To appear in *IMA Journal of Numerical Analysis* in 1997.

On the stability and convergence of discretisations of initial value p.d.e.'s

M. B. Giles
Oxford University Computing Laboratory
Numerical Analysis Group

This paper examines the stability and convergence of discretisations of initial value p.d.e.'s using spatial discretisation followed by time integration with an explicit one-step method. A Cauchy integral representation is used to bound the growth in the discrete solution. New results are obtained regarding sufficient conditions for both algebraic and strong stability. Sufficient conditions are also derived for convergence on a finite time interval.

Subject classifications: AMS(MOS): 65L05,65M10,65M20

Key words and phrases: Numerical analysis, stability, convergence, initial value p.d.e.

Oxford University Computing Laboratory
Numerical Analysis Group
Wolfson Building
Parks Road
Oxford, England OX1 3QD
E-mail: giles@comlab.oxford.ac.uk

April, 1997

1 Introduction

Fourier analysis is the standard method for analysing the stability of discretisations of an initial value p.d.e. on a regular structured grid. For each point in the computational grid, a linear model problem is constructed on an infinite grid with uniform grid spacing and coefficients matching those of the chosen point. This model problem has Fourier eigenmodes whose stability is relatively easily analysed. If they are stable at all points in the grid, and the discretisation of the boundary conditions is also stable (which can be analysed using Godunov-Ryabenkii [3] or GKS [4] stability theory) then for most applications the overall discretisation is stable, in the sense of Lax [9]. The Lax Equivalence theorem then applies, that if the discretisation is consistent (for a dense subset of sufficiently smooth initial conditions) then the discrete solution will approach the analytic solution for all initial conditions as the grid spacing and timestep are reduced to zero.

However, engineering applications of CFD are increasingly using finite volume and finite element methods based on unstructured grids. For these, Fourier stability analysis is not applicable, and one must instead consider the full discrete matrix that arises from the combined spatial and temporal discretisation of the p.d.e. and associated boundary conditions. This paper addresses theoretical aspects of this analysis. Reference [2] gives an example of its application to the stability of a Galerkin discretisation of the Navier-Stokes equations on an unstructured tetrahedral grid.

We consider a particular class of discretisations of the initial value p.d.e.

$$\frac{\partial u}{\partial t} = Q(u), \tag{1.1}$$

where Q is a time-invariant linear differential operator which in three dimensions would be of the form

$$Q(u) = \sum_{i,j,k} a_{i,j,k}(x,y,z) \frac{\partial^i}{\partial x^i} \frac{\partial^j}{\partial y^j} \frac{\partial^k}{\partial z^k} u(x,y,z). \tag{1.2}$$

The first stage in discretising this p.d.e. is to perform a spatial approximation to produce the semi-discrete system of coupled o.d.e.'s,

$$\frac{du_h}{dt} = Q_h u_h(t). \tag{1.3}$$

Here $u_h(t)$ is to approximate the value of $u(x,t)$ at a set of discrete points and Q_h is the time-invariant matrix which approximates $Q(u)$. h represents the spatial grid resolution and we will consider a family of such discretisations for a sequence of values of h tending to zero. Note that as $h \to 0$, the dimension of Q_h will increase without bound. It is this feature which makes it difficult to derive stability bounds for the whole family of discretisations.

The second stage in the discretisation is to approximate the semi-discrete equations using an explicit one-step Runge-Kutta method with timestep k, to give

$$u_{h,n+1} = \varphi(kQ_h)\, u_{h,n}, \tag{1.4}$$

where $u_{h,n}$ represents $u_h(t)$ at time $t = nk$, and $\varphi(z)$ is a polynomial function of degree p,

$$\varphi(z) = \sum_{j=0}^{p} a_j z^j, \quad a_0 = a_1 = 1, \quad a_p \neq 0. \tag{1.5}$$

In the family of fully-discrete discretisations, we assume an implicit relationship between h and k, such that $k \to 0$ as $h \to 0$. Given the association between h and k, it is convenient to change notation, replacing u_h, Q_h by u_k, Q_k.

Central to any stability analysis is the stability region associated with $\varphi(z)$, defined as

$$S = \{z : |\varphi(z)| \leq 1\}. \tag{1.6}$$

The aim of this paper is to investigate the conditions required for stability and convergence of the fully discrete approximation. The objective in the stability analysis is to construct an upper bound for the growth of the solution for arbitrary initial conditions. Following the terminology of Spijker *et al* [5, 6, 12], a discretisation of an initial value problem (not necessarily arising from the spatial discretisation of a p.d.e.) is defined to be *strongly stable* if there a positive constant γ such that

$$|u_n| \leq \gamma |u_0|, \quad \forall n > 0, \tag{1.7}$$

and it is defined to be *algebraically stable* if there are positive constants γ, q such that

$$|u_n| \leq \gamma n^q |u_0|, \quad \forall n > 0. \tag{1.8}$$

There has been considerable research on the conditions under which the discretisation of a system of o.d.e.'s, or a family of such systems, is stable in either of the above senses. Spijker *et al* have shown the important role of the numerical range $\tau(kQ_k)$ of the matrix kQ_k [5, 6, 12]. They prove that there are many equivalent characterisations of the numerical range for arbitrary norms, but the definition which is most useful in proving stability is the following based on the resolvent:

Definition 1.1 *The numerical range $\tau(A)$ of the square matrix A is the smallest closed convex set $V \subset \mathbb{C}$ such that*

$$\left\| (zI - A)^{-1} \right\| \leq d(z, V)^{-1}, \ \forall z \notin V$$

where

$$d(z, V) \equiv \inf_{\zeta \in V} |z - \zeta|$$

When using the L_2 norm, this can be proved to be equivalent to the classical numerical range defined as

$$\tau(A) = \{x^* A x : x^* x = 1\}.$$

Previous papers by Spijker *et al* [5, 6, 12], Reddy and Trefethen [8] and Lubich and Nevanlinna [7] have proved algebraic stability with $q=1$ when $\tau(kQ_k)\subset S, \forall k$, and with improved exponents $q<1$ under various additional conditions. In particular, strong stability ($q=0$) can be proved under more restrictive conditions. The stability results in Section 2 of the present paper add to this literature by proving new sufficient conditions for both algebraic stability (with $0<q<1$) and strong stability.

In convergence analysis for a finite time interval, $0\le t\le 1$, the question is whether the discrete solution $u_{k,n}$ approaches the analytic solution $u(x,t)$ uniformly as $k\to 0$. The main result of Section 3 is that sufficient conditions for convergence are that $\tau(kQ_k)\subset S, \forall k$ and the full discretisation has a truncation error which decays faster than $|\log k|^{-1}$ as $k\to 0$ and satisfies a Lipschitz condition. Under additional conditions it is shown that the logarithmic term can be omitted. Section 3 concludes with a discussion of the relationship of these results to the Lax Equivalence Theorem [9].

2 Stability

The stability estimates are all based on the use of the Cauchy integral formula,

$$f(A)=\frac{1}{2\pi \mathrm{i}}\int_\Gamma f(z)(zI-A)^{-1}\mathrm{d}z \tag{2.1}$$

where $f(z)$ is an analytic function and the contour Γ encloses the spectrum of the square matrix A. In the context of the p.d.e. discretisation discussed in the Introduction, the matrix A corresponds to kQ_k for some particular k.

Lemma 2.1 *For a given $\varphi(z)$ and associated stability region, S, there exists a real constant a, such that $\forall n$,*

$$|\varphi^n(z)|\le a,\quad \forall z\in\Gamma_n,$$

where Γ_n is defined as

$$\Gamma_n=\left\{z : d(z,S)=n^{-1}\right\}$$

or equivalently as the boundary of S_n, defined as

$$S_n=\left\{z : d(z,S)\le n^{-1}\right\}.$$

Proof $|\varphi(z)|=1$ on the boundary of the stability region, ∂S and $\varphi'(z)$ is bounded in $\{z : d(z,S)\le 1\}$, so there exists a positive constant b such that

$$|\varphi(z)|\le \exp\left(b\, d(z,S)\right),\ \forall z : d(z,S)\le 1$$

The result then follows directly, setting $a=\mathrm{e}^b$. $\square$

Theorem 2.2 *For a given $\varphi(z)$, there exists a constant M such that if $\tau(A) \subset S$ then $\|\varphi^n(A)\| \leq Mn$*

Remark: This result is due to Lenferink and Spijker [6] and Reddy and Trefethen [8]; a closely related result has been proved by Lubich and Nevanlinna [7]. The theorem is included here for completeness and to introduce the method of proof used in the subsequent new results.

Proof Using the Cauchy integral formula,

$$\varphi^n(A) = \frac{1}{2\pi \mathrm{i}} \int_{\Gamma_n} \varphi^n(z)(zI - A)^{-1} \mathrm{d}z$$

For $z \in \Gamma_n$, $|\varphi^n(z)| \leq a$ by the previous lemma, and $\left\|(zI - A)^{-1}\right\| \leq n$ because of the resolvent condition in the definition of the numerical range. Therefore,

$$\|\varphi^n(A)\| \leq \frac{1}{2\pi} \int_{\Gamma_n} |\varphi^n(z)| \left\|(zI - A)^{-1}\right\| |\mathrm{d}z| \leq \frac{Pan}{2\pi},$$

where P is the length of the contour Γ_1. $\square$

Since the boundary of the stability region, ∂S, has finite curvature at $z = 0$, in a neighbourhood of $z = 0$ it can be described by $x = x_S(y)$. The next results consider matrices A for which $\tau(A) \subset V$ with V being a closed convex set satisfying the following conditions:

i) $V \subset \mathrm{int}(S) \cup \{0\}$

ii) there are positive real constants c, ϵ and $r \geq 0$ such that for $|z| \leq \epsilon$, ∂V can be described by $x = x_V(y)$, where $c\,|y|^{r+1} \leq x_S(y) - x_V(y) \leq 2c\,|y|^{r+1}$, and $\left|\dfrac{dx_V}{dy}\right| \leq 1$.

Given such a set V, we define associated sets V_n by

$$V_n = \left\{z : d(z, V) \leq n^{-1}\right\}, \quad n > 0$$

Since V is convex, V_n is also convex and therefore has a rectifiable boundary ∂V_n [10]. Furthermore, from condition ii) above, it follows that $n^{-1} \in \partial V_n$ and that in the neighbourhood of n^{-1} the boundary can be represented parametrically as

$$\left(x_V(y) + n^{-1}\cos(\theta(y)), \quad y + n^{-1}\sin(\theta(y))\right)$$

where $(x_V(y), y)$ are the coordinates of the nearest point on ∂V and $\theta(y)$ is the angle of the outward normal to ∂V (and the corresponding point on ∂V_n) given by

$$\tan(\theta(y)) = -\frac{dx_V}{dy}.$$

Figure 1 illustrates the set V_5 for the case in which V is a half-disk and S is the stability region of the four stage Runge-Kutta method commonly used in CFD computations.

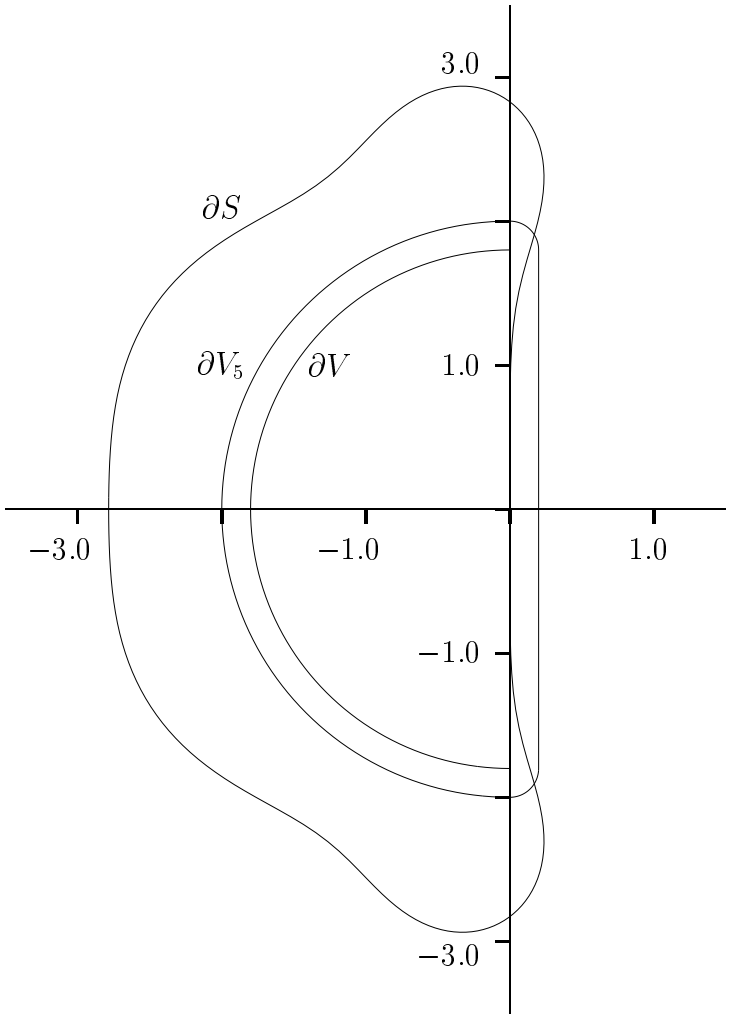

Figure 1: The stability region S for the 4-stage Runge-Kutta method, a half-disk V and the set $V_5 = \left\{z : d(z, V) \leq \frac{1}{5}\right\}$.

Lemma 2.3 *For a given $\varphi(z)$ and V satisfying the above conditions, there exist strictly positive real constants δ, a, b and a positive integer n_1 such that $\forall n \geq n_1$,*

$$|\varphi^n(z)| \leq \begin{cases} a\mathrm{e}^{-\frac{1}{2}nc|y|^{r+1}}, & \forall z \in \Gamma_n^{(1)} \equiv \Gamma_n \cap D_\delta \\ \mathrm{e}^{-bn}, & \forall z \in \Gamma_n^{(2)} \equiv \Gamma_n \cap D_\delta^c \end{cases}$$

where Γ_n is now defined to be the boundary of V_n, and the closed disc D_δ and its closed complement D_δ^c are defined as

$$D_\delta = \{z : |z| \leq \delta\}, \quad D_\delta^c = \{z : |z| \geq \delta\}.$$

Proof $\varphi'(z)=1$ at $z=0$ so we can choose δ, n_0 such that for $|y| \leq \delta \leq \epsilon$

$$|\varphi| \leq \begin{cases} \mathrm{e}^{2(x-x_S(y))}, & x_S(y) \leq x \leq x_S(y) + 2n_0^{-1}, \\ \mathrm{e}^{\frac{1}{2}(x-x_S(y))}, & x_V(y) \leq x \leq x_S(y). \end{cases}$$

These two inequalities can be combined to give the result that for $z \in \Gamma_n^{(1)}$, $n \geq n_0$,

$$|\varphi^n| \leq \mathrm{e}^3 \mathrm{e}^{\frac{1}{2}n(x-x_S(y))} \leq \mathrm{e}^4 \mathrm{e}^{-\frac{1}{2}nc|y|^{r+1}}$$

since

$$x - x_S(y) \;=\; (x - x_V(y)) \;-\; (x_S(y) - x_V(y)) \;\leq\; 2n^{-1} - c|y|^{r+1}.$$

To prove the result for $z \in \Gamma_n^{(2)}$, note that since $(V \cap D_\delta^c) \subset \mathrm{int}(S)$, it is possible to choose $n_1 \geq n_0$ such that $(V_{n_1} \cap D_\delta^c) \subset \mathrm{int}(S)$. The constant b can then be defined by

$$\mathrm{e}^{-b} = \sup_{V_{n_1} \cap D_\delta^c} |\varphi(z)| < 1.$$

□

Theorem 2.4 *For a given $\varphi(z)$ and V satisfying the same conditions as in Lemma 2.3, there exists a constant M such that if $\tau(A) \subset V$ then*

$$\|\varphi^n(A)\| \leq M\, n^q, \qquad q = 1 - \tfrac{1}{r+1}.$$

Remark: an alternative proof of this theorem has recently appeared in a paper by Spijker and Straetemans [11]. It is also related to Theorem 3.3 of Lubich and Nevanlinna [7].

Proof We start with the standard Cauchy integral formula,

$$\varphi^n(A) = \frac{1}{2\pi \mathrm{i}} \int_{\Gamma_n} \varphi^n(z)(zI - A)^{-1} \mathrm{d}z$$

where Γ_n is as defined in the last lemma. For all $z \in \Gamma_n$,

$$\left\| (zI - A)^{-1} \right\| \leq n.$$

Using the last lemma, for $z \in \Gamma_n^{(2)}$ $|\varphi^n(z)| \leq \mathrm{e}^{-bn}$, and so

$$\int_{\Gamma_n^{(2)}} |\varphi^n(z)| \left\| (zI - A)^{-1} \right\| |\mathrm{d}z| \leq Pn\mathrm{e}^{-bn} < \frac{P}{b}$$

For $z \in \Gamma_n^{(1)}$, $|\varphi^n(z)| \leq a\mathrm{e}^{-\frac{1}{2}nc|y|^{r+1}}$, and $|\mathrm{d}z| \leq \sqrt{2}\,|\mathrm{d}y|$ since $\left|\frac{dx_V}{dy}\right| \leq 1$. Hence,

$$\int_{\Gamma_n^{(1)}} |\varphi^n(z)| \left\| (zI - A)^{-1} \right\| |\mathrm{d}z| \leq \sqrt{2}\,an \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}nc|y|^{r+1}} \mathrm{d}y = \sqrt{2}\,an^{1-\frac{1}{r+1}} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}c|w|^{r+1}} \mathrm{d}w.$$

□

The final results of this section obtain even stronger stability by placing an additional restriction on $\varphi(z)$.

Lemma 2.5 *For a given $\varphi(z)$ and V satisfying the same conditions as in Lemma 2.3, with the added condition that $\varphi(z) = \mathrm{e}^z + O(z^{s+1})$ with $s > r$, there exist strictly positive real constants δ, a, b and a positive integer n_1 such that $\forall n \geq n_1$,*

$$\max\left\{|\varphi^n|, |\mathrm{e}^{nz}|\right\} \leq \begin{cases} a\mathrm{e}^{-\frac{1}{2}nc|y|^{r+1}}, & \forall z \in \Gamma_n^{(1)} \equiv \Gamma_n \cap D_\delta \\ \mathrm{e}^{-bn}, & \forall z \in \Gamma_n^{(2)} \equiv \Gamma_n \cap D_\delta^c \end{cases}$$

where again $\Gamma_n = \partial V_n$.

Proof Since $s > r$, the degree of tangency between V and the imaginary axis at $z=0$ is the same as the degree of tangency between V and S. Furthermore, $V \subset \mathbb{C}^-$ because of the convexity of V and the fact that $V \subset S$ and ∂S is tangent to the imaginary axis at $z = 0$. The remainder of the proof is similar to that of Lemma 2.3. □

Theorem 2.6 *For a given $\varphi(z)$ and V satisfying the same conditions as in Lemma 2.5, there exists a constant M such that if $\tau(A) \subset V$ then*

$$\|\varphi^n(A)\| \leq M\, n^q, \qquad q = \max(0, 2 - \tfrac{s+2}{r+1}).$$

First remark: an important feature of this theorem is that it proves that $s \geq 2r$ is a sufficient condition for strong stability, so this theorem adds new classes of discretisation to those in the literature which have previously been proved to be strongly stable. In particular, first order 'upwind' discretisations of hyperbolic p.d.e.'s are often of a form for which $r=1$. This theorem therefore gives $s \geq 2$ as a sufficient condition for strong stability for such discretisations; this condition is often satisfied by the methods in common use in CFD computations.

Second remark: Brenner and Thomée [1] have proved a similar result with the improved bound $q = \max(0, \frac{1}{2}(1 - \frac{s+1}{r+1}))$ for A-stable implicit methods for which the stability region S includes the entire left half-plane. Another related result, due to Kraaijevanger *et al* [5], proves strong stability in the maximum norm when $\tau(A)$ is defined using the maximum norm and V is a disk of the form $\{z : |z + \rho| \leq \rho\}$.

Proof

$$\|\varphi^n(A)\| \;\leq\; \|\varphi^n(A) - \exp(nA)\| \;+ \|\exp(nA)\|\,.$$

Given the equivalent definitions of the range of values proved by Spijker in Theorem 5.1 [12], in particular applying condition (iii) with $\zeta = n, \xi = 0, \theta = 0$, it follows that $\|\exp(nA)\| \leq 1$ because $\tau(A) \subset V \subset \mathbb{C}^-$.

To bound $\|\varphi^n(A) - \exp(nA)\|$ we start with the Cauchy integral formula,

$$(\varphi^n(A) - \exp(nA)) = \frac{1}{2\pi \mathrm{i}} \int_{\Gamma_n} (\varphi^n(z) - \exp(nz))\,(zI - A)^{-1} \mathrm{d}z.$$

and again separate the contributions from the two segments, $\Gamma_n^{(1)}$ and $\Gamma_n^{(2)}$.

For $z \in \Gamma_n^{(2)}$, using the last lemma, $|\varphi^n(z) - \exp(nz)| \leq 2\mathrm{e}^{-bn}$ and so

$$\int_{\Gamma_n^{(2)}} |\varphi^n(z) - \exp(nz)| \left\|(zI - A)^{-1}\right\| |\mathrm{d}z| \leq 2Pn\mathrm{e}^{-bn} < \frac{2P}{b}$$

The contour $\Gamma_n^{(1)}$ is now itself broken into three pieces,

$$\begin{aligned}
\Gamma_n^{(1a)} &= \left\{ z \in \Gamma_n^{(1)} : |y| \geq n^{-\frac{1}{t+1}} \right\}, \\
\Gamma_n^{(1b)} &= \left\{ z \in \Gamma_n^{(1)} : n^{-1} \leq |y| \leq n^{-\frac{1}{t+1}} \right\}, \\
\Gamma_n^{(1c)} &= \left\{ z \in \Gamma_n^{(1)} : |y| \leq n^{-1} \right\},
\end{aligned}$$

with t being a constant chosen such that $s > t > r$.

For $z \in \Gamma_n^{(1a)}$, $|\varphi^n(z) - \exp(nz)| \leq 2a\mathrm{e}^{-\frac{1}{2}nc|y|^{r+1}}$ and so

$$\int_{\Gamma_n^{(1a)}} |\varphi^n(z) - \exp(nz)| \left\|(zI - A)^{-1}\right\| |\mathrm{d}z| \leq 4\sqrt{2}\,an \int_{n^{-\frac{1}{t+1}}}^{\infty} \mathrm{e}^{-\frac{1}{2}ncy^{r+1}} \mathrm{d}y$$

Since

$$n \int_{n^{-\frac{1}{t+1}}}^{\infty} \mathrm{e}^{-\frac{1}{2}ncy^{r+1}} \mathrm{d}y = n^{1-\frac{1}{r+1}} \int_{n^{\frac{1}{r+1}-\frac{1}{t+1}}}^{\infty} \mathrm{e}^{-\frac{1}{2}cw^{r+1}} \mathrm{d}w,$$

and this tends to zero as $n \to \infty$, the contribution from $\Gamma_n^{(1a)}$ is bounded.

For $z \in \Gamma_n^{(1b)}$,

$$\mathrm{e}^{-z}\varphi(z) = 1 + O(z^{s+1}) \quad \Longrightarrow \quad \mathrm{e}^{-nz}\varphi^n(z) = 1 + O(nz^{s+1}),$$

with the choice of constant t ensuring that nz^{s+1} remains bounded for all n. In addition, $|z| \leq \sqrt{2}\,|y|$ and $\exp(nz) < a\mathrm{e}^{-\frac{1}{2}nc|y|^{r+1}}$. Hence, there exists a constant g such that

$$|\varphi^n(z) - \exp(nz)| \leq ga\,n|y|^{s+1}\mathrm{e}^{-\frac{1}{2}nc|y|^{r+1}}.$$

Therefore, it follows that

$$\begin{aligned}
\int_{\Gamma_n^{(1b)}} |\varphi^n(z) - \exp(nz)| \left\|(zI - A)^{-1}\right\| |\mathrm{d}z| &\leq 2\sqrt{2}\,ga\,n^2 \int_0^{\infty} y^{s+1}\mathrm{e}^{-\frac{1}{2}ncy^{r+1}} \mathrm{d}y \\
&= 2\sqrt{2}\,ga\,n^{2-\frac{s+2}{r+1}} \int_0^{\infty} w^{s+1}\mathrm{e}^{-\frac{1}{2}cw^{r+1}} \mathrm{d}w.
\end{aligned}$$

Finally, for $z \in \Gamma_n^{(1c)}$, $|z| < \sqrt{2}\, n^{-1}$ and so

$$|\varphi^n(z) - \exp(nz)| \leq g n^{-s} a,$$

and so

$$\int_{\Gamma_n^{(1c)}} |\varphi^n(z) - \exp(nz)| \left\| (zI - A)^{-1} \right\| |\mathrm{d}z| < 2\sqrt{2}\, g a\, n^{-s}.$$

Summing the upper bounds on the magnitudes of each of the contributions to the Cauchy integral completes the proof. □

3 Convergence

To prove convergence for the full discretisation of the p.d.e. presented in the Introduction, under fairly weak sufficient conditions, requires a new form of stability result.

Theorem 3.1 *Provided the roots of $\varphi(z)=1$ are simple, there exists a constant M depending solely on $\varphi(z)$ such that if $\tau(A) \subset S$ then* $\left\| \sum_{j=0}^{n-1} \varphi^j(A) \right\| \leq Mn \log n$

Proof Using the Cauchy integral formula with $\Gamma_n \equiv \partial S_n$,

$$\sum_{j=0}^{n-1} \varphi^j(A) \;=\; \frac{1}{2\pi \mathrm{i}} \int_{\Gamma_n} \sum_{j=0}^{n-1} \varphi^j(z)(zI - A)^{-1} \mathrm{d}z \;=\; \frac{1}{2\pi \mathrm{i}} \int_{\Gamma_n} \frac{\varphi^n(z) - 1}{\varphi(z) - 1} (zI - A)^{-1}\, \mathrm{d}z.$$

Using Lemma 2.1, for $z \in \Gamma_n$,

$$\left| \frac{\varphi^n(z) - 1}{\varphi(z) - 1} \right| \leq \frac{a+1}{|\varphi(z) - 1|}.$$

Since $\varphi'(0)=1$, it is possible to find $\epsilon > 0$ such that for all $z \in D_\epsilon$,

$$|\varphi(z) - 1| > \tfrac{1}{2}|z|$$

and for all n, and all $z \in \Gamma_n \cap D_\epsilon$,

$$\left| \frac{dx}{dy} \right| \leq 1$$

and hence

$$|z| \geq \max\left\{ |y|, \frac{1}{\sqrt{2}\, n} \right\} \geq \tfrac{1}{2}\sqrt{y^2 + n^{-2}}.$$

Therefore,

$$\int_{\Gamma_n \cap D_\epsilon} \left| \frac{\varphi^n(z) - 1}{\varphi(z) - 1} \right| \left\| (zI - A)^{-1} \right\| |\mathrm{d}z| < \int_{-\epsilon}^{\epsilon} \frac{4\sqrt{2}(a+1)n}{\sqrt{y^2 + n^{-2}}}\, \mathrm{d}y < 8\sqrt{2}(a+1)n(\epsilon + \log n),$$

since

$$\int_0^{\epsilon} \frac{\mathrm{d}y}{\sqrt{y^2+n^{-2}}} \;=\; \int_0^{n\epsilon} \frac{\mathrm{d}w}{\sqrt{w^2+1}} \;<\; \int_0^{\epsilon} \mathrm{d}w + \int_{\epsilon}^{n\epsilon} \frac{\mathrm{d}w}{w} \;=\; \epsilon + \log n.$$

Similar neighbourhoods can be constructed around each of the other $p-1$ distinct roots of $\varphi(z)=1$ on ∂S, resulting in similar $O(n \log n)$ bounds on the contribution to the Cauchy integral. The contribution from the remainder of Γ_n is only $O(n)$ since $|\varphi(z) - 1|$ is bounded away from zero and so the integrand is $O(n)$. $\square$

For the family of discretisations described in the Introduction, the solution error $e_n = u(x_k, nk) - u_{k,n}$ satisfies the difference equation

$$e_{k,n+1} = \varphi(kQ_k)\, e_{k,n} + kT_{k,n}, \tag{3.1}$$

with $T_{k,n}$ being the truncation error and with the initial error $e_{k,0}$ being zero. Given these definitions we now prove the following theorem.

Theorem 3.2 *If the roots of $\varphi(z)=1$ are simple, and*

i) $\tau(kQ_k) \subset S, \quad \forall k$

ii) $T_{k,0} = o\left(\dfrac{1}{\log(k^{-1})}\right)$

iii) $\max\limits_{0\le mk\le 1} |T_{k,m} - T_{k,m-1}| = o\left(\dfrac{k}{\log(k^{-1})}\right)$

then $e_{k,n} \to 0$ as $k \to 0,\ nk \to t$, for $0 \le t \le 1$.

Proof Defining

$$B_{k,n} = \sum_{m=0}^{n} \varphi^m(kQ_k)$$

then

$$e_{k,n} = k \sum_{m=0}^{n-1} \varphi^{n-1-m}(kQ_k) T_{k,m} = k\left(B_{k,n-1}T_{k,0} + \sum_{m=1}^{n-1} B_{k,n-1-m}(T_{k,m} - T_{k,m-1})\right)$$

and so

$$|e_{k,n}| \le k\left(\|B_{k,n-1}\|\,|T_{k,0}| + \sum_{m=1}^{n-1} \|B_{k,n-1-m}\|\,|T_{k,m} - T_{k,m-1}|\right)$$

Applying Theorem 3.1 and the conditions on the truncation error completes the proof. $\square$

To weaken the consistency conditions sufficient for convergence, it is necessary to tighten the stability result. We first define the function,

$$s(n,z) \equiv \sum_{j=0}^{\infty} \frac{n^{j+1} z^j}{(j+1)!}, \tag{3.2}$$

with the obvious properties that it is analytic, $\frac{\partial s}{\partial n} = \exp(nz)$, $s(0,z) = 0$ and when $z \neq 0$, $s(n,z) = z^{-1}(\exp(nz) - 1)$.

Similarly, for a square matrix A we define

$$s(n,A) \equiv \sum_{j=0}^{\infty} \frac{n^{j+1}}{(j+1)!} A^j, \tag{3.3}$$

for which $\frac{\partial s}{\partial n} = \exp(nA)$ and $s(0,A) = 0$.

Lemma 3.3 *For a given $\varphi(z)$ and V satisfying the same conditions as in Lemma 2.3, with the added conditions that $\varphi(z) = \mathrm{e}^z + O(z^{r+1})$ and the degree of tangency between V and the imaginary axis at $z = 0$ is also r, there exist strictly positive real constants δ, a, b and a positive integer n_1 such that $\forall n \geq n_1$,*

$$\max\left\{\left|\varphi^n\right|, \left|\mathrm{e}^{nz}\right|\right\} \leq \begin{cases} a\mathrm{e}^{-\frac{1}{2}nc|y|^{r+1}}, & \forall z \in \Gamma_n^{(1)} \equiv \Gamma_n \cap D_\delta \\ \mathrm{e}^{-bn}, & \forall z \in \Gamma_n^{(2)} \equiv \Gamma_n \cap D_\delta^c \end{cases}$$

where again $\Gamma_n = \partial V_n$.

Proof The proof is again similar to that in Lemma 2.3. □

Theorem 3.4 *For a given $\varphi(z)$ and V satisfying the same conditions as in Lemma 3.3 there exists a constant M such that if $\tau(A) \subset V$ then*

$$\left\| \sum_{j=0}^{n-1} \varphi^j(A) \right\| \leq Mn$$

Remark: Using Lemma 2.5 it is straightforward to prove the same result when $\varphi(z) = \mathrm{e}^z + O(z^{s+1})$ with $s > r$.

Proof

$$\left\| \sum_{j=0}^{n-1} \varphi^j(A) \right\| \leq \left\| \sum_{j=0}^{n-1} \varphi^j(A) - s(n,A) \right\| + \left\| s(n,A) \right\|.$$

As in Theorem 2.6, $\|\exp(nA)\| \leq 1$ and so

$$\left| \frac{\mathrm{d}}{\mathrm{d}n} \left\| s(n,A) \right\| \right| \leq \left\| \frac{\mathrm{d}}{\mathrm{d}n} s(n,A) \right\| \leq 1.$$

Since $s(n,A) = 0$ when $n = 0$, it follows that for arbitrary positive n, $\|s(n,A)\| \leq n$.

To bound the other term we use the Cauchy integral formula with $\Gamma_n = \partial V_n$,

$$\sum_{j=0}^{n-1} \varphi^j(A) - s(n,A) = \frac{1}{2\pi \mathrm{i}} \int_{\Gamma_n} \left(\sum_{j=0}^{n-1} \varphi^j(z) - s(n,z) \right) (zI - A)^{-1} \mathrm{d}z,$$

and again separate the contributions from the two segments, $\Gamma_n^{(1)}$ and $\Gamma_n^{(2)}$.

For $z \in \Gamma_n^{(2)}$, using the last lemma,

$$\left|\sum_{j=0}^{n-1} \varphi^j(z)\right| \leq \sum_{j=0}^{n-1} \mathrm{e}^{-bj} < \frac{1}{1-\mathrm{e}^{-b}}$$

and

$$|s(n,z)| = \left|\frac{\mathrm{e}^{nz}-1}{z}\right| \leq \frac{2}{\delta}$$

and hence

$$\int_{\Gamma_n^{(2)}} \left|\sum_{j=0}^{n-1} \varphi^j(z) - s(n,z)\right| \left\|(zI-A)^{-1}\right\| |\mathrm{d}z| \leq Pn\left(\frac{1}{1-\mathrm{e}^{-b}} + \frac{2}{\delta}\right)$$

For $z \in \Gamma_n^{(1)}$, we first note that $z \neq 0$ and $\varphi(z) \neq 0$ and so

$$\sum_{j=0}^{n-1} \varphi^j(z) - s(n,z) = \frac{\varphi^n(z)-1}{\varphi(z)-1} - \frac{\mathrm{e}^{nz}-1}{z} = \frac{\varphi^n(z)-\mathrm{e}^{nz}}{z} - \frac{(\varphi(z)-1-z)(\varphi^n(z)-1)}{z(\varphi(z)-1)}$$

The second term is uniformly bounded $\forall n, z \in \Gamma_n^{(1)}$ and so

$$\int_{\Gamma_n^{(1)}} \left|\frac{(\varphi(z)-1-z)(\varphi^n(z)-1)}{z(\varphi(z)-1)}\right| \left\|(zI-A)^{-1}\right\| |\mathrm{d}z| \leq dn$$

for some constant d.

The corresponding integral for the first term has to be broken into three pieces, as in the proof of Theorem 2.6,

$$\begin{aligned}
\Gamma_n^{(1a)} &= \left\{z \in \Gamma_n^{(1)} : |y| \geq n^{-\frac{1}{r+1}}\right\}. \\
\Gamma_n^{(1b)} &= \left\{z \in \Gamma_n^{(1)} : n^{-1} \leq |y| \leq n^{-\frac{1}{r+1}}\right\}, \\
\Gamma_n^{(1c)} &= \left\{z \in \Gamma_n^{(1)} : |y| \leq n^{-1}\right\}.
\end{aligned}$$

For $z \in \Gamma_n^{(1a)}$,

$$\left|\frac{\varphi^n(z)-\mathrm{e}^{nz}}{z}\right| \leq \frac{2a\mathrm{e}^{-\frac{1}{2}nc|y|^{r+1}}}{|y|}$$

and so

$$\begin{aligned}
\int_{\Gamma_n^{(1a)}} \left|\frac{\varphi^n(z)-\mathrm{e}^{nz}}{z}\right| \left\|(zI-A)^{-1}\right\| |\mathrm{d}z| &\leq 4\sqrt{2}\, an \int_{n^{-\frac{1}{r+1}}}^{\infty} \frac{\mathrm{e}^{-\frac{1}{2}ncy^{r+1}}}{y}\,\mathrm{d}y \\
&= 4\sqrt{2}\, an \int_1^{\infty} \frac{\mathrm{e}^{-\frac{1}{2}cw^{r+1}}}{w}\,\mathrm{d}w.
\end{aligned}$$

For $z \in \Gamma_n^{(1b)}$, following a similar argument to that in the proof of Theorem 2.6, there exists a constant g such that

$$\left|\frac{\varphi^n(z)-\mathrm{e}^{nz}}{z}\right| < ga\, n|y|^r \mathrm{e}^{-\frac{1}{2}nc|y|^{r+1}} < ga\, n|y|^r,$$

and so

$$\int_{\Gamma_n^{(1b)}} \left| \frac{\varphi^n(z) - \mathrm{e}^{nz}}{z} \right| \left\| (zI - A)^{-1} \right\| |\mathrm{d}z| \leq 2\sqrt{2}\, ga\, n^2 \int_0^{n^{-\frac{1}{r+1}}} y^r \,\mathrm{d}y = 2\sqrt{2}\, ga\, n \int_0^1 w^r \,\mathrm{d}w.$$

Similarly, for $z \in \Gamma_n^{(1c)}$,

$$\left| \frac{\varphi^n(z) - \mathrm{e}^{nz}}{z} \right| < gan^{1-r},$$

and so

$$\int_{\Gamma_n^{(1c)}} \left| \frac{\varphi^n(z) - \mathrm{e}^{nz}}{z} \right| \left\| (zI - A)^{-1} \right\| |\mathrm{d}z| < 2\sqrt{2}\, ga\, n^{1-r}.$$

Summing the upper bounds on the magnitudes of each of the contributions to the Cauchy integral completes the proof. □

Theorem 3.5 *For a given $\varphi(z)$ and V satisfying the same conditions as in Lemma 3.3, if*

i) $\tau(kQ_k) \subset V, \quad \forall k$

ii) $T_{k,0} \to 0$ *as* $k \to 0$

iii) $\max\limits_{0 \leq mk \leq 1} |T_{k,m} - T_{k,m-1}| = o(k)$

then $e_{k,n} \to 0$ *as* $k \to 0,\ nk \to t$, *for* $0 \leq t \leq 1$.

Proof The proof is almost identical to that of Theorem 3.2. □

It is important to place the above results in the context of the Lax Equivalence Theorem [9] which proves that strong stability is a necessary and sufficient condition for convergence for all initial data, provided that the discretisation is also consistent for a dense subset of the initial data. In the results in this section, convergence is only proved for a subset of the initial data for which the discretisation is consistent. There is no guarantee of convergence for initial data for which the discretisation is not consistent, or which violates the Lipschitz conditions of the above theorems. This is the natural consequence of the use of the weaker algebraic stability rather than strong stability.

For smooth initial data, convergence in theory is sometimes not achieved in practice because of the explosive growth of rounding errors due to finite precision computer arithmetic. The simplest example of this phenomenon is a discretisation of the simple convection p.d.e. on an infinite domain using a uniform grid. Provided the spatial discretisation and one-step Runge-Kutta time discretisation are consistent, convergence will be achieved in theory for initial data comprising a single Fourier mode. However, if the discretisation does not satisfy the Fourier

stability requirement for all Fourier modes (and so is not strongly stable) then rounding errors will grow exponentially.

This potential problem, of convergence in theory but not in practice, does not arise with the results of this section because the sufficient conditions for convergence for certain initial data are also sufficient conditions for algebraic stability for all other initial data. Therefore, the potential growth of rounding errors is limited. Given the increasing use of 64-bit floating point arithmetic, leading to very small initial rounding errors, it is very unlikely that these will grow to a noticeable level.

In many applications the initial data of interest will satisfy the consistency and Lipschitz conditions of the above theorems. Hence, the convergence results of this section, together with the algebraic stability results of the previous section, give strong support to the idea that stability based on the range of values of the discretisation matrix, is a useful definition of stability.

4 Conclusions

This paper has derived new algebraic stability results bounding the growth of families of matrices of the form $\varphi^n(A)$, in which $\varphi(z)$ is a polynomial arising from an explicit one-step Runge-Kutta time integration. Bounds on $\sum_0^{n-1}\varphi^j(A)$ are also derived to determine conditions for the convergence of approximations of initial value p.d.e.'s based on a spatial discretisation followed by Runge-Kutta time integration.

These results show that although algebraic stability is a weaker definition of stability than the strong stability of Lax, it is nevertheless sufficient to ensure convergence, both in theory and in practice, for a large class of initial conditions. Accordingly, it is a useful practical definition of stability for many applications. Furthermore, expanding upon previous results in the literature, strong stability is proved for certain new classes of discretisations.

The analysis in this paper remains valid when $\varphi(z)$ is a rational function arising from an implicit one-step time integration method. However, direct solution of the implicit equations that arise in three-dimensional applications is so costly that additional approximations, such as approximate factorisation, are often used, limiting the applicability of the results for rational $\varphi(z)$.

References

[1] P. Brenner and V. Thomée. On rational approximations of semigroups. *SIAM Journal of Numerical Analysis*, 16:683–694, 1979.

[2] M.B. Giles. Stability analysis of Galerkin/Runge–Kutta Navier–Stokes discretisations on unstructured grids. AIAA Paper 95-1753, 1995.

[3] S.K. Godunov and V.S. Ryabenkii. *The Theory of Difference Schemes–An Introduction*. North Holland, Amsterdam, 1964.

[4] B. Gustafsson, H.-O. Kreiss, and A. Sundström. Stability theory of difference approximations for mixed initial boundary value problems. II. *Mathematics of Computation*, 26(119):649–686, Jul 1972.

[5] J.F.B.M. Kraaijevanger, H.W.J. Lenferink, and M.N. Spijker. Stepsize restrictions for stability in the numerical solution of ordinary and partial differential equations. *Journal of Computational and Applied Mathematics*, 20:67–81, Nov 1987.

[6] H.W.J. Lenferink and M.N. Spijker. On the use of stability regions in the numerical analysis of initial value problems. *Mathematics of Computation*, 57(195):221–237, 1991.

[7] C. Lubich and O. Nevanlinna. On resolvent conditions and stability estimates. *BIT*, 31:293–313, 1991.

[8] S.C. Reddy and L.N. Trefethen. Stability of the method of lines. *Numerische Mathematik*, 62:235–267, 1992.

[9] R.D. Richtmyer and K.W. Morton. *Difference Methods for Initial-Value Problems*. Wiley-Interscience, 2nd edition, 1967. Reprint edn (1994) Krieger Publishing Company, Malabar.

[10] R.T. Rockafellar. *Convex Analysis*. Princeton, 1970.

[11] M. Spijker and F. Straetemans. Error growth analysis, via stability regions, for discretizations of initial value problems. To appear in BIT, 1997.

[12] M.N. Spijker. Numerical ranges and stability estimates. *Applied Numerical Mathematics*, 13:241–249, 1993.