COMPUTER METHODS IN APPLIED MECHANICS AND ENGINEERING 80 (1990) 147–164
NORTH-HOLLAND

LAX-STABILITY OF FULLY DISCRETE SPECTRAL METHODS VIA STABILITY REGIONS AND PSEUDO-EIGENVALUES*

Satish C. REDDY and Lloyd N. TREFETHEN
Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.

Received 26 June 1989
Revised manuscript received December 1989

In many calculations, spectral discretization in space is coupled with a standard ordinary differential equation formula in time. To analyze the stability of such a combination, one would like simply to test whether the eigenvalues of the spatial discretization operator (appropriately scaled by the time step k) lie in the stability region for the o.d.e. formula, but it is well known that this kind of analysis is in general invalid. In the present paper we rehabilitate the use of stability regions by proving that a discrete linear multistep 'method of lines' approximation to a partial differential equation is Lax-stable, within a small algebraic factor, if and only if all of the ε-pseudo-eigenvalues of the spatial discretization operator lie within $O(\varepsilon)$ of the stability region as $\varepsilon \to 0$. An ε-pseudo-eigenvalue of a matrix A is any number that is an eigenvalue of some matrix $A+E$ with $\|E\| \leq \varepsilon$; our arguments make use of resolvents and are closely related to the Kreiss matrix theorem. As an application of our general result, we show that an explicit N-point Chebyshev collocation approximation of $u_t = -xu_x$ on $[-1, 1]$ is Lax-stable if and only if the time step satisfies $k = O(N^{-2})$, although eigenvalue analysis would suggest a much weaker restriction of the form $k \leq CN^{-1}$.

1. Introduction

Most spectral calculations for time-dependent partial differential equations consist of spectral discretization of the space derivatives coupled with a standard discrete o.d.e. formula for the time-stepping, such as an Adams–Bashforth, backwards differentiation, or Runge–Kutta formula [1, 2]. This decoupling of space and time is known as the *method of lines*, and among other advantages it has the virtue of simplicity. Though the order of accuracy with respect to the time step k may be low, it is usually practical to obtain acceptable results by taking k fairly small, since the computational work grows only linearly with k^{-1} and the storage requirements do not grow at all.

Numerical stability, however, presents serious difficulties in spectral calculations of this kind—more serious than with finite difference methods, for at least three reasons. First, stability has proven more difficult to analyze for spectral than for finite difference methods,

* Work supported by a U.S. National Science Foundation Presidential Young Investigator Award to L.N. Trefethen and by U.S. Air Force Grant AFOSR-87-0102. AMS subject classification: 65M10, 15A18.

since the operators involved are not translation-invariant and are often far from normal.[1] Second, the time step limits for stability are far more restrictive: typically $k \leq C/N^2$ and $k \leq C/N^4$ for explicit discretizations of first- and second-order problems on a grid of N points per space dimension, respectively, which is much stricter than would generally be required by accuracy considerations [2]. Finally, if one chooses to circumvent the problem of stability by employing an A-stable implicit time-stepping formula like the trapezoid or backward Euler formulas, the matrices to be inverted are dense and ill-conditioned. In certain cases preconditioned matrix iterations are effective for these problems, sometimes in combination with multigridding, but the resulting algorithms can hardly be called simple [3, 4].

This paper introduces a new method for dealing with the first of these difficulties, the analysis of stability for spectral methods. In any discretization by the method of lines, one has two elements to couple together: a family of grids and associated spatial discretization operators $\{L_k\}$, conveniently indexed by the time step k, and a stability region for the time-stepping formula. The standard heuristic for integrations over an infinite time interval is that the discretization will be stable if and only if all the eigenvalues of kL_k lie inside the stability region, for all sufficiently small k. However, unless the operators $\{L_k\}$ are normal, it is well known that this condition is in fact necessary but not sufficient for Lax-stability. In the case of spectral methods, L_k may grow rapidly further from normal as $k \to 0$ in the sense that the condition number of the matrix of eigenvectors grows exponentially, and the heuristic can be spectacularly wrong. For example, if $u_t = u_x$ on $[-1, 1]$, with boundary condition $u(1, t) = 0$, is discretized by an explicit spectral collocation method in Gauss–Legendre points, then eigenvalue analysis suggests a stability restriction $k \leq CN^{-1}$ [5, 6], whereas numerical experiments indicate that the actual Lax-stability restriction is $k = O(N^{-2})$ [7, 8].

This paper will show that the use of stability regions becomes valid if one modifies the standard heuristic by replacing eigenvalues by ε-pseudo-eigenvalues. This notion is defined in Section 2: an ε-pseudo-eigenvalue of a matrix A is any number $z \in \mathbb{C}$ which is an eigenvalue of $A + E$ for some perturbation matrix E with $\|E\| \leq \varepsilon$. Section 3 presents our main theorem: for Lax-stability, it is necessary and sufficient that all of the *ε-pseudo-eigenvalues* of kL_k lie within a distance $O(\varepsilon)$ of the stability region as $\varepsilon \to 0$, for all sufficiently small k.[2] Section 4 presents three examples of spectral differentiation operators whose pseudo-eigenvalues differ markedly from their eigenvalues, so that one needs this theorem to reach realistic conclusions about stability. Section 5 focuses on one particular example, analogous to the one above, and proves that for this example the condition $k = O(N^{-2})$ is indeed both necessary and sufficient for Lax-stability.

2. Definition of pseudo-eigenvalues

Let $\mathcal{B}$ be a Banach space with norm $\|\cdot\|$, and let $A: \mathcal{B} \to \mathcal{B}$ be a bounded linear operator; in many applications $\mathcal{B}$ is a space of vectors of finite length and A is a matrix. The definition of pseudo-eigenvalues for matrices is as follows. For operators, the definition is condition (iii), although the other definitions can also be appropriately modified.

[1] A normal matrix or operator is one which has a complete set of orthogonal eigenvectors, and consequently, eigenvalues that are well-conditioned with respect to perturbations. Symmetric and skew-symmetric matrices are the most familiar examples.

[2] The precise result gives Lax-stability up to a mild algebraic factor (Theorem 2), and assumes that the boundary of the stability region contains no cusps (Assumption A).

DEFINITION. Given $\varepsilon > 0$, the number $\lambda \in \mathbb{C}$ is an ε-pseudo-eigenvalue of A if any of the following equivalent conditions is satisfied:

(i) λ is an eigenvalue of $A + E$ for some matrix E with $\|E\| \leq \varepsilon$;

(ii) $\exists u \in \mathcal{B}$ with $\|u\| = 1$ such that $\|(\lambda I - A)u\| \leq \varepsilon$;

(iii) $\|(\lambda I - A)^{-1}\| \geq \varepsilon^{-1}$.

The set of all ε-pseudo-eigenvalues of A is denoted by $\Lambda_\varepsilon(A)$ or simply Λ_ε.

Thus a pseudo-eigenvalue of a matrix or operator A is an eigenvalue of some nearby operator $A + E$. The vector u in (ii) is called a (normalized) *ε-pseudo-eigenvector* of A. The matrix or operator $(\lambda I - A)^{-1}$ in (iii) is the *resolvent* of A at the point λ; all of our results are closely tied to the theory of resolvents.

If A is a normal matrix or operator, the ε-pseudo-spectrum Λ_ε consists of the union of the ε-balls about all of the points of the spectrum of A—which is not very interesting. If A is far from normal, however, Λ_ε may be much larger than this and may have a very different shape. One interesting and relatively easy to understand set of examples is the family of nonsymmetric Toeplitz matrices [9]. The matrices that arise in spectral methods appear to be less easy to understand, but they are even more interesting.

In the remainder of this paper, for simplicity, we shall assume that $\mathcal{B}$ is a Hilbert space of finite or infinite dimension whose norm $\|\cdot\|$ is the usual 2-norm $\|\cdot\|_2$. If $\mathcal{B}$ is of finite dimension N, then A is an $N \times N$ matrix, and an equivalent condition for λ to be an ε-pseudo-eigenvalue of A is

(iv) $\sigma_N(\lambda I - A) \leq \varepsilon$,

where σ_N denotes the smallest singular value.

The idea of pseudo-eigenvalues was introduced in [10][3] in connection with applications to matrix iterations. Even in that paper, however, the definition was motivated in part by the matrices that arise in spectral methods. We have found that pseudo-eigenvalues shed light on a wide variety of theoretical and practical problems in numerical analysis, and a survey of their uses is in preparation [12].

3. Main theorem

Before turning to Lax-stability for discretizations of partial differential equations, we begin with the simpler problem of power-boundedness for families of matrices. Let D denote the open unit disk in the complex plane, $\bar{D}$ its closure, and ∂D the unit circle. The following result is essentially one part of the Kreiss matrix theorem, sharpened as described in [13].

THEOREM 1. Let $\{A_\nu\}$ be a family of matrices or bounded linear operators of dimensions $N_\nu \leq \infty$. If the powers of these matrices satisfy

$$\|A_\nu^n\| \leq C \quad \forall n \geq 0 \tag{3.1}$$

for some constant C, independent of ν, then their ε-pseudo-eigenvalues $\{\lambda_\varepsilon\}$ satisfy

$$\operatorname{dist}(\lambda_\varepsilon, D) \leq C\varepsilon \quad \forall \varepsilon \geq 0\,. \tag{3.2}$$

[3] Under the name 'approximate eigenvalues'. We have subsequently learned that the same definition appeared earlier in [27].

Conversely, (3.2) implies

$$\|A_\nu^n\| \leq 2eC \min\{N_\nu, n\} \quad \forall n > 0 . \tag{3.3}$$

PROOF. By condition (iii) of the last section, (3.2) is equivalent to the estimate

$$\|(\lambda I - A_\nu)^{-1}\| \leq \frac{C}{\operatorname{dist}(\lambda, D)} \quad \forall \lambda \in \mathbb{C}\backslash\bar{D} ; \tag{3.4}$$

of course, for $\lambda \in \mathbb{C}\backslash\bar{D}$, $\operatorname{dist}(\lambda, D)$ is another way of writing $|\lambda| - 1$. If (3.1) holds, then (3.4) follows readily from a Taylor expansion of the resolvent. Conversely, if (3.4) holds, then (3.3) can be derived by means of the resolvent integral for A_ν^n. The bound $2eCn$ comes from estimating the integral directly, and the bound $2eCN_\nu$ follows with a bit more work after an integration by parts. The details are given in [13], which is a sharpening of an earlier paper by Tadmor [14]. □

If $\{A_\nu\}$ is a family of matrices of fixed dimension N, then $2eCN$ is a constant and Theorem 1 becomes a criterion for power-boundedness—a restatement of one of the assertions of the Kreiss matrix theorem. Even for a single matrix A, this criterion is quite different in style from the usual criterion for power-boundedness based on the spectral radius $\rho(A)$, which requires $\rho(A) \leq 1$ together with the condition that any eigenvalues on the unit circle be nondefective. Of course the two criteria are mathematically equivalent, but here, the nondefectiveness condition is implicit in (3.2) rather than explicit.

Now consider an explicit s-step linear multistep approximation to an autonomous ordinary differential equation $u_t = f(u)$,

$$\sum_{j=0}^{s} \alpha_j v^{n+j} - k \sum_{j=0}^{s-1} \beta_j f^{n+j} = 0 , \tag{3.5}$$

normalized by $\alpha_s = 1$ and $|\alpha_0| + |\beta_0| \neq 0$, with the usual notation $v^n \approx u(nk)$ and $f^n = f(v^n)$, where k is the time step. For each $w \in \mathbb{C}$, let $\pi_w(z)$ denote the associated *stability polynomial*

$$\pi_w(z) = \rho(z) - w\sigma(z) , \tag{3.6}$$

of degree exactly s, where $\rho(z)$ and $\sigma(z)$ are defined by

$$\rho(z) = \sum_{j=0}^{s} \alpha_j z^j , \qquad \sigma(z) = \sum_{j=0}^{s-1} \beta_j z^j . \tag{3.7}$$

Let $S \subseteq \mathbb{C}$ denote the *stability region* in the w-plane for the linear multistep formula (3.5): the set of all $w \in \mathbb{C}$ for which all roots z of $\pi_w(z)$ satisfy $|z| \leq 1$, with simple roots only for $|z| = 1$. Equivalently, (3.5) can be rewritten as a one-step formula involving vectors of length s,

$$\boldsymbol{v}^{n+1} = A\boldsymbol{v}^n + B\boldsymbol{f}^n , \tag{3.8}$$

that is,

$$\begin{bmatrix} v^{n+s} \\ \cdot \\ \cdot \\ \cdot \\ v^{n+1} \end{bmatrix} = \begin{bmatrix} -\alpha_{s-1} & \cdots & -\alpha_1 & -\alpha_0 \\ 1 & & & \\ & \ddots & & \\ & & 1 & \end{bmatrix} \begin{bmatrix} v^{n+s-1} \\ \cdot \\ \cdot \\ \cdot \\ v^n \end{bmatrix} + \begin{bmatrix} \beta_{s-1} & \cdots & \beta_1 & \beta_0 \\ 0 & & & \\ & \ddots & & \\ & & 0 & \end{bmatrix} \begin{bmatrix} f^{n+s-1} \\ \cdot \\ \cdot \\ \cdot \\ f^n \end{bmatrix}. \tag{3.9}$$

Then S is the set of all $w \in \mathbb{C}$ for which the $s \times s$ companion matrix

$$G(w) = A + wB \tag{3.10}$$

is power-bounded.[4] The idea behind the definition of S is that if (3.5) is applied to the linear model problem $u_t = au$, where $a \in \mathbb{C}$ is a constant, then the resulting sequence $\{\boldsymbol{v}^n\}$ satisfies

$$\boldsymbol{v}^n = [G(ak)]^n \boldsymbol{v}^0, \quad n \geq 0, \tag{3.11}$$

which is bounded for all initial vectors $\boldsymbol{v}^0$ if and only if $ak \in S$.

To determine the stability region S, one may consider the rational function

$$r(z) = \rho(z)/\sigma(z), \tag{3.12}$$

which maps the z-plane in an s-to-1 fashion onto the w-plane. The image under $r(z)$ of the unit circle ∂D in the z-plane is a curve which divides the w-plane into a number of components, in each of which the number of roots of $\pi_w(z)$ that lie in D is a constant, and the interior of S is the union of those components for which this number is s. For simplicity, we make the following assumptions concerning the linear multistep formula, which are satisfied by most explicit linear multistep formulas but not, for example, by the midpoint rule:

Assumption A. $r(z) \neq \infty$ and $r'(z) \neq 0$ for $z \in \partial D$.

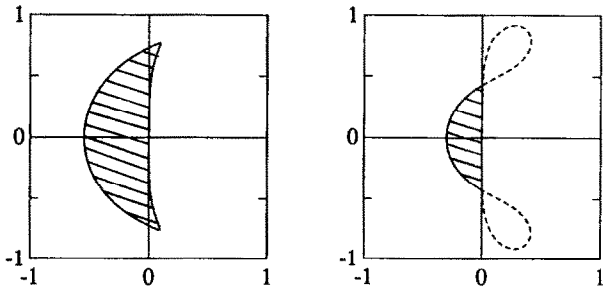
Since $r(\infty) = \infty$, the component containing the point $w = \infty$ is not in S, and thus Assumption A implies that S is bounded. Figure 1 shows two examples of stability regions satisfying this assumption.

Now, following the standard formulation of the Lax–Richtmyer stability theory [16], suppose we are given an autonomous linear partial differential equation

$$u_t = \mathcal{L}u, \tag{3.13}$$

where $u(t)$ is a function of one or more space variables on a bounded or unbounded domain and $\mathcal{L}$ is a differential operator, independent of t. For each sufficiently small time step $k > 0$, let a corresponding finite or infinite spatial grid be defined and let (3.13) first be discretized

[4] The power-boundedness of such matrices is considered in [15], where the analysis also leads to a condition like Assumption A.



(a) 3rd order (b) 4th order

Fig. 1. Stability regions for Adams–Bashforth formulas.

with respect to the space variables only, so that it becomes a system of ordinary differential equations,

$$v_t = L_k v\,, \tag{3.14}$$

where $v(t)$ is a vector of dimension $N_k \leq \infty$ and L_k is a matrix or bounded linear operator. With the space discretization determined in this way, let (3.14) then be discretized with respect to t by the linear multistep formula (3.5) with time step k. The resulting fully discrete approximation to (3.13) can be written

$$\boldsymbol{v}^{n+1} = A_k \boldsymbol{v}^n\,, \tag{3.15}$$

where A_k is the block companion matrix

$$A_k = G(kL_k) = \begin{bmatrix} \beta_{s-1}kL_k - \alpha_{s-1}\mathrm{I} & \cdots & \beta_1 kL_k - \alpha_1 \mathrm{I} & \beta_0 kL_k - \alpha_0 \mathrm{I} \\ \mathrm{I} & & & \\ & \ddots & & \\ & & \mathrm{I} & \end{bmatrix}, \tag{3.16}$$

with I denoting the identity operator on vectors of length N_k.

Here is the main theorem of this paper.

THEOREM 2. If the method of lines discretization described above satisfies

$$\|A_k^n\| \leq C_1 \quad \forall n \geq 0\,, \tag{3.17}$$

then the ε-pseudo-eigenvalues $\{\mu_\varepsilon\}$ of the operators $\{kL_k\}$ satisfy

$$\operatorname{dist}(\mu_\varepsilon, S) \leq C_2 \varepsilon \quad \forall \varepsilon \geq 0\,. \tag{3.18}$$

Conversely, (3.18) implies

$$\|A_k^n\| \leq C_3 \min\{N_k, n\} \quad \forall n > 0 . \tag{3.19}$$

The constants C_j are independent of k, and the ratios C_2/C_1 and C_3/C_2 may be chosen to depend only on the linear multistep formula, which is assumed to satisfy Assumption A.

PROOF. By Theorem 1, these assertions are certainly true if (3.18) is replaced by the estimate

$$\operatorname{dist}(\lambda_\varepsilon, D) \leq C_2' \varepsilon \quad \forall \varepsilon \geq 0 \tag{3.20}$$

on the ε-pseudo-eigenvalues $\{\lambda_\varepsilon\}$ of A_k. Therefore we are done if we can show that (3.18) and (3.20) are equivalent. Restating these conditions in terms of resolvents, as in the proof of Theorem 1, we find that we are done if we can show the equivalence of

$$\|(\mu I - kL_k)^{-1}\| \leq \frac{C_2}{\operatorname{dist}(\mu, S)} \quad \forall \mu \in \mathbb{C} \backslash \bar{S} \tag{3.21}$$

and

$$\|(\lambda I - A_k)^{-1}\| \leq \frac{C_2'}{\operatorname{dist}(\lambda, D)} \quad \forall \lambda \in \mathbb{C} \backslash \bar{D} . \tag{3.22}$$

First we shall prove that (3.21) implies (3.22). We begin by assuming that λ is close to D. Let Ω_z be an annulus

$$\Omega_z = \{z \in \mathbb{C} \mid 0 < \operatorname{dist}(z, D) < \tau_z\} , \tag{3.23}$$

with $\tau_z > 0$ chosen small enough so that $r(z) \neq \infty$ for all $z \in \bar{\Omega}_z$ (possible by Assumption A). Let $\lambda \in \Omega_z$ be arbitrary, and define $\mu = r(\lambda)$ (Fig. 2). Since $A_k = G(kL_k)$ by (3.16), we can write $(\lambda I - A_k)^{-1}$ as a resolvent integral with respect to w,

$$(\lambda I - A_k)^{-1} = \frac{1}{2\pi i} \int_\Gamma (\lambda I - G(w))^{-1} \otimes (wI - kL_k)^{-1} \, dw , \tag{3.24}$$

where Γ is any simple closed contour in the w-plane contained in the region of analyticity of $(\lambda I - G(w))^{-1}$ and enclosing the spectrum of kL_k [17]. As indicated by the symbol $\otimes$, the integrand is a matrix or operator of dimension sN_k obtained as the tensor product of a matrix of dimension s and a matrix or operator of dimension N_k. Now, $(\lambda I - G(w))^{-1}$ is an $s \times s$ matrix-valued function of w that is analytic throughout the w-plane except for a pole at $\mu = r(\lambda)$, which must be simple because $\det(\lambda I - G(w))$ is an affine function of w by (3.9) and (3.10). Accordingly, as suggested in Fig. 2, we can evaluate the integral by choosing Γ to be the union of Γ_1, a small circle about μ traversed negatively, and Γ_2, a large circle enclosing both μ and the spectrum of kL_k. The contribution from Γ_1 is

$$-R(\mu) \otimes (\mu I - kL_k)^{-1} ,$$

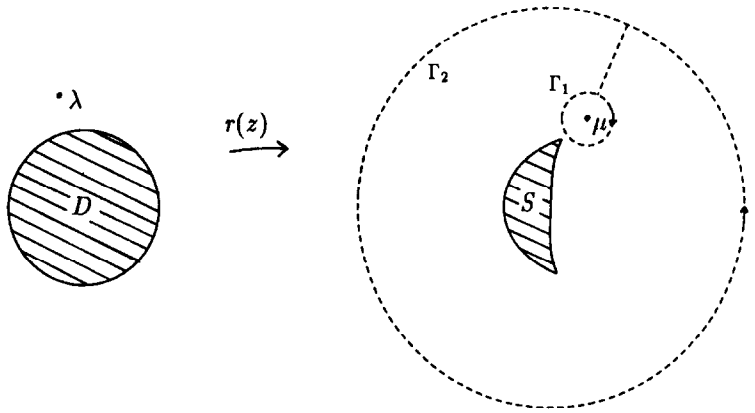


Fig. 2. Integration contours for the proof of Theorem 2.

where $R(\mu)$ denotes the residue of $(\lambda I - G(w))^{-1}$ at $w = \mu$, a nonzero matrix of dimension s. The contribution from Γ_2 is

$$(\lambda I - G(\infty))^{-1} \otimes I ,$$

where $(\lambda I - G(\infty))^{-1}$ is an abbreviation for $\lim_{w\to\infty}(\lambda I - G(w))^{-1}$; the term I appears as the residue of $(wI - kL_k)^{-1}$ at $w = \infty$. All together, we have

$$(\lambda I - A_k)^{-1} = (\lambda I - G(\infty))^{-1} \otimes I - R(\mu) \otimes (\mu I - kL_k)^{-1} , \tag{3.25}$$

and therefore

$$\|(\lambda I - A_k)^{-1}\| \leq C_4 + C_5 \|(\mu I - kL_k)^{-1}\| , \tag{3.26}$$

where C_4 and C_5 are defined by

$$C_4 = \sup_{z\in\bar{\Omega}_z} \|(zI - G(\infty))^{-1}\| , \qquad C_5 = \sup_{z\in\bar{\Omega}_z} \|R(r(z))\| ; \tag{3.27}$$

both constants are finite since the suprema involve continuous functions on the compact domain $\bar{\Omega}_z$. Combining (3.21) and (3.26) gives

$$\|(\lambda I - A_k)^{-1}\| \leq C_4 + \frac{C_2 C_5}{\operatorname{dist}(\mu, S)} . \tag{3.28}$$

Now from Assumption A it is readily verified that we have

$$\operatorname{dist}(\lambda, D) \leq C_6 \operatorname{dist}(\mu, S) \quad \forall \lambda \in \Omega_z , \tag{3.29}$$

for some constant C_6, independent of λ. Since we also have $\operatorname{dist}(\lambda, D) \leq \tau_z$ for $\lambda \in \Omega_z$, the combination of (3.28) and (3.29) implies

$$\|(\lambda I - A_k)^{-1}\| \leq \tilde{C}_2'/\mathrm{dist}(\lambda, D) \quad \forall \lambda \in \Omega_z$$

with $\tilde{C}_2' = C_4\tau_z + C_2C_5C_6$.

This shows that (3.21) implies the inequality of (3.22) for $\lambda \in \Omega_z$. The remaining values $\lambda \in \mathbb{C}\backslash\bar{D}$, namely those with $\mathrm{dist}(\lambda, D) \geq \tau_z$, can be handled as a corollary of this result by making use of the resolvent integral

$$(\lambda I - A_k)^{-1} = \frac{1}{2\pi i}\int_\Gamma (\lambda - z)^{-1}(zI - A_k)^{-1}\,dz \tag{3.30}$$

in the z-plane. Choosing the contour of integration to be $\Gamma = \{z \in \mathbb{C} \,|\, |z| = 1 + \frac{1}{2}\tau_z\}$ gives the bounds

$$|(\lambda - z)^{-1}| \leq 2/\mathrm{dist}(\lambda, D)$$

and

$$\|(zI - A_k)^{-1}\| \leq 2\tilde{C}_2'/\tau_z$$

for the two factors of this integrand, for all λ with $\mathrm{dist}(\lambda, D) \geq \tau_z$, and therefore (3.30) implies

$$\|(\lambda I - A_k)^{-1}\| \leq \frac{4(1 + \frac{1}{2}\tau_z)\tilde{C}_2'}{\tau_z\,\mathrm{dist}(\lambda, D)}\,, \tag{3.31}$$

since the arc length of Γ is $2\pi(1 + \frac{1}{2}\tau_z)$. In other words, (3.22) holds for all $\lambda \in \mathbb{C}\backslash\bar{D}$ if we take $C_2' = 4(1 + \frac{1}{2}\tau_z)\tilde{C}_2'/\tau_z$.

Now we shall prove the converse, that (3.22) implies (3.21). As before, we begin by assuming that μ is close to S. Let Ω_z again be the annulus defined in (3.23) and define

$$\Omega_w = \{w \in \mathbb{C} \,|\, 0 < \mathrm{dist}(w, S) < \tau_w\} \tag{3.32}$$

for some $\tau_w > 0$. From Assumption A it can be shown that one can pick τ_w small enough so that for each $\mu \in \Omega_w$, there is a number $\lambda \in \Omega_z$ satisfying $r(\lambda) = \mu$ and, in analogy to (3.29),

$$\mathrm{dist}(\mu, S) \leq C_6'\,\mathrm{dist}(\lambda, D)\,. \tag{3.33}$$

(There are s values of $r^{-1}(w)$ to choose from, among which we may take λ to be one that is largest in modulus.) This pair λ, μ will again satisfy (3.25), which implies

$$\|(\mu I - kL_k)^{-1}\| \leq C_5'(C_4 + \|(\lambda I - A_k)^{-1}\|)\,, \tag{3.34}$$

with

$$C_5' = [\inf_{z \in \Omega_z} \sup_{1 \leq i,j \leq s} |R(r(z))_{ij}|]^{-1}\,; \tag{3.35}$$

we have $C_5' < \infty$ since $R(r(z))$ is a continuous nonzero function on Ω_z. Combining (3.22), (3.33) and (3.34) implies

$$\|(\mu I - kL_k)^{-1}\| \leq \tilde{C}_2/\mathrm{dist}(\mu, S) \quad \forall \mu \in \Omega_w ,$$

with $\tilde{C}_2 = C_5' C_6' (C_4 \tau_z + C_2')$, since $\sup_{z \in \Omega_z} \mathrm{dist}(w(z), S) \leq C_6' \tau_z$.

Thus (3.22) implies the inequality of (3.21) for $\mu \in \Omega_w$, and to complete the proof, all that remains is to extend this conclusion to values $\mu \in \mathbb{C} \backslash \bar{S}$ with $\mathrm{dist}(\mu, S) \geq \tau_w$. This can be carried out much as before by means of the resolvent integral

$$(\mu I - kL_k)^{-1} = \frac{1}{2\pi i} \int_\Gamma (\mu - w)^{-1} (wI - kL_k)^{-1} \, dw , \tag{3.36}$$

with the contour of integration chosen as $\Gamma = \{w \in \mathbb{C} \,|\, \mathrm{dist}(w, S) = \frac{1}{2}\tau_w\}$. We omit the details. □

Many extensions of Theorem 2 are possible. For example, it is natural to consider

1. Stability for initial-value problems defined on a finite time interval $0 \leq t \leq T$ instead of the infinite interval $0 \leq t < \infty$;
2. Weighted norms $\| \cdot \|$ rather than just the 2-norm, to permit sharp stability estimates for nonuniform grids;
3. Weaker definitions of stability (e.g., algebraic stability in the sense of [1]), with correspondingly weaker conditions on the pseudo-eigenvalues;
4. Implicit as well as explicit time-stepping formulas;
5. Runge–Kutta and other classes of time-stepping formulas;
6. Time-stepping formulas that violate Assumption A.

Most of these generalizations will be treated in our upcoming paper [11]. As for the sharpness of Theorems 1 and 2, it is an open question whether the factors n and N in (3.3) and (3.19) can be improved, but for many practical purposes, including the applications of the next two sections, they are nearly as good as constants.

Theorem 2 is certainly not the first result in the literature concerning stability of method of lines discretizations. Close in spirit to our own work are [18, 19], based on the idea of the spectrum of a family of matrices [20, 21], which corresponds roughly to a limit of ε-pseudo-spectra as $N \to \infty$ and $\varepsilon \to 0$. Also closely related are the various results of Lenferink et al. [22, 23], some of which are valid for nonlinear as well as linear operators $\mathcal{L}$ in (3.13). As far as we know, however, none of these results include necessary and sufficient stability criteria for multistep method of lines calculations.

4. Examples of pseudo-spectra

In this section we present three examples of pseudo-spectra of spectral differentiation matrices. The interesting cases are those in which the pseudo-eigenvalues differ markedly from the exact ones, and this is the typical situation for spectral differentiation of odd order on domains with boundaries.

In particular, consider two model first-order initial boundary value problems defined for $x \in [-1, 1]$, $t \geq 0$:

$$u_t = u_x\,, \qquad u(x,0) = f(x)\,, \quad u(1,t) = 0 \tag{4.1}$$

and

$$u_t = -xu_x\,, \qquad u(x,0) = f(x)\,, \quad \text{no boundary conditions}\,. \tag{4.2}$$

Both these problems are well-posed in various norms. The solution to (4.1) is $u(x,t) = f(x+t)$ if $x+t \leq 1$, $u(x,t) = 0$ otherwise, and the solution to (4.2) is $u(x,t) = f(e^{-t}x)$. Let D_N be the $N \times N$ differentiation matrix obtained by a spectral collocation discretization of the operator ∂_x or $-x\partial_x$ in a grid of distinct points $x_1, \ldots, x_N$. Then both (4.1) and (4.2) take the semidiscrete form

$$v_t = D_N v\,, \tag{4.3}$$

a special case of (3.14). In the case of problem (4.1) the grid points lie in $[-1,1)$, and D_N is defined by first interpolating the data $\{v_j\}$ by a polynomial $p(x)$ of degree N that satisfies the boundary condition $p(1) = 0$, then setting $(D_N v)_j = p'(x_j)$. (Thus the jth column of D_N consists of samples of the derivative of the Lagrange interpolating polynomial to the discrete delta function located at x_j.) In the case of problem (4.2) the points lie in $[-1,1]$, $p(x)$ becomes a polynomial of degree $N-1$ since there are no boundary conditions, and D_N is defined by $(D_N v)_j = -x_j p'(x_j)$.

Many grids $\{x_j\}$ may be chosen for these differentiation processes, of which we shall consider two. Our Legendre grid is defined by taking x_j to be the jth zero of the Legendre polynomial $P_N(x)$. Our Chebyshev grid is defined by taking x_j to be the jth extreme point of the Chebyshev polynomial $T_{N-1}(x)$ for problem (4.2), or the jth extreme point of $T_N(x)$ (other than $x = 1$) for problem (4.1). (That is, $x_j = \cos((j-1)\pi/(N-1))$ for (4.2), $\cos(j\pi/N)$ for (4.1).) Other grids could equally well be considered. Any choice with density proportional to $(1-x^2)^{-1/2}$ in the limit $N \to \infty$ is a reasonable candidate, but other limits such as a uniform density lead to differentiation operators with eigenvalues in the right half-plane, making them exponentially ill-posed even in semidiscrete form, before the time-discretization is introduced.

For our first example, Fig. 3 illustrates the spectral behavior of D_N/N^2 for problem (4.1) on a Legendre grid in the limit $N \to \infty$ (essentially the same pictures would be obtained with $N = 64$ or 128). The important thing to notice is that we have divided D_N by N^2, and even so, interesting behavior remains. The solid dot at the origin reflects the fact that in the limit $N \to \infty$, the only eigenvalue of D_N/N^2 is 0; the eigenvalues of D_N itself have magnitude $O(N)$ [5, 6]. The pseudo-eigenvalues of D_N/N^2 are quite another matter, however, as is revealed by the curves representing the boundaries of the sets Λ_ε for $\varepsilon = 10^{-5}$, 10^{-10} and 10^{-15}.[5] Evidently for large N and small ε, the ε-pseudo-spectrum of D_N/N^2 includes a large lobe to the left of the origin. This lobe is approximately circular in shape, with a radius that decreases approximately in proportion to $|\log \varepsilon|^{-1}$ but independently of N. These pseudo-eigenvalues will have practical consequences just like exact eigenvalues, and indeed, in floating-point

[5] To determine curves like these numerically, having picked a reasonably large value of N, one can evaluate the smallest singular value $\sigma_N(w\mathrm{I} - D_N)$ at a grid of points in the w-plane and then draw a contour plot of the results, but this process requires a great deal of computing. Instead, we have perturbed D_N by a random matrix of norm ε and computed the corresponding eigenvalues, which tend to delineate the boundary of Λ_ε inexactly but cheaply. (For examples of such plots see [7–9].) Figures 3–5 are approximate sketches based on a few such computations with $N = 32$, 64 and 128.

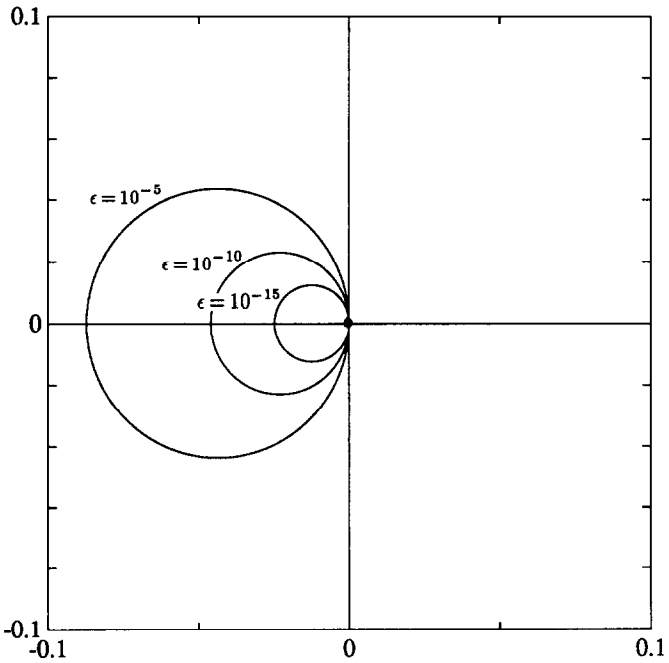


Fig. 3. Eigenvalue (the solid dot at the origin) and ε-pseudo-eigenvalue domains (bounded by curves) of D_N/N^2 for problem (4.1) on a Legendre grid in the limit $N \to \infty$. The curves are approximate.

arithmetic with machine precision 10^{-10}, for example, D_N/N^2 would behave numerically as if the middle circle in Fig. 3 were entirely filled with solid dots.

For the matrix D_N instead of D_N/N^2, we see from Fig. 3 that the pseudo-eigenvalues fill much of the left half-plane, scaling in proportion to N^2 even though the eigenvalues scale in proportion to N.

Theorem 2 enables us to draw conclusions concerning stability from these pseudo-eigenvalue domains. Assuming that Fig. 3 is correct, the theorem indicates that a Legendre method-of-lines discretization of (4.1) will be Lax-stable if and only if the stability region contains *any* disk in the left half-plane whose boundary passes through the origin. For most explicit formulas, the resulting time step constraint will be $k = O(N^{-2})$, where the constant implicit in the 'big O' is arbitrary, a conclusion supported abundantly by numerical experiments in [8]. Thus for this problem, the time step restriction for Lax-stability is determined entirely by pseudo-eigenvalues rather than by eigenvalues.

For our second example, Fig. 4 illustrates the spectral behavior of problem (4.1) on a Chebyshev instead of a Legendre grid. Again, the figure shows the eigenvalues and pseudo-eigenvalues of D_N/N^2 in the limit $N \to \infty$. The pseudo-eigenvalues look much as before: they are of magnitude O(1), indicating that the pseudo-eigenvalues of D_N itself scale again as $O(N^2)$. Now, however, there are nonzero eigenvalues too, for some of the eigenvalues of D_N scale as N^2. By Theorem 2, a Chebyshev method of lines discretization of (4.1) will be stable if and only if the stability region contains these eigenvalues and in addition some disk in the left half-plane whose boundary passes through the origin. For most explicit linear multistep formulas, the former condition will be more restrictive than the latter, and we will end up with a stability restriction $k \leq CN^{-2}$ for some particular constant C.

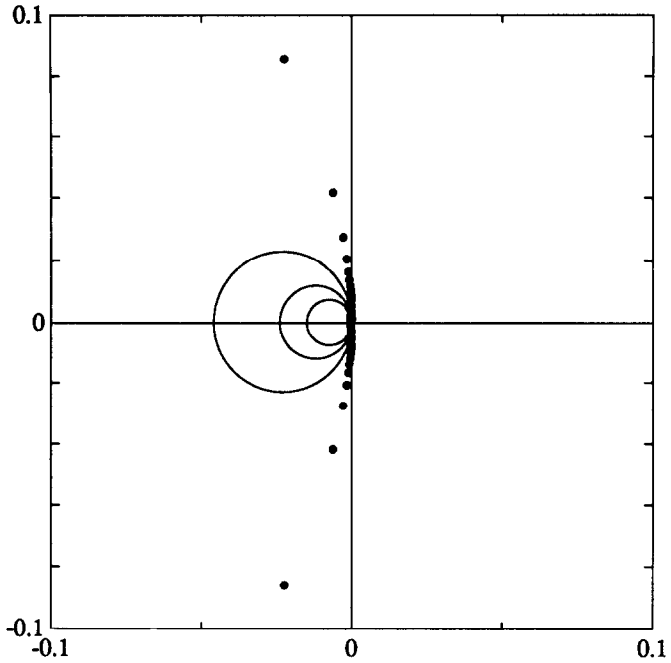


Fig. 4. Same as Fig. 3, but for problem (4.1) on a Chebyshev grid. Now some nonzero eigenvalues as well as pseudo-eigenvalues are present (solid dots).

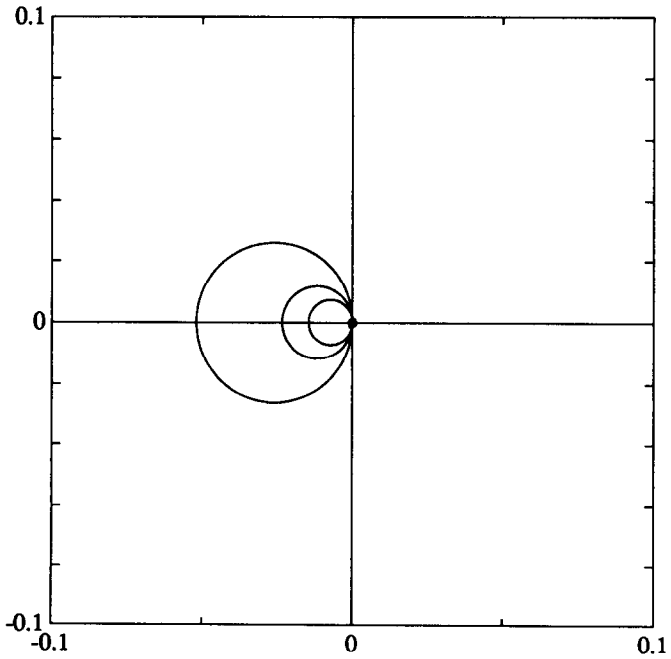


Fig. 5. Same as Figs. 3 and 4, but for problem (4.2) on a Legendre or Chebyshev grid. As in Fig. 3, the only eigenvalue is 0.

Finally, Fig. 5 is analogous to Figs. 3 and 4 but represents problem (4.2) on either a Chebyshev or a Legendre grid (the results for the two grids differ little). Except for a scale factor of about 2, the picture is much as in Fig. 3: the eigenvalues of D_N/N^2 all approach 0 in the limit $N \to \infty$ (easy to prove; the eigenvalues of D_N are in fact $0, -1, \ldots, 1-N$), but the pseudo-eigenvalues occupy approximate disks in the left half-plane. Again the Lax-stability restriction will be $k = O(N^{-2})$, with an arbitrary constant implicit in the O. Experiments with this spectral method are reported in [7, 24].

Many theoretical and empirical results have been derived previously concerning the eigenvalues of the differentiation matrices just considered, of which the most precise are those of Dubiner [5, 6], who proved that the eigenvalues of D_N have magnitude $O(N)$ for problem (4.1) with a Legendre grid, and related their distribution to the zeros of a Hankel function, and Vandeven [25], who has shown more recently that the eigenvalues of D_N are essentially zeros of Padé approximants. Since the eigenvalues of these matrices are so sensitive to perturbations, however, the consideration of them alone may be misleading; one must take the pseudo-eigenvalues into account too.

5. An application

The pseudo-eigenvalue domains presented in the last section were computed numerically. In combination with Theorem 2, these results immediately suggest time step limits for Lax-stability, but before these limits can be stated as theorems one must prove that the pseudo-eigenvalues behave as they appear to.

At present we have just a few results in this direction, and therefore in this final section, we shall establish only a single stability estimate for a special case, leaving sharper and more general estimates to a later paper. As far as we know, the theoretical determination of pseudo-eigenvalues is not a routine matter like Von Neumann analysis for finite difference formulas. This is unfortunate, but it does not imply that analysis of pseudo-eigenvalues is the wrong approach to stability. After all, Theorem 2 shows that one *has* to consider pseudo-eigenvalues, explicitly or implicitly, since eigenvalues alone may give the wrong answer. A more positive observation is that unlike exact eigenvalues, pseudo-eigenvalues can reliably be determined numerically (because the effect of small perturbations is built into their definition), and as a result, practical stability analysis based upon them can be carried out with the aid of the computer. As for theoretical stability analysis, we suspect that the analytical estimation of pseudo-eigenvalues will in the long run prove no harder than the analytical estimation of exact eigenvalues, and possibly easier, since the estimates in question are less sensitive to perturbations. For some ideas in this direction we refer to the final section of [7].

The example we shall consider is problem (4.2) on the Chebyshev grid, whose pseudo-spectrum was depicted in Fig. 5. To make our proof go through we shall restrict our attention to a special class of linear multistep formulas, which includes for example the Adams–Bashforth formulas of orders 3 and 4 (Fig. 1) but not of orders 1, 2 or 5:

Assumption B. The stability region S contains the intersection of the left half-plane with some disk Δ centered at the origin.

Roughly speaking, the following theorem asserts that the spectral method in question is algebraically Lax-stable provided that the time step satisfies $k = O(N^{-2})$, where the constant in the 'big O' can in principle be arbitrarily large. Our stability estimate (5.2) is certainly not sharp, and Assumption B is possibly unnecessary.

Reversing the convention of Section 3 to conform to the more standard usage in the spectral literature, in what follows we speak of N as the independent parameter and of k as a function of N, and for consistency, the solution operator A_k gets the new name A_N.

THEOREM 3. Let problem (4.2) be modeled by spectral differentiation in Chebyshev extreme points, as described in the last section, coupled with a linear multistep formula satisfying Assumptions A and B. If $k = k(N)$ satisfies

$$k \leq C_1 N^{-2} \tag{5.1}$$

for some constant C_1, then we have the stability estimate

$$\|A_N^n\| \leq C_2 N^{3/2} \log N \quad \forall n \geq 0 \tag{5.2}$$

for some constant C_2, for all N larger than some constant N_0.

PROOF. The estimate (5.2) is the same as (3.19), except weakened by a factor of $N^{1/2} \log N$. Therefore, by a trivial modification of Theorem 2 to take this factor into account, we are done if we can show that (5.1) implies the correspondingly modified version of (3.18),

$$\operatorname{dist}(\mu_\varepsilon, S) \leq C_3 \varepsilon N^{1/2} \log N \quad \forall \varepsilon \geq 0\,, \tag{5.3}$$

where μ_ε ranges over the ε-pseudo-eigenvalues of the operators $\{kD_N\}$. Equivalently, it is enough to show that (5.1) implies

$$\|(\mu I - kD_N)^{-1}\| \leq C_3 N^{1/2} \log N / \operatorname{dist}(\mu, S) \quad \forall \mu \in \mathbb{C} \backslash \bar{S}\,, \tag{5.4}$$

in analogy to (3.21).

First, suppose that $\mu \in \mathbb{C} \backslash \bar{S}$ lies in Δ, hence in the right half-plane by Assumption B. The continuous problem (4.2) has solution operator $\exp(-x\partial_x t)$ and solution $u(x, t) = u(e^{-t}x, 0)$, from which we readily derive the identity

$$\|\exp(-x\partial_x t)\| = e^{t/2} \tag{5.5}$$

in the 2-norm on $[-1, 1]$. The semidiscrete problem (4.3) has solution operator $\exp(D_N t)$, and in a suitably weighted discrete 2-norm we could derive a well-posedness estimate closely analogous to (5.5). Since Theorem 2 was stated for simplicity in the unweighted discrete 2-norm, however, we shall make do instead with the estimate

$$\|\exp(D_N t)\| \leq C_3 N^{1/2} \log N \quad \forall t \geq 0 \tag{5.6}$$

for some C_3, or equivalently,

$$\|\exp(kD_N t)\| \leq C_3 N^{1/2} \log N \quad \forall t \geq 0 . \tag{5.7}$$

To establish (5.6), note that the solution operator $\exp(D_N t)$ can be interpreted as the composition of polynomial interpolation in the grid points $\{x_j\}$, dilation by the factor e^t, and resampling at the grid points. In the supremum norm $\|\cdot\|_\infty$ these three operations have norms bounded by Λ_N, 1 and 1, respectively, where Λ_N is the so-called Lebesgue constant for the set $\{x_j\}$. For Chebyshev extreme points it is known that the Lebesgue constants satisfy $\Lambda_N \leq C_3 \log N$ for some C_3 [26], and we lose the additional factor of $N^{1/2}$ in converting back from $\|\cdot\|_\infty$ to $\|\cdot\| = \|\cdot\|_2$.

By the exponential version of the Kreiss matrix theorem—the analogue of the relationship (3.1)$\Rightarrow$(3.4) for exponentials of matrices rather than powers—it now follows from (5.7) that the resolvent satisfies

$$\|(\mu I - kD_N)^{-1}\| \leq C_3 N^{1/2} \log N / \operatorname{Re} \mu , \tag{5.8}$$

and this in turn implies (5.4), since $\operatorname{dist}(\mu, S) \leq \operatorname{Re} \mu$ for $\mu \in \mathbb{C} \backslash \bar{S}$ by Assumption B. Note that this part of the argument holds regardless of the size of k; we did not make use of (5.1).

On the other hand, suppose $\mu \in \mathbb{C} \backslash \bar{S}$ lies outside of Δ. Explicit formulas for the entries of the matrix D_N are given in [2, p. 69] among other places, and the largest of these entries scale as N^2, so if (5.1) holds, the entries of kD_N are bounded independently of N. By elementary consideration of the formulas for these elements, it can be shown that if kD_N is viewed in a natural way as an infinite matrix with all entries equal to 0 except in an $N \times N$ submatrix, then

$$\|kD_N - kD_\infty\| \to 0 \quad \text{as } N \to \infty , \tag{5.9}$$

where kD_∞ is the infinite matrix of limiting entries, and moreover, the spectrum of kD_∞ is the single point $\{0\}$. Now choose $\varepsilon_0 > 0$ small enough so that the pseudo-eigenvalue domain $\Lambda_{2\varepsilon_0}(kD_\infty)$ is contained in Δ, and choose N_0 large enough so that $\|kD_N - kD_\infty\| \leq \varepsilon_0$ for $N \geq N_0$. Then for $N \geq N_0$ we have

$$\Lambda_{\varepsilon_0}(kD_N) \subseteq \Lambda_{\varepsilon_0 + \|kD_N - kD_\infty\|}(kD_\infty) \subseteq \Lambda_{2\varepsilon_0}(kD_\infty) \subseteq \Delta , \tag{5.10}$$

or to put it in terms of the resolvent,

$$\|(\mu I - kD_N)^{-1}\| < \varepsilon_0^{-1} \quad \forall N \geq N_0 , \tag{5.11}$$

since $\mu \notin \Delta$. This establishes (5.4) even without the factor $N^{1/2} \log N$. □

As for the converse of Theorem 3, it is shown in the Theorem of [8] that if $k \neq O(N^{-2})$ as $N \to \infty$, then the spectral discretization is at least algebraically unstable. Although we shall not prove it here, the instability is actually much more dramatic than this, as can be seen by considering the logarithmic spacing of the pseudo-eigenvalue domains in Fig. 5. If $k = N^{-\alpha}$

with $\alpha < 2$, for example, then for each small ε, the ε-pseudo-eigenvalue domains of kD_N grow in size as $N \to \infty$, quickly reaching outside any bounded stability region. Concomitantly, given any fixed stability region S, the values of ε for which the ε-pseudo-eigenvalues of kD_N lie within $O(\varepsilon)$ of S decrease exponentially with N, and by Theorem 2, the result is an exponential instability.

Note added in proof

Recently we have received several new papers dealing with stability of the method of lines via resolvents. Sufficient conditions for stability are obtained for one-step methods with bounded stability regions in [28, 29] and for A-stable methods in [30].

Acknowledgment

We are grateful for advice and references to A. Iserles, J.F.B.M. Kraaijevanger, O. Nevanlinna, J.M. Sanz-Serna, M.N. Spijker and J.G. Verwer.

References

[1] D. Gottlieb and S.A. Orszag, Numerical Analysis of Spectral Methods: Theory and Applications (SIAM, Philadelphia, PA, 1977).
[2] C. Canuto, M.Y. Hussaini, A. Quarteroni and T.A. Zang, Spectral Methods in Fluid Dynamics (Springer, New York, 1988).
[3] E.M. Rønquist and A.T. Patera, Spectral element multigrid I: Formulation and numerical results, J. Sci. Comput. 2 (1987) 389–406.
[4] Y. Maday and R. Muñoz, Spectral element multigrid II: Theoretical justification, J. Sci. Comput. 3 (1988) 323–353.
[5] M. Dubiner, Asymptotic analysis of spectral methods, J. Sci. Comput. 2 (1987) 3–31.
[6] H. Tal-Ezer, A pseudospectral Legendre method for hyperbolic equations with an improved stability condition, J. Comput. Phys. 67 (1986) 145–172.
[7] L.N. Trefethen and M.R. Trummer, An instability phenomenon in spectral methods, SIAM J. Numer. Anal. 24 (1987) 1008–1023.
[8] L.N. Trefethen, Lax-stability vs. eigenvalue stability of spectral methods, in: K.W. Morton and M.J. Baines, eds., Numerical Methods in Fluid Dynamics III (Clarendon Press, Oxford, 1988) 237–253.
[9] L. Reichel and L.N. Trefethen, Eigenvalues and pseudo-eigenvalues of Toeplitz matrices, Linear Algebra Appl., submitted.
[10] L.N. Trefethen, Approximation theory and numerical linear algebra, in: J.C. Mason and M.G. Cox, eds., Algorithms for Approximation II (Chapman & Hall, London, 1990) 336–360.
[11] S.C. Reddy and L.N. Trefethen, Lax-stability of the method of lines, in preparation.
[12] L.N. Trefethen, Non-normal matrices and pseudo-eigenvalues in numerical analysis, in preparation.
[13] R.J. LeVeque and L.N. Trefethen, On the resolvent condition in the Kreiss matrix theorem, BIT 24 (1984) 584–591.
[14] E. Tadmor, The equivalence of L_2-stability, the resolvent condition, and strict H-stability, Linear Algebra Appl. 41 (1981) 151–159.

[15] G. Dahlquist, H. Mingyou and R. LeVeque, On the uniform power-boundedness of a family of matrices and the applications to one-leg and linear multistep methods, Numer. Math. 42 (1983) 1–13.
[16] R.D. Richtmyer and K.W. Morton, Difference Methods for Initial-Value Problems, 2nd ed. (Wiley/Interscience, New York, 1967).
[17] T. Kato, Perturbation Theory for Linear Operators (Springer, New York, 1976).
[18] G. Di Lena and D. Trigiante, On the stability and convergence of lines method, Rend. Mat. 24 (1983) 113–126.
[19] V. Lakshimkantham and D. Trigiante, Theory of Difference Equations: Numerical Methods and Applications (Academic Press, New York, 1988).
[20] N.S. Bakhvalov, Numerical Methods (Mir, Moscow, 1977).
[21] S.K. Godunov and V.S. Ryabenki, Theory of Difference Schemes (North-Holland, Amsterdam, 1964).
[22] H.W.J. Lenferink and M.N. Spijker, The relevance of stability regions in the numerical solution of initial value problems, in: K. Strehmel, ed., Numerical Treatment of Differential Equations (Teubner, Leipzig, 1988).
[23] J.M. Sanz-Serna and J.G. Verwer, Stability and convergence at the PDE/stiff ODE interface, Appl. Numer. Math. 5 (1989) 117–132.
[24] A. Solomonoff and E. Turkel, Global properties of pseudospectral methods, J. Comput. Phys. 81 (1989) 239–276.
[25] H. Vandeven, to appear.
[26] H. Ellich and K. Zeller, Auswertung der Normen von Interpolationsoperatoren, Math. Ann. 164 (1966) 105–112.
[27] J.M. Varah, On the separation of two matrices, SIAM J. Numer. Anal. 16 (1979) 216–222.
[28] H.W.J. Lenferink and M.N. Spijker, On the use of stability regions in the numerical analysis of initial value problems, Report TW-89-07, Dept. Math. and Comput. Sci., University of Leiden, 1989.
[29] H.W.J. Lenferink and M.N. Spijker, On a generalization of the resolvent condition in the Kreiss matrix theorem, Report TW-89-08, Dept. Math. and Comput. Sci., University of Leiden, 1989.
[30] C. Lubich and O. Nevanlinna, On resolvent conditions and stability estimates, BIT, submitted.