them to the simple function $\varphi(z)$ given by

$$\varphi(z) = 1 + z + \tfrac{1}{2} z^2 + \omega z^3, \tag{4.1}$$

where ω denotes a real parameter. The corresponding stability region is denoted by S.

We first consider the situation where $V = S$ and hA is an arbitrary $s \times s$ matrix satisfying the resolvent condition (1.3). In this case Theorem 3.2 guarantees the existence of a constant γ such that the stability estimate

$$\|\varphi(hA)^n\| \leq \gamma \cdot K \cdot \min\{s, n\}$$

is valid. Here γ only depends on the above parameter ω (and not on $n \geq 1$, $s \geq 1$, hA or $K \geq \|I\|$).

The above stability estimate is valid in particular for the case where $\omega = 0.0625$. The value 0.0625 is of special interest since it equals the unique value of the parameter ω which maximizes the quantity

$$M(\omega) = \max\{\xi : \ \xi \in \mathbb{R} \ \text{ with } \ [-\xi, 0] \subset S\}$$

(see, e.g., Houwen [8, 1977]). For $\omega = 0.0625$ we have $M(\omega) = 6.2608$ (rounded to 5 decimal digits), and $\varphi'(\xi) = 0$ with $\xi = -4 \in \partial S$. We emphasize that, as $\varphi'(z)$ vanishes at some point of ∂S, the above stability estimate does not follow from any of the results in the literature mentioned in Section 1.4.

Next we consider the situation where V is a cigar-shaped region of the form

$$V(\rho, \sigma) = \{z : \ z = x + y \ \text{ with } \ x \in \mathbb{R}, \ y \in \mathbb{C}, \ -\sigma - \rho \leq x \leq -\rho, \ |y| \leq \rho\},$$

where $\rho > 0$, $\sigma > 0$ are given parameters (cf. Lenferink and Spijker [11, 1991]). We shall illustrate Theorem 3.4, with a set V of the above form, in the numerical solution of the initial-boundary value problem

$$(4.2) \qquad \frac{\partial}{\partial t}U(x,t) = \frac{\partial}{\partial x}\left(a(x)\frac{\partial}{\partial x}U(x,t)\right) + \frac{\partial}{\partial x}\Big(b(x)U(x,t)\Big) + c(x)U(x,t),$$

$$U(0,t) = U(1,t) = 0, \quad U(x,0) = U_0(x).$$

Here $0 \leq x \leq 1$, $t \geq 0$, and a, b, c, U_0 are given functions, with $a(x) > 0$, $c(x) \leq 0$, and $b(x) \geq b(y)$ for $x \leq y$.

We shall approximate the solution $U(x,t)$ to problem (4.2) by a numerical process which involves increments $\Delta t = h > 0$ and $\Delta x = (1+s)^{-1}$. Let $A = (\alpha_{i,j})_{i,j=1}^{s}$ be the tridiagonal $s \times s$ matrix with

$$\begin{aligned}
\alpha_{i,i-1} &= (\Delta x)^{-2} a_{i-1/2} + (2\Delta x)^{-1}\{-b_{i-1/2} + \varepsilon_{i-1/2}|b_{i-1/2}|\} \qquad (1 \leq i \leq s),\\
\alpha_{i,i} &= -\alpha_{i,i-1} - \alpha_{i,i+1} + (\Delta x)^{-1}\{-b_{i-1/2} + b_{i+1/2}\} + c_i \qquad (1 \leq i \leq s),\\
\alpha_{i,i+1} &= (\Delta x)^{-2} a_{i+1/2} + (2\Delta x)^{-1}\{b_{i+1/2} + \varepsilon_{i+1/2}|b_{i+1/2}|\} \qquad (1 \leq i \leq s).
\end{aligned}$$

Here $\alpha_{1,0}$, $\alpha_{s,s+1}$ are defined only for notational convenience, and $a_m = a(m\Delta x)$, $b_m = b(m\Delta x)$, $c_m = c(m\Delta x)$. Further, ε_m are so-called upwind parameters (cf. Griffiths et al. [5, 1980], Grossmann and Roos [6, 1992, Section 7.1], Noye [14, 1984, p. 226], Stoyan [23, 1984]), which are chosen such that

$$0 \leq \varepsilon_m \leq 1, \qquad 1 - 2a_m(|b_m|\Delta x)^{-1} \leq \varepsilon_m$$

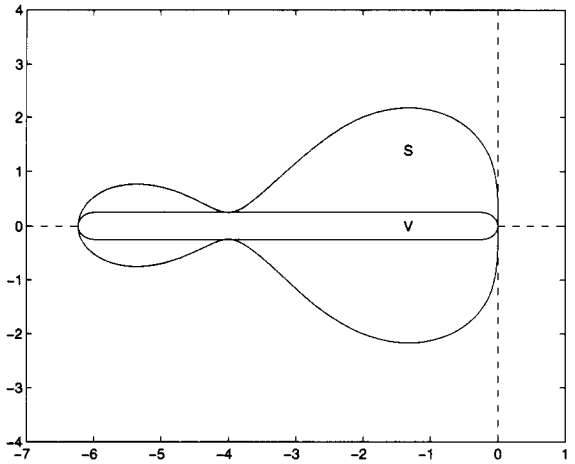

Figure 4.1: The sets $V = V(\rho, \sigma)$ and S for $\omega = 0.06275$.

(the last inequality guarantees the off-diagonal elements of the matrix A to be nonnegative). Put $f_n = 0$, and let $u_n \in \mathbb{C}^s$ be computed from (1.1) (with φ and A as specified above) starting with a vector $u_0 \in \mathbb{C}^s$, the i-th component of which equals $U_0(i\Delta x)$ $(1 \leq i \leq s)$. Then the i-th component of u_n provides us with a numerical approximation to the true value $U(x,t)$ for $x = i\Delta x$, $t = n\Delta t$, $1 \leq i \leq s$, $n \geq 1$.

We shall establish the Kreiss resolvent condition (1.3) using, for $B \in \mathbb{C}^{s,s}$, the norm

$$\|B\| = \max\{|Bx|/|x| : \ x \in \mathbb{C}^s \ \text{ with } \ x \neq 0\},$$

where $|\cdot|$ denotes the *maximum norm* defined by

$$|x| = \max_{1 \leq i \leq s} |\xi_i| \quad (\text{for } x \in \mathbb{C}^s \text{ with components } \xi_i).$$

We put

$$\alpha = \max_{1 \leq i \leq s} \{\alpha_{i,i-1} + \alpha_{i,i+1}\}, \qquad \beta = \max_{1 \leq i \leq s} \left|c_i + (\Delta x)^{-1}(-b_{i-1/2} + b_{i+1/2})\right|.$$

From well known properties of numerical ranges (see, e.g., Lenferink and Spijker [9, 1990], Spijker [20, 1993]) one easily concludes that the matrix hA satisfies condition (1.3), with constant $K = 1$, with respect to the set

$$V(h\alpha, h\beta).$$

Let $\rho > 0$, $\sigma > 0$ be quantities (depending on the parameter ω in (4.1)) such that

(4.3a) $V(\rho, \sigma) \subset S$;

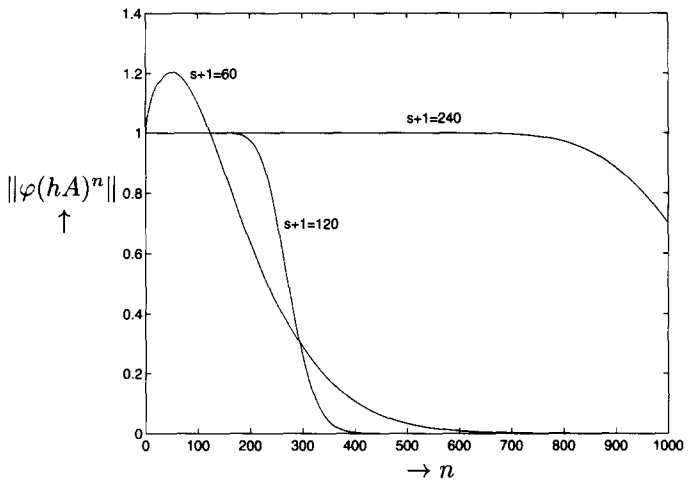

Figure 4.2: $\|\varphi(hA)^n\|$ as a function of n for $0 \leq n \leq 1000$, where φ is given by (4.1) with $\omega = 0.06275$. Values of $\|\varphi(hA)^n\|$, with $h < h_0$, have been plotted for:
$s+1 = 60$, with $h_0 = 2.1294 \times 10^{-5}$, $h = 2.1 \times 10^{-5}$;
$s+1 = 120$, with $h_0 = 8.8724 \times 10^{-6}$, $h = 8.8 \times 10^{-6}$;
$s+1 = 240$, with $h_0 = 2.2181 \times 10^{-6}$, $h = 2.2 \times 10^{-6}$.

(4.3b) The intersection of $\partial[V(\rho,\sigma)]$ and ∂S consists of a finite number of points $x_1, x_2, \ldots, x_r$;

(4.3c) At the points x_i the set $\partial[V(\rho,\sigma)]$ has contact of order 1 with ∂S (for $i = 1, 2, \ldots, r$).

We see that the stepsize restriction

$$0 < h \leq h_0 = \min\left\{\frac{\rho}{\alpha}, \frac{\sigma}{\beta}\right\} \tag{4.4}$$

implies the inclusion $V(h\alpha, h\beta) \subset V(\rho,\sigma)$. Clearly this inclusion proves that hA satisfies the Kreiss condition (1.3), with $K = 1$, also with respect to the set

$$V = V(\rho,\sigma).$$

By virtue of the Theorems 3.2, 3.4 we thus can conclude that, under the stepsize restriction (4.4), there is a nice stability behaviour in that

$$\|\varphi(hA)^n\| \leq \gamma \cdot \min\{s, \sqrt{n}\}$$

holds. Here γ only depends on the parameter ω (and not on $n \geq 1$, $s \geq 1$ or $h \in (0, h_0]$).

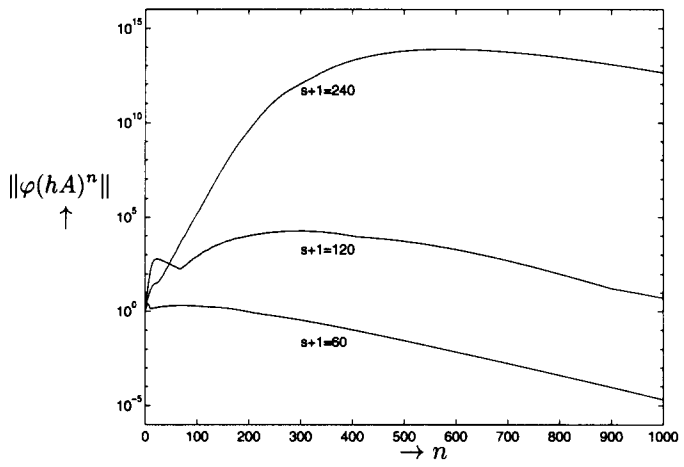

Figure 4.3: $\|\varphi(hA)^n\|$ as a function of n for $0 \leq n \leq 1000$, where φ is given by (4.1) with $\omega = 0.06275$. Values of $\|\varphi(hA)^n\|$, with $h < h_1$, have been plotted for:
$s+1 = 60$, with $h_1 = 2.9769 \times 10^{-5}$, $h = 2.9 \times 10^{-5}$;
$s+1 = 120$, with $h_1 = 2.6059 \times 10^{-5}$, $h = 2.6 \times 10^{-5}$;
$s+1 = 240$, with $h_1 = 1.5079 \times 10^{-5}$, $h = 1.5 \times 10^{-5}$.

4.2 A numerical illustration.

In the following we illustrate the stepsize restriction (4.4) by considering a situation which is a variant to one dealt with in Lenferink and Spijker [11, 1991]. We define the function $\varphi(z)$ by (4.1), with $\omega = 0.06275$. We have verified, by invoking Lemma 2.4, that the conditions (4.3) are now fulfilled with $r = 4$ and with $\rho = 0.25553$, $\sigma = 5.7205$ (rounded to 5 decimal digits). Figure 4.1 explains the geometrical situation.

We deal with problem (4.2), where the functions a, b, c are specified by

$$a(x) = 1, \quad b(x) = -200, \quad c(x) = \begin{cases} 0 & (0 \leq x \leq 0.5) \\ -40000 - 160000x & (0.5 < x \leq 1). \end{cases}$$

We choose the upwind parameters equal to

$$\varepsilon_m = \max\{0, 1 - 2a_m(|b_m|\Delta x)^{-1}\}.$$

Figure 4.2 shows, for $s = 59, 119, 239$, numerically computed norms $\|\varphi(hA)^n\|$ as functions of n, with $0 \leq n \leq 1000$. The corresponding values of h_0, defined in (4.4) (rounded to 5 decimal digits), and $h < h_0$ are specified in the caption below the figure. We see a nice stability behaviour of the numerical process, in that $\|\varphi(hA)^n\|$ is of moderate size, when the stepsize restriction (4.4) is fulfilled.

We conclude this section by comparing shortly the stepsize restriction (4.4) to the stepsize restriction at which one would arrive by using the eigenvalue

condition mentioned in Section 1.3. Let h_1 denote the largest value with the property that $\sigma[hA] \subset \text{int}(S)$ for all h with

$$0 < h < h_1. \tag{4.5}$$

Figure 4.3 shows, for $s = 59, 119, 239$, numerically computed norms $\|\varphi(hA)^n\|$ as functions of n, with $0 \leq n \leq 1000$. The corresponding threshold values h_1 (rounded to 5 decimal digits), and h with $h_0 < h < h_1$ are specified in the caption below the figure. For increasing values of s a very poor stability behaviour manifests itself.

From a comparison of the Figures 4.2 and 4.3 it is clear that the stepsize restriction along the lines of Section 4.1 is much more reliable than restriction (4.5).

Acknowledgement.

We would like to thank the referee for his comments.

REFERENCES

1. J. C. Butcher, *The Numerical Analysis of Ordinary Differential Equations*, John Wiley, Chicester, 1987.
2. M. Crouzeix, S. Larsson, S. Piskarev, and V. Thomée, *The stability of rational approximations of analytic semigroups*, BIT, 33 (1993), pp. 74–84.
3. J. L. M. van Dorsselaer, J. F. B. M. Kraaijevanger, and M. N. Spijker, *Linear stability analysis in the numerical solution of initial value problems*, Acta Numerica, 1993, pp. 199–237.
4. M. B. Giles, *Stability and convergence of discretizations of initial value p.d.e.'s*, Report 96/06, Oxford University Comp. Lab., 1996.
5. D. F. Griffiths, I. Christie, and A. R. Mitchell, *Analysis of error growth for explicit difference schemes in conduction-convection problems*, Internat. J. Numer. Meth. Engrg, 15 (1980), pp. 1075–1081.
6. Ch. Grossmann and H.-G. Roos, *Numerik partieller Differentialgleichungen*, Teubner, Stuttgart, 1992.
7. E. Hairer and G. Wanner, *Solving Ordinary Differential Equations*, Vol. II, Springer, Berlin, 1991.
8. P. J. van der Houwen, *Construction of Integration Formulas for Initial Value Problems*, North-Holland, Amsterdam, New York, Oxford, 1977.
9. H. W. J. Lenferink and M. N. Spijker, *A generalization of the numerical range of a matrix*, Linear Algebra Appl., 140 (1990), pp. 251–266.
10. H. W. J. Lenferink and M. N. Spijker, *On a generalization of the resolvent condition in the Kreiss matrix theorem*, Math. Comp., 57 (1991), pp. 211–220.
11. H. W. J. Lenferink and M. N. Spijker, *On the use of stability regions in the numerical analysis of initial value problems*, Math. Comp., 57 (1991), pp. 221–237.
12. C. Lubich and O. Nevanlinna, *On resolvent conditions and stability estimates*, BIT 31 (1991), pp. 293–313.
13. K. W. Morton, *Stability of finite-difference approximations to a diffusion-convection equation*, Int. J. Num. Meth. Engrg., 15 (1980), pp. 677–683.

14. J. Noye, *Computational Techniques for Differential Equations*, North-Holland, Amsterdam, 1984.
15. C. Palencia, *Stability of rational multistep approximations of holomorphic semigroups*, Math. Comp., 64 (1995), pp. 591–599.
16. S. V. Parter, *Stability, convergence, and pseudo-stability of finite-difference equations for an over-determined problem*, Numer. Math., 4 (1962), pp. 277–292.
17. S. C. Reddy and L. N. Trefethen, *Stability of the method of lines*, Numer. Math., 62 (1992), pp. 235–267.
18. R. D. Richtmyer and K. W. Morton, *Difference Methods for Initial Value Problems*, 2nd ed., John Wiley and Sons, New York, 1967.
19. M. N. Spijker, *Stepsize restrictions for stability of one-step methods in the numerical solution of initial value problems*, Math. Comp., 45 (1985), pp. 377–392.
20. M. N. Spijker, *Numerical ranges and stability estimates*, Appl. Numer. Math. 13 (1993), pp. 241–249.
21. M. N. Spijker and F. A. J. Straetemans, *Stability estimates for families of matrices of nonuniformly bounded order*, Linear Algebra Appl., 239 (1996), pp. 77–102.
22. M. N. Spijker and F. A. J. Straetemans, *A note on the order of contact between sets in the complex plane*, Report TW-96-06, Mathem. Instit., Leiden University, 1996.
23. G. Stoyan, *On Monotone Difference Schemes for Weakly Coupled Systems of Partial Differential Equations*, Computational Mathematics, Banach Center Publications, Vol. 13, PWN-Polish Scientific Publishers, Warsaw, 1984.