Global dissipativity for A–stable methods

A. T. Hill

School of Mathematical Sciences, University of Bath, Claverton Down, Bath
BA2 7AY U.K.
e-mail: ath@uk.ac.bath.maths

Abstract: This paper concerns the discretization of the initial value problem $u_t = f(u)$, under the three structural conditions
(i) $f : \mathbb{C}^N \longrightarrow \mathbb{C}^N, \quad \Re e\langle f(u),\, u\rangle \le a - b|u|^2, \quad a \ge 0,\, b > 0, \quad \text{for all } u \in \mathbb{C}^N;$
(ii) $f : \mathbb{C}^N \longrightarrow \mathbb{C}^N, \quad \Re e\langle f(u),\, u\rangle < 0 \quad \text{for } u \in \mathbb{C}^N \setminus B(0,\, R) \quad \text{and } R > 0;$
(iii) $f : W \longrightarrow H, \quad \Re e\langle f(w),\, w\rangle_H \le a - b|w|_H^2, \quad a \ge 0,\, b > 0, \quad \text{for all } w \in W,$
for complex Hilbert spaces $W \subseteq H$.

Dahlquist's G–stability theory is used to show that linear multistep and one–leg methods yield dissipative discretizations for all f satisfying **(i)** if and only if the method $(\rho,\, \sigma)$ is A–stable. Extensions of G–stability theory are made to find necessary and sufficient conditions on $(\rho,\, \sigma)$ for similar properties to hold in cases **(ii)** and **(iii)**. In every case, conditions are found for the strict contractivity of solutions for large initial data, and bounds for the asymptotic rate of decay are calculated in cases **(i)** and **(iii)**.

Keywords and phrases: Multistep Methods, G–stability, Dissipativity.

AMS(MOS) Subject classification: 65L07, 34D05.

1 Introduction

In this paper we consider linear multistep and one–leg approximations of dissipative differential equations. In particular, we study systems where an energy inequality is the mechanism of dissipation, and those numerical methods which preserve that mechanism. For both ODEs and PDEs, dissipativity of the solution is an essential prerequisite for the existence of a compact global attractor. Under certain circumstances, therefore, it may be an important property to be preserved in discretizations. The following standard analysis, which establishes that the solution of a differential equation is dissipative in the presence of a suitable energy inequality, reveals the structural properties of the underlying system that must be exploited in the analysis of its approximations.

Let the inner product $\langle \cdot, \cdot \rangle$ and its corresponding norm $|\cdot|$ be defined on $\mathbb{C}^N$, and let $f : \mathbb{C}^N \longrightarrow \mathbb{C}^N$ be a locally Lipschitz continuous function. Suppose that the differential equation

$$u_t = f(u), \qquad t > 0, \qquad u(0) = u_0 \in \mathbb{C}^N, \tag{1.1}$$

defined on $\mathbb{C}^N$, possesses a unique continuous solution $u \in C[[0, \infty); \mathbb{C}^N]$.

Here, we suppose that f satisfies the structural condition,

$$\Re e \langle f(u), u \rangle \;\le\; a - b|u|^2, \qquad \text{for all } u \in \mathbb{C}^N, \tag{1.2}$$

where $a \geq 0$ and $b > 0$. This condition arises (in an infinite–dimensional context) in the study of several frequently analysed PDEs, including the Navier–Stokes equations, (see [16]). Taking the real part of the inner product of u with (1.1), one obtains

$$\frac{1}{2}\frac{d}{dt}|u(t)|^2 = \Re e\langle f(u), u\rangle \le a - b|u|^2. \tag{1.3}$$

Multiplying by e^{2bt} and integrating from 0 to t, yields

$$|u(t)|^2 \le e^{-2bt}|u_0|^2 + \frac{a}{b}(1 - e^{-2bt}). \tag{1.4}$$

We deduce from (1.4), and definition below, that (1.1, 1.2) is *dissipative* in $\mathbb{C}^N$, and that, for all $\epsilon > 0$, $B(0, \sqrt{a/b} + \epsilon)$ is an absorbing set.

Definition 1.1 *The evolutionary equation (1.1) is* **dissipative** *in $\mathbb{C}^N$ if there is a bounded set $\mathcal{B} \subset \mathbb{C}^N$ such that, for all bounded sets $D \subset \mathbb{C}^N$, there is a time $t_0(D)$, such that, for all initial data $u_0 \in D$, the corresponding solution $u(t)$ is contained in $\mathcal{B}$ for all $t \geq t_0$. $\mathcal{B}$ is called an* **absorbing set** *in $\mathbb{C}^N$.*

In approximating (1.1, 1.2), our objective is to classify numerical methods that are dissipative, according to a discrete version of Definition 1.1. In the context of Runge–Kutta methods, this has recently been achieved by Humphries and Stuart. Here, we consider the case of linear multistep approximations:

$$\sum_{i=0}^{q} \alpha_i U_{n+i} = k \sum_{i=0}^{q} \beta_i f(U_{n+i}). \tag{1.5}$$

The multiple function evaluations on the right–hand side of (1.5) make a direct application of condition (1.2) non–productive, in general. Remarkably, Dahlquist established in [6] that all trajectories of linear multistep methods are also trajectories of the corresponding one–leg method (for different initial data), and vice–versa. The one–leg method

$$\sum_{i=0}^{q} \alpha_i U_{n+i} = k f\left(\sum_{i=0}^{q} \beta_i U_{n+i}\right) \tag{1.6}$$

is more suited to energy analysis, since, as its name suggests, it has only one function evaluation. Dissipativity results, established initially in this context, may thus be extended to (1.5) using Dahlquist's equivalence.

In Sections 2, 3 and 4, we consider approximations in three different contexts:
(i) ODEs in $\mathbb{C}^N$, where f satisfies (1.2);
(ii) ODEs in $\mathbb{C}^N$, where f satisfies the more general condition

$$\Re e\langle f(u),\, u\rangle < 0 \qquad \text{for all } u \in \mathbb{C}^N \setminus B(0,\, R), \tag{1.7}$$

for some $R > 0$;
(iii) ODEs on a complex Hilbert space H, where f satisfies an analogue of (1.2).

In Section 2, we show that irreducible one–leg and linear multistep methods are dissipative for all f satisfying (1.2), if and only if the method is A–stable. The essential feature of the proof is Dahlquist's G–stability result [7], which permits one to write the left–hand side of the real part of the inner product of (1.6) with $\sum_{i=0}^{q} \beta_i U_{n+i}$ as a sort of discrete derivative, and enables the use of a discrete evolutionary energy analysis analgous to that commonly used in continuous problems, e.g. for the Navier–Stokes equations [16, Chapter 3].

In Section 3, we consider case **(ii)**. We are not able to classify dissipative methods. Instead, we identify those one–leg methods whose solutions strictly decrease in a G–norm, (see Definition 2.4), exterior to some absorbing set. This implies dissipativity of those methods, and their linear multistep counterparts. Necessary and sufficient conditions, strictly stronger than A–stability, are obtained for this strict decay property to hold.

In Section 4, we show that the one–leg method (1.6) is dissipative for (1.1, 1.2), when $\mathbb{C}^N$ is replaced by a general complex Hilbert space, if and only if the method is A–stable and its whose stability domain contains the point at infinity. It is also shown that e^{-bt}, the decay rate for large initial data, given for the continuous problem by (1.4), may be emulated by such methods under a further technical assumption. This latter result is also novel in $\mathbb{C}^N$, and is applicable to monotone problems — see (1.8) below.

Most previous work on asymptotic stability of which we are aware has been carried out in the context of ODEs in $\mathbb{C}^N$. Investigations into the asymptotic approximation properties of linear multistep methods were initiated by Dahlquist in [5], where he established the concept of A–stability. Later work in [6] and [7] was devoted to equations of the form (1.1) in $\mathbb{C}^N$, for f satisfying the *monotone* structural condition,

$$\langle f(u) - f(v),\, u - v\rangle \leq -b|u-v|^2, \quad b \geq 0 \tag{1.8}$$

which implies (weak) contractivity for the ODE. Having introduced G–norms in [6], Dahlquist established in [7] that, for linear multistep methods,

A–stability $\iff$ Stability of one–leg approximations of (1.1, 1.8) in a G–norm.

Dahlquist's proof of this result relied on a decomposition of A–stable methods, based on the theory of harmonic functions. This decomposition is essential to the work in Section 3 of this paper. A later, more direct, proof was given by Baiocchi and Crouzeix [1].

Parallel work in the context of Runge–Kutta methods was carried out by Burrage and Butcher [3, 2], who defined the condition of algebraic stability in terms of the Butcher tableau, and showed that it implied the stability of approximations of (1.1, 1.8). Recently, Humphries and Stuart [12] showed that, for DJ–irreducible Runge–Kutta methods, algebraic stability is also sufficient to imply the dissipativity of solutions of (1.1, 1.2) in $\mathbb{C}^N$.

In the above context, the results of Section 2 of this paper, which show that

A–stability $\iff$ Dissipativity of one–leg methods for (1.1, 1.2) in $\mathbb{C}^N$,

in some sense completes a set of nonlinear stability results for Runge–Kutta and linear multistep methods and the structural conditions (1.8) and (1.2).

2 Finite–dimensional dissipativity I

The linear multistep method (1.5), or its one–leg counterpart (1.6), will be said to be determined by 'the method $(\rho,\, \sigma)$', where ρ and σ are defined by

$$\rho(\xi) \equiv \sum_{i=0}^{q} \alpha_i \xi^i \quad \sigma(\xi) \equiv \sum_{i=0}^{q} \beta_i \xi^i; \quad |\alpha_0| + |\beta_0| > 0, \quad \alpha_q \neq 0, \quad \alpha_i,\, \beta_i \in \mathbb{R}, \tag{2.1}$$

with the additional requirement that

$$\sigma(1) = \sum_{i=0}^{q} \beta_i \neq 0. \tag{2.2}$$

Below, we collect several definitions and results that are required in the sequel.

Definition 2.1 *A method* (ρ, σ) *is* **irreducible** *if* $\rho(\xi)$ *and* $\sigma(\xi)$ *are coprime polynomials. A method is* **strictly stable** *if*

$$\rho(\xi) = 0 \quad \textit{for} \quad |\xi| = 1 \implies \xi = 1 \quad \textit{and}\ \rho'(1) \neq 0. \tag{2.3}$$

A polynomial $P(\xi)$ *satisfies the* **root condition** *if all of its roots lie in the closed unit disk* $\overline{B}(0, 1)$*, and all roots on* $\partial B(0, 1)$ *are simple. The* **stability domain** *for a method* (ρ, σ) *is the set*

$$S \equiv \{\lambda \in \mathbb{C} \text{ s.t. } \rho(\xi) - \lambda\sigma(\xi) \text{ obeys the root condition.}\} \tag{2.4}$$

A method is **A–stable** *if* $S \supseteq \mathbb{C}^- \equiv \{\lambda \in \mathbb{C} \ s.t.\ \Re e[\lambda] \leq 0\}$*. A method is* **strongly A–stable** *if it is A–stable and* $\sigma(\xi)$ *has no roots on* $\partial B(0, 1)$*.*

Definition 2.2 *A method* (ρ, σ) *is consistent (and at least first–order) if*

$$\rho(1) = 0 \qquad \textit{and} \qquad \rho'(1) = \sigma(1). \tag{2.5}$$

If, in addition to (2.5),

$$\rho'(1) + \rho''(1) = 2\sigma'(1), \tag{2.6}$$

then the method is at least second–order accurate.

Theorem 2.1 (Dahlquist [5]) *If* (ρ, σ) *is an A–stable method, it must be of order* $p \leq 2$*. If* $p = 2$*, then*

$$\rho(e^z) - z\sigma(e^z) = c_3 z^3 + c_4 z^4 + O(z^5), \tag{2.7}$$

where $c_3/\sigma(1) \geq -1/12$*. The only A–stable method for which* $c_3/\sigma(1) = -1/12$ *is the trapezoidal rule. For irreducible methods, A–stability is equivalent to*

$$\Re e[\rho(\xi)\overline{\sigma(\xi)}] > 0 \qquad \textit{for all}\ |\xi| > 1. \tag{2.8}$$

For a q–step method, $\sigma(\xi)$ *is of degree q and* $\alpha_q/\beta_q > 0$*.*

Definition 2.3 *The shift operator E is defined on a sequence* (x_n) *by*

$$Ex_n \mapsto x_{n+1}. \tag{2.9}$$

Our main results in this paper depend upon Dahlquist's G–stability theorem. In order to state this, we need the following definition, which differs slightly from standard terminology.

Definition 2.4 *Suppose that* $G = (g_{ij})$ *is a real* $q \times q$ *symmetric positive definite matrix, and that, for some* $N \in \mathbb{N}$*,* $\langle\cdot, \cdot\rangle$ *is an inner product on* $\mathbb{C}^N$*, then a norm is defined on* $(\mathbb{C}^N)^q$ *by*

$$\|\mathbf{U}\|_G^2 = \sum_{i=1}^{q}\sum_{j=1}^{q} g_{ij}\langle U_i, U_j\rangle, \qquad \textit{for}\ \mathbf{U} \equiv [U_1, \ldots, U_q]^T, \quad U_i \in \mathbb{C}^N. \tag{2.10}$$

If, for a given method (ρ, σ), there is such a matrix G, such that, for all $N \in \mathbb{N}$, all inner products on $\mathbb{C}^N$, and any sequence (U_n) defined on $\mathbb{C}^N$,

$$2\Re e \, \langle \rho(E)U_n, \, \sigma(E)U_n \rangle \geq \|\mathbf{U}_{n+1}\|_G^2 - \|\mathbf{U}_n\|_G^2 \qquad \textit{for all } n \geq 0, \tag{2.11}$$

then we say that the method (ρ, σ) is **G–stable***, and that $\| \cdot \|_G$ is a* **G–norm** *for the method.*

Theorem 2.2 (Dahlquist [7]) *Suppose that the method (ρ, σ) is irreducible and A–stable. Then, there exist a real symmetric $q \times q$ positive definite matrix $G = (g_{ij})$, and a real symmetric $(q+1) \times (q+1)$ non–negative definite matrix S, such that, for all $\xi, \omega \in \mathbb{C}$,*

$$\rho(\xi)\sigma(\omega) + \rho(\omega)\sigma(\xi) = (\xi\omega - 1) \sum_{i,j=1}^{q} g_{ij}\xi^{i-1}\omega^{j-1} + \sum_{i,j=1}^{q+1} s_{ij}\xi^{i-1}\omega^{j-1}. \tag{2.12}$$

Equivalently, an irreducible method is G–stable if and only if it is A–stable.

$\rho(E)U_n/(k\sigma(1))$, is an approximation to du/dt in (1.5). Bearing this in mind, it can be seen that, for the one–step formulation of the linear multistep method, (2.11) is in some sense a discrete version of the identity

$$2\Re e\langle u_t, \, u\rangle = \frac{d}{dt}|u|^2, \tag{2.13}$$

which is of fundamental importance in energy analysis. (2.11) is of similar significance for the main results in this section and the next.

The system we consider approximating here is (1.1, 1.2). To avoid the problems involved in a direct analysis of the linear multistep approximation (1.5, 1.2), referred to in the Introduction, we initially analyse a one–leg approximation of the form (1.6),

$$\rho(E)U_n = kf(\sigma(E)U_n).$$

(1.6) possesses an extra degree of freedom compared with (1.5), due to the loss of linearity, and it is therefore unnecessary to normalize here in order to determine a unique representation for the method. However, a consistent method requires $\sigma(1) = 1$, in addition to (2.5). The order of accuracy of (1.6) is never greater, and sometimes less, than that of the corresponding method (1.5), see [6, 10].

Our main objective in this section is to show that when solutions exist for (1.6, 1.2), they are dissipative if and only if the method (ρ, σ) is A–stable. In the literature of discrete dynamical systems, dissipativity is defined in terms of a discrete semigroup, see [13]. In this finite–dimensional context, however, the following definition suffices.

Definition 2.5 *A multistep method of the form (1.5), or a one–leg method of the form (1.6), is* **dissipative** *under the structural condition (1.2), (respectively (1.7)), if for all* $N \in \mathbb{N}$, *all* $f : \mathbb{C}^N \longrightarrow \mathbb{C}^N$ *satisfying (1.2), (respectively (1.7)), and all* $k > 0$, *there is a bounded set* $\mathcal{B} \in (\mathbb{C}^N)^q$ *such that, given any solution sequence of the numerical approximation of (1.1, 1.2), (respectively (1.1, 1.7)),* $\mathbf{U}_0, \mathbf{U}_1 \ldots$ *in* $(\mathbb{C}^N)^q$, *there is an* $n_0 \in \mathbb{N}$ *such that* $\mathbf{U}_n \in \mathcal{B}$ *for all* $n \geq n_0$.

The above definition applies whether or not any solution sequences actually exist, and this is in line with the general philosophy of this paper, which is to find conditions for dissipativity of solutions whenever they exist, rather than to concentrate on the practical construction of solutions. We note that the contraction mapping theorem ensures the existence of a solution U_{n+q} for (1.6), which is unique in a neighbourhood of $-\sum_{i=0}^{q-1} \alpha_i U_{n+i}/\alpha_q$, provided that

$$k \leq \frac{\alpha_q}{L\beta_q}, \tag{2.14}$$

where L is a local Lipschitz constant for f. However, in view of our objectives, which are to consider the properties of (1.6) for a fixed step–size k and arbitrarily large $\mathbf{U}_n$, such a restriction is intolerable. In certain circumstances, say if f also satisfies (1.8), unique existence of solutions to (1.6) follows without any bound on k, (see [6]), but we know of no general uniqueness theorem for systems of the form (1.6, 1.2).

However, in the sequel we will establish results of the form: if U_{n+q} exists, then $\|\mathbf{U}_{n+1}\|_G^2 \leq \|\mathbf{U}_n\|_G^2 + 2ka$. Similar results are also obtained in Section 3. Given such an *a priori* bound, one may then apply the (nonconstructive) Brouwer fixed point theorem to deduce the existence of at least one solution U_{n+q}, as we show below. Such a solution may not be constructible in general.

Proposition 2.3 *Let* (ρ, σ) *be a consistent method such that* $\alpha_q/\beta_q > 0$; *let* $f : \mathbb{C}^N \longrightarrow \mathbb{C}^N$ *be continuous, and suppose that* $M > 0$ *and* $U_0, U_1, \ldots, U_{q-1} \in \overline{B}(0, M/A) \subset \mathbb{C}^N$ *are given quantities, where* $A = \sum_{i=0}^{q} |\alpha_i| > 0$. *Suppose, in addition, that any possible solution* U_q *of (1.6), with* $n = 0$, *must satisfy the* a priori *bound* $|U_q| \leq M/A$, *uniformly for all* $k \in (0, k_0]$, *for some* $k_0 > 0$. *Then, there exists at least one solution* U_q *of (1.6) for all* $k \in (0, k_0]$.

Proof: Using the change of variables, $Y = \sum_{i=0}^{q} \alpha_i U_i$, $Z = \sum_{i=0}^{q-1} (\beta_i - \beta_q \alpha_i/\alpha_q) U_i$, we may rewrite (1.6) as the fixed point equation

$$Y = kf(\beta_q Y/\alpha_q + Z), \tag{2.15}$$

where Z is known. Any solution Y satisfies $|Y| \leq M$, for all $k \in (0, k_0]$. Changing variables again, let $y = Y/(2M)$ and define

$$G(y) \equiv \begin{cases} \dfrac{kf(2\beta_q My/\alpha_q + Z)}{2M}, & \text{for} \quad \dfrac{|kf(2\beta_q My/\alpha_q + Z)|}{2M} < 1 \\[2ex] \dfrac{f(2\beta_q My/\alpha_q + Z)}{|f(2\beta_q My/\alpha_q + Z)|}, & \text{for} \quad \dfrac{|kf(2\beta_q My/\alpha_q + Z)|}{2M} \geq 1. \end{cases} \tag{2.16}$$

$G : \overline{B}(0, 1) \longrightarrow \overline{B}(0, 1)$ is a continuous mapping from a closed convex set to itself, and therefore, by the Brouwer fixed point theorem, (see [15, p. 639]), possesses a fixed point $y^* \in \overline{B}(0, 1)$. If $|y^*| = 1$, we may define $Y^* \equiv 2My^*$, so that $|Y^*| = 2M$ and Y^* satisfies

$$\begin{aligned} Y^* &= \frac{2Mf(\beta_q Y^*/\alpha_q + Z)}{|f(\beta_q Y^*/\alpha_q + Z)|} \\ &= hf(\beta_q Y^*/\alpha_q + Z), \qquad \text{for } h \equiv \frac{2M}{|f(\beta_q Y^*/\alpha_q + Z)|} \leq k \leq k_0. \end{aligned}$$

Hence, by the *a priori* bound $|Y^*| \leq M$ for (2.15), with h replacing k, we obtain a contradiction.

Thus, $|y^*| < 1$, and, reversing the change of variables, we see that $Y^* \equiv 2My^*$ is a solution of (2.15) and $U_q \equiv (Y^* - \sum_{i=0}^{q-1} \alpha_i U_i)/\alpha_q$ is a solution of (1.6).

Q.E.D.

As a first step towards the main result of this section, we show that A–stability is a necessary property for dissipativity.

Lemma 2.4 *If the method (ρ, σ) is irreducible, and either the linear multistep method (1.5), or the one–leg method (1.6), is dissipative, then (ρ, σ) is A–stable.*

Proof: Let us choose any λ such that $\Re e[\lambda] < 0$, and define $f(u) \equiv \lambda u$, for $u \in \mathbb{C}^N$. Now, $\Re e\langle f(u), u\rangle = \Re e[\lambda]|u|^2$, and therefore the equation $u_t = \lambda u$ is of the form (1.1, 1.2). Since f is linear, (1.5) and (1.6) are the same, and both have a solution for all initial data and all $n \geq 0$. If any solution of the equation

$$\rho(\xi) - \lambda\sigma(\xi) = 0 \tag{2.17}$$

satisfied $|\xi| \geq 1$ for any λ such that $\Re e[\lambda] < 0$, one could derive a contradiction to dissipativity by considering such an f.

Thus, $\sigma(\eta)$ may not have a root $|\eta| > 1$, since Rouché's theorem would otherwise imply that (2.17) possesses a root $|\xi(\lambda)| > 1$ for all real negative λ sufficiently large. Hence, ρ/σ is analytic for $|\xi| > 1$, and $\Re e[\rho(\xi)/\sigma(\xi)] \geq 0$ for $|\xi| > 1$. By the minimum principle for harmonic functions, we deduce that

$$0 \leq \liminf_{\eta \to \overline{B}(0,1)} \Re e[\rho(\eta)/\sigma(\eta)] < \Re e[\rho(\xi)/\sigma(\xi)], \qquad \text{for } 1 < |\xi|. \tag{2.18}$$

By Theorem 2.1, this implies that the method (ρ, σ) is A–stable. **Q.E.D.**

The equivalence between one–leg and linear multistep methods is expressed in the following theorem.

Theorem 2.5 (Dahlquist[6]) *Let $(\rho,\, \sigma)$ be an irreducible method, and suppose that $U_0,\, U_1,\, \ldots \in \mathbb{C}^N$ is a sequence which satisfies (1.6). Then, the sequence $V_0,\, V_1,\, \ldots$, where $V_n \equiv \sigma(E)U_n$, satisfies (1.5).*

Conversely, there are real coefficients $\{a_i\}_0^{q-1}$ and $\{b_i\}_0^{q-1}$, such that if the sequence $V_0,\, V_1,\, \ldots \in \mathbb{C}^N$ satisfies (1.5), then (1.6) is satisfied by the sequence $U_0,\, U_1,\, \ldots$, where

$$U_n \equiv \sum_{i=0}^{q-1} a_i V_{n+i} - k \sum_{i=0}^{q-1} b_i f(V_{n+i}). \tag{2.19}$$

The following lemma is needed for the proof of our main result.

Lemma 2.6 *Suppose that for $i = 1, \ldots, r$, the polynomials $p_i(\xi) = \sum_{j=1}^{r} p_{ij}\xi^{j-1}$ form a basis for $\mathbb{P}^{r-1}$, the space of polynomials of degree less than r, and that the values $f_1,\, \ldots,\, f_r \in \mathbb{C}^N$ are given. Then, the system of equations*

$$\sum_{j=1}^{r} p_{ij} U_j = f_i, \qquad i = 1,\, \ldots,\, r, \tag{2.20}$$

possesses a unique solution $U_1,\, \ldots,\, U_r \in \mathbb{C}^N$. Furthermore, there is a constant $C(p_{ij})$, such that any sequence $V_1,\, \ldots,\, V_r \in \mathbb{C}^N$ obeying the inequality

$$\left| \sum_{j=1}^{r} p_{ij} V_j \right| \le |f_i|, \qquad i = 1,\, \ldots,\, r, \tag{2.21}$$

also satisfies

$$\sup_{1\le i\le r} |V_i| \le C \sup_{1\le i\le r} |f_i|. \tag{2.22}$$

Proof: Because $\{p_i(\xi)\}_{i=1}^{r}$ is a basis, there are unique real coefficients a_{ik} for each $i \in \{1,\, \ldots,\, r\}$, such that $\sum_{k=1}^{r} a_{ik} \sum_{j=1}^{r} p_{kj}\xi^{j-1} = \xi^{i-1}$. Considering the bijection $U_i \longleftrightarrow \xi^{i-1}$, we deduce that the solution of (2.20) is

$$U_i = \sum_{k=1}^{r} a_{ik} f_k, \quad i = 1,\, \ldots,\, r.$$

In (2.22), we may take $C(p_{ij}) = \|A\|_\infty$, where $A = (a_{ij})$. **Q.E.D.**

Theorem 2.7 *Let $(\rho,\, \sigma)$ be an irreducible method. Then, the linear multistep approximation (1.5), (respectively the one–leg approximation (1.6)), is dissipative if and only if $(\rho,\, \sigma)$ is A–stable.*

Proof: By virtue of Lemma 2.4 and Theorem 2.5, it only remains to prove that A–stability is a sufficient condition for dissipativity in the one–leg case. Combining (2.11), (1.6) and (1.2), we obtain

$$\begin{aligned} \|\mathbf{U}_{n+1}\|_G^2 - \|\mathbf{U}_n\|_G^2 &\le 2\Re e\, \langle \rho(E)U_n,\, \sigma(E)U_n \rangle \\ &= 2k\Re e\, \langle f(\sigma(E)U_n),\, \sigma(E)U_n \rangle \\ &\le 2k \left(a - b\, |\sigma(E)U_n|^2 \right). \end{aligned} \tag{2.23}$$

Let us take $m \equiv [\|\mathbf{U}_0\|_G^2/(2qka)]+1$ and $M = mq$, $([x] \equiv \sup\{n \in \mathbb{N} \mid n \le x\})$.

Summing (2.23) over M terms, we obtain

$$\|\mathbf{U}_M\|_G^2 - \|\mathbf{U}_0\|_G^2 \;\le\; 2Mka - 2bk \sum_{n=0}^{M-1} |\sigma(E)U_n|^2 \tag{2.24}$$

$$bk \sum_{r=0}^{m-1} \sum_{n=rq}^{(r+1)q-1} |\sigma(E)U_n|^2 \;\le\; Mka + \|\mathbf{U}_0\|_G^2/2. \tag{2.25}$$

Hence, there exists $s \in [0,\, m-1]$ such that

$$\sum_{n=sq}^{(s+1)q-1} |\sigma(E)U_n|^2 \le \frac{Mka + \|\mathbf{U}_0\|_G^2/2}{mbk} \le 2aq/b.$$

Therefore, for all $n \in [sq,\, (s+1)q-1]$,

$$|\sigma(E)U_n| \le \sqrt{\frac{2aq}{b}}, \qquad |\rho(E)U_n| \le kL, \qquad \text{where } L \equiv \sup_{|u| \le \sqrt{2aq/b}} |f(u)|. \tag{2.26}$$

Now, ρ and σ are coprime, and both are of degree q. Hence, $\{\xi^i\rho(\xi),\, \xi^i\sigma(\xi);\ i = 0,\, \ldots,\, q-1\}$ form a basis for $\mathbb{P}^{2q-1}$. Applying Lemma 2.6 to (2.26), we deduce that there is a constant $R_0(k,\, L)$, independent of $\mathbf{U}_0$, such that $\|\mathbf{U}_{sq}\|_G \le R_0$.

Suppose that for some $r \in \mathbb{N}$, $\mathbf{U}_r \in \overline{B}(0,\, R_0) \subset ((\mathbb{C}^N)^q,\, \|\cdot\|_G)$. A repetition of the above analysis implies that the sequence $(\mathbf{U}_n)$ will return to $\overline{B}(0,\, R_0)$ within $q[R_0^2/(2qka)]$ terms. For the intervening terms, (2.24) implies that

$$\|\mathbf{U}_n\|_G^2 \le 2(n-r)ka + \|\mathbf{U}_s\|_G^2 \le 2R_0^2.$$

Hence, for all $n \ge \|\mathbf{U}_0\|_G^2/(2ka) + q$, $\|\mathbf{U}_n\|_G^2 \le 2R_0^2 + 2qka$. **Q.E.D.**

Two quantitative questions associated with the above qualitative result are,
(a) How close is the numerical absorbing ball to the true one?
(b) How does the decay rate of numerical solutions with large initial data compared with e^{-bt}, the estimate for the true rate, given by (1.4)?

(b) will be answered in Section 4. As for **(a)**, once one absorbing set has been found, the problem of whether there is a smaller one becomes local in $(\mathbb{C}^N)^q$. More refined estimates for the size of an absorbing set may then be obtained, either by the method of Hale, Lin and Raugel [14, Proposition 2.1], (which was generalized to multistep methods in [11, Section 4]), which depends on accurately approximating the true solution, or by the technique of Eirola and Nevanlinna [8, Theorem 1, part 4], which analyses the numerical solution directly, but only applies when the method $(\rho,\, \sigma)$ is strictly stable.

3 Finite–dimensional dissipativity II

In $\mathbb{C}^N$, (1.2) is a stronger condition than is necessary in order to establish the dissipativity of the ODE (1.1). A weaker sufficient condition is (1.7): $f : \mathbb{C}^N \longrightarrow \mathbb{C}^N$ is Lipschitz continuous and, for some $R > 0$,

$$\Re e\langle f(u),\, u\rangle < 0 \qquad \text{for all } u \in \mathbb{C}^N \setminus B(0,\, R).$$

Note that if f satisfies (1.2), it satisfies (1.7) for $R = \sqrt{a/b} + \epsilon$ and any $\epsilon > 0$. By virtue of Theorem 2.5, an analysis of linear multistep approximations (1.5) may be reduced to that of the corresponding one–leg method

$$\sum_{i=0}^{q} \alpha_i U_{n+i} = kf\left(\sum_{i=0}^{q} \beta_i U_{n+i}\right).$$

As in Section 2, our ultimate goal is to identify the class of methods that are dissipative according to Definition 2.5. Lemma 2.4 implies that, if the method $(\rho,\, \sigma)$ is irreducible, then it must also be A–stable in order for the corresponding one–leg or multistep method to be dissipative. However, we have not been able to show that A–stability is a sufficient condition for dissipativity, nor have we found a counterexample, and, to our knowledge, the problem remains open. Answering a slightly different proplem, we have sought to find those methods that satisfy a certain sufficient condition for dissipativity, which we call strict dissipativity. This property is very much analogous to the notion of B–stability, coined by Butcher [3] for Runge–Kutta approximations of (1.1, 1.8).

Definition 3.1 *If, for a given irreducible method $(\rho,\, \sigma)$, there is a real $q \times q$ symmetric positive definite matrix G, such that $\|\cdot\|_G$ is a G–norm for $(\rho,\, \sigma)$, and such that, for all $N \in \mathbb{N}$, all inner products on $\mathbb{C}^N$, all $f : \mathbb{C}^N \to \mathbb{C}^N$ and all $k > 0$, there is a constant $R_1(k)$ such that, for all solution sequences $\mathbf{U}_0, \mathbf{U}_1, \ldots \in \mathbb{C}^N$ of the corresponding one–leg method (1.6, 1.7),*

$$\|\mathbf{U}_{n+1}\|_G^2 \geq \|\mathbf{U}_n\|_G^2 \Longrightarrow \|\mathbf{U}_{n+1}\|_G^2 \leq R_1,$$

then we say that the method $(\rho,\, \sigma)$ is **strictly dissipative***, and we call $\|\cdot\|_G$ a* **strict G–norm** *for the method $(\rho,\, \sigma)$.*

Strict dissipativity corresponds to a property of the solutions of (1.1, 1.7), which decrease strictly in norm when $|u(t)| > R$. We remark that every G–norm is not necessarily a strict G–norm, even when $(\rho,\, \sigma)$ is strictly dissipative. Furthermore, strict dissipativity says nothing about the decrease of $\|\mathbf{U}_n\|_G$, when $(\mathbf{U}_n)$ is a solution sequence for the linear multistep method (1.6, 1.7).

Strict dissipativity implies the bound

$$\|\mathbf{U}_{n+1}\|_G \leq \max\{R_1,\, \|\mathbf{U}_n\|_G\}. \tag{3.1}$$

If $R_1(k)$ is also uniformly bounded as k tends to 0, then (3.1) implies that strictly dissipative methods satisfy the hypotheses of Proposition 2.3, and thus possess solution sequences for all $\mathbf{U}_0 \in (\mathbb{C}^N)^q$. A second consequence of (3.1) is that

$$\mathbf{U}_{n_0} \in \overline{B}(0,\, R_1) \Longrightarrow \mathbf{U}_n \in \overline{B}(0,\, R_1), \qquad \text{for all } n \geq n_0. \tag{3.2}$$

This property is exploited below in the proof of the following result.

Proposition 3.1 *The methods (1.5, 1.7) and (1.6, 1.7) are both dissipative if the method* $(\rho,\, \sigma)$ *is strictly dissipative.*

Proof: After Theorem 2.5, it is sufficient for us to consider the dissipativity of (1.6, 1.7). (3.2) implies that dissipativity can fail only if, for some strict G–norm $\|\cdot\|_G$ and some f, k and $\mathbf{U}_0$, there is a solution sequence $(\mathbf{U}_n)$, such that $\mathbf{U}_n \notin \overline{B}(0,\, R_1)$ for all $n \geq 0$. We suppose that such a sequence exists.

Now, (3.1) implies that $(\|\mathbf{U}_n\|_G)$ is a nonincreasing sequence, which must tend to some limit $R_2 \geq R_1$. By sequential compactness, there exists a subsequence $(\mathbf{U}_{\mu(n)})$, such that

$$\mathbf{U}_{\mu(n)} \longrightarrow \mathbf{V}_0 \qquad \text{and} \qquad \mathbf{U}_{\mu(n)+1} \longrightarrow \mathbf{V}_1, \qquad \text{as } n \to \infty. \tag{3.3}$$

Here, the limit pair $(\mathbf{V}_0,\, \mathbf{V}_1)$ is such that $\mathbf{V}_0 = [V_0,\, \ldots,\, V_{q-1}]^T$, $\mathbf{V}_1 = [V_1,\, \ldots,\, V_q]^T$, and $\|\mathbf{V}_0\|_G = \|\mathbf{V}_1\|_G = R_2$.

(3.3) and the continuity of f imply that

$$\sum_{i=0}^{q} \alpha_i V_i = \lim_{n\to\infty} \sum_{i=0}^{q} \alpha_i U_{\mu(n)+i} = k \lim_{n\to\infty} f\left(\sum_{i=0}^{q} \beta_i U_{\mu(n)+i}\right) = kf\left(\sum_{i=0}^{q} \beta_i V_i\right). \tag{3.4}$$

Hence, $V_0,\, \ldots,\, V_q$ satisfy (1.6), and so $\|\mathbf{V}_1\|_G \leq R_1$. We deduce that $R_2 = R_1$. Thus, for any $\epsilon > 0$, there is an $n_0 \in \mathbb{N}$ such that $\mathbf{U}_n \in B(0,\, R_1 + \epsilon)$ for all $n \geq n_0$. Thus, the method is dissipative by Definition 2.5. **Q.E.D.**

We now consider the problem of classifying strictly dissipative methods. The method $(\rho,\, \sigma)$ must be such that a matrix G can be found for which there is sufficient information in the single inequality

$$\|\mathbf{U}_{n+1}\|_G \geq \|\mathbf{U}_n\|_G, \tag{3.5}$$

to imply the existence of a constant dependent only on ρ, σ, N, f and k, which bounds each of the the values $|U_n|,\, \ldots,\, |U_{n+q}|$. Assuming that (3.5) holds for a given method, our objective will be to obtain $q+1$ independent linear constraints on $U_n,\, \ldots,\, U_{n+q}$. Lemma 2.6 would then imply the existence of such a constant.

For a given irreducible A–stable method $(\rho,\, \sigma)$, and any corresponding G–norm $\|\cdot\|_G$, we obtain the following from (2.11) and (1.6):

$$\begin{aligned} \|\mathbf{U}_{n+1}\|_G^2 - \|\mathbf{U}_n\|_G^2 &\leq 2\Re e\langle \rho(E)U_n,\, \sigma(E)U_n\rangle \\ &= 2k\Re e\langle f(\sigma(E)U_n),\, \sigma(E)U_n\rangle. \end{aligned}$$

Thus, (3.5) implies $\Re e\langle f(\sigma(E)U_n),\, \sigma(E)U_n\rangle \geq 0$, and we deduce from (1.7) that

$$|\sigma(E)U_n| < R. \tag{3.6}$$

This is our first constraint.

The second constraint follows from (3.6) and (1.6), since, when $|\sigma(E)U_n| < R$,

$$|\rho(E)U_n| = k|f(\sigma(E)U_n)| \leq kL, \quad \text{where } L = \sup\{|f(u)|\,\text{s.t.}\,|u| < R\}. \tag{3.7}$$

Assuming that the method $(\rho,\, \sigma)$ is irreducible, (3.6) and (3.7) are linearly independent, and are thus sufficient to imply the strict dissipativity of any irreducible one–step A–stable method. In general, however, a further $q-1$ constraints are required to bound $\mathbf{U}_{n+1}$. Replacing $\xi^i\omega^j$ by $\langle U_{n+i},\, U_{n+j}\rangle$ in (2.12), and applying (3.5)–(3.7), we obtain

$$\begin{aligned} \sum_{i,\,j=1}^{q+1} s_{ij}\langle U_{n+i},\, U_{n+j}\rangle &= \|\mathbf{U}_n\|_G^2 - \|\mathbf{U}_{n+1}\|_G^2 + 2\Re e\langle \rho(E)U_n,\, \sigma(E)U_n\rangle \\ &\leq 2kRL. \end{aligned} \tag{3.8}$$

As we shall prove below, (3.6)–(3.8) capture all the possible information to be gleaned from (3.5), and hence all the additional constraints that we require must come, if anywhere, from (3.8). Since S is symmetric non–negative definite, there is a set of orthonormal eigenvectors $\mathbf{a}_1,\, \ldots,\, \mathbf{a}_{q+1}$ of S, with corresponding eigenvalues $\{\mu_m^2\}_{m=1}^{q+1}$, such that

$$\mu_1,\, \mu_2,\, \ldots,\, \mu_r > 0 \qquad \text{and} \qquad \mu_{r+1} = \ldots = \mu_{q+1} = 0, \tag{3.9}$$

where r is the rank of S. The following theorem states necessary and sufficient conditions on S for the method $(\rho,\, \sigma)$ to be strictly dissipative.

Theorem 3.2 *Suppose that $(\rho,\, \sigma)$ is a consistent irreducible A–stable method, and that G and S satisfy (2.12). If $q = 1$, then the method is strictly dissipative. If $q \geq 2$, then $\|\cdot\|_G$ is a strict G–norm if and only if there are orthogonal eigenvectors of S, $\mathbf{a}_1,\, \ldots,\, \mathbf{a}_{q-1} \notin Ker(S)$, such that*

$$\mathrm{Span}\{\sigma(\xi),\, \rho(\xi),\, a_1(\xi),\, \ldots,\, a_{q-1}(\xi)\} = \mathbb{P}^q, \tag{3.10}$$

where $a_i(\xi) \equiv \sum_{j=1}^{q+1} a_{ij}\xi^{j-1}$.

Proof: The case $q = 1$ follows from the discussion after (3.7). Henceforth, we assume $q \geq 2$. If (3.10) holds, then, for given $N \in \mathbb{N}$,

$$|\rho(E)U_n|^2 + |\sigma(E)U_n|^2 + \sum_{i,\,j=1}^{q+1} s_{ij}\langle U_{n+i},\, U_{n+j}\rangle$$

describes a norm on any vector $[U_n, \dots, U_{n+q}]^T \in (\mathbb{C}^N)^{q+1}$, because, by (3.9),

$$\sum_{i,j=1}^{q+1} s_{ij}\langle U_{n+i}, U_{n+j}\rangle = \sum_{m=1}^{r} |\mu_m a_m(E)U_n|^2 \tag{3.11}$$

and $r \geq q-1$. By the equivalence of finite–dimensional norms, there is a positive constant c_0, independent of $\mathbf{U}_n$ and $\mathbf{U}_{n+1}$, such that

$$\begin{aligned} |U_n|^2 + \|\mathbf{U}_{n+1}\|_G^2 &\leq c_0^2 \left\{ |\rho(E)U_n|^2 + |\sigma(E)U_n|^2 + \sum_{i,j=1}^{q+1} s_{ij}\langle U_{n+i}, U_{n+j}\rangle \right\} \\ &\leq c_0^2(R^2 + k^2L^2 + 2kL) = c_0^2(R+kL)^2, \end{aligned}$$

when (3.5) holds. (Here, we have applied (3.6)–(3.8).) Hence, the method is strictly dissipative, with constant $R_1(k) = c_0(R+kL)$.

Now, suppose that (3.10) does not hold, and consider $f(z) \equiv 2z - |z|^2 z$. Clearly, $f : \mathbb{C} \to \mathbb{C}$ is Lipschitz continuous, $f(1) = 1$ and f satisfies (1.2), and thus (1.7), since

$$\langle f(z), z\rangle = f(z)\overline{z} = 2|z|^2 - |z|^4 \leq 4 - 2|z|^2. \tag{3.12}$$

Below, we will obtain a contradiction to strict dissipativity by showing that we can satisfy (1.6) and (3.5) simultaneously with $\|\mathbf{U}_{n+1}\|_G$ of arbitrary size.

Putting $\xi = \omega = 1$ in (2.12), and recalling that $\rho(1) = 0$ for a consistent method, we conclude that $\mathbf{1}^T S \mathbf{1} = 0$, for $\mathbf{1} = [1, \dots, 1]^T$. Substituting $[U_0, \dots, U_q]^T = \mathbf{1}$ into (3.11), we conclude, noting (3.8), that 1 is a root of $a_m(\xi)$, for $m = 1, \dots, r$, where $r = \mathrm{Rank}(S)$. Thus, by (2.2),

$$\sigma(\xi) \notin V_r \equiv \mathrm{Span}\{a_1(\xi), \dots, a_r(\xi)\}. \tag{3.13}$$

By assumption, there is a non–trivial polynomial

$$b(\xi) \in \mathbb{P}^q \ominus \mathrm{Span}\{\sigma(\xi), \rho(\xi), a_1(\xi), \dots, a_r(\xi)\}.$$

When $\rho \notin V_r$, we may find a solution $[U_n, \dots, U_{n+q}]^T \in \mathbb{C}^{q+1}$ to the equations

$$b(E)U_n = L, \;\; \sigma(E)U_n = 1, \;\; \rho(E)U_n = k, \;\; a_m(E)U_n = 0, \;\; m = 1, \dots, r, \tag{3.14}$$

for some large positive L, which, by (3.13), are linearly independent.

When $\rho \in V_r$, $\sum_{m=1}^{r} A_m a_m(\xi) = A_0 \rho(\xi)$, and we may assume that $A_0, A_r \neq 0$. Here, let us find a solution of the linearly independent system

$$b(E)U_n = L, \; \sigma(E)U_n = 1, \; \rho(E)U_n = k, \; a_m(E)U_n = 0, \; m = 1, \dots, r-1. \tag{3.15}$$

All solutions of (3.14) or (3.15) satisfy (1.6) for the f considered. Furthermore, if we take $A_0/A_r = 0$ when $\rho \notin V_r$, (3.8), (3.9) and (3.11) imply

$$\begin{aligned} \|\mathbf{U}_{n+1}\|_G^2 - \|\mathbf{U}_n\|_G^2 &= 2\Re e\langle \rho(E)U_n,\, \sigma(E)U_n\rangle - \sum_{i,\,j=1}^{q+1} s_{ij}\langle U_{n+i},\, U_{n+j}\rangle \\ &= 2k - \sum_{m=1}^{r} |\mu_m a_m(E)U_n|^2 \\ &= 2k - \mu_r^2 A_0^2 k^2 / A_r^2. \end{aligned} \tag{3.16}$$

The right–hand side is always non–negative for sufficiently small k, so (3.14) and (3.15) contradict (3.1), when L is large. **Q.E.D.**

For consistent methods, we have shown that whether a particular G–norm is a strict G–norm depends upon the image space of the corresponding matrix S. Below, we formulate general necessary and sufficient conditions for the strict dissipativity of the method $(\rho,\, \sigma)$ in terms of the function

$$P(\theta) \equiv 2\Re e[\rho(e^{i\theta})\sigma(e^{-i\theta})] \qquad \text{for } \ \theta \in \mathbb{R}, \tag{3.17}$$

which, by (2.12) and (3.11), satisfies the identity

$$P(\theta) = \sum_{j,\,k=1}^{q+1} s_{jk} e^{ij\theta} e^{-ik\theta} = \overline{\mathbf{v}(\theta)}^T S \mathbf{v}(\theta) = \sum_{m=1}^{r} |\mu_m a_m(e^{i\theta})|^2, \tag{3.18}$$

for $\mathbf{v}(\theta) \equiv [1,\, e^{i\theta},\, \ldots,\, e^{iq\theta}]^T$.

Definition 3.2 *We define an irreducible method $(\rho,\, \sigma)$ to be* **strictly A–stable** ($\overline{\mathbf{A}}$**–stable**)*, if it is A–stable, strictly stable, if $P(\theta) > 0$ for $\theta \in \mathbb{R} \setminus \pi\mathbb{Z}$, and*

$$P(\pi + \phi)P(\phi) \neq O(\phi^6). \tag{3.19}$$

The main result of this section is as follows.

Theorem 3.3 *An irreducible consistent method $(\rho,\, \sigma)$ is strictly dissipative if and only if it is strictly A–stable, or one–step and A–stable.*

The first step in the proof of Theorem 3.3 is to establish necessity.

Lemma 3.4 *If an irreducible consistent method $(\rho,\, \sigma)$ is strictly dissipative, then it must be strictly A–stable, or one–step and A–stable.*

Proof: We assume that $(\rho,\, \sigma)$ is a given strictly dissipative method. Propositions 3.1 and 2.4 imply that $(\rho,\, \sigma)$ must be A–stable. There is no more to prove when $q = 1$, so we assume that $q \geq 2$ below. Theorem 3.2 implies that there is a pair $(G,\, S)$ satisfying (2.12), such that Dim(Ker(S)) ≤ 2. On the other hand, we observe, from (3.18), that if $P(\theta) = 0$ for some $\theta \in \mathbb{R}$, then $\mathbf{v}(\theta)$ must be orthogonal to $\mathrm{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_r\}$, and thus $\mathbf{v}(\theta) \in \mathrm{Ker(S)}$. Since S is real symmetric, $\mathbf{v}(-\theta) = \overline{\mathbf{v}(\theta)} \in \mathrm{Ker(S)}$. $\rho(1) = 0$, by consistency; thus, $P(0) = 0$ and $\mathbf{v}(0) \in \mathrm{Ker(S)}$. Hence, by Theorem 3.2, $P(\theta) > 0$ for $\theta \in (0, \pi)$, else

$$\mathrm{Dim(Ker(S))} \geq \mathrm{Dim\ Span}\{\mathbf{v}(0), \mathbf{v}(\theta), \mathbf{v}(-\theta)\} = 3.$$

It remains to consider $P(\theta)$ when $\theta = 0$ and $\theta = \pi$. If $P(\pi + \phi)P(\phi) = O(\phi^6)$, then, since $P(\phi) = O(\phi^{2n})$ for $n \geq 1$, at least one of the following is also true:
(a) $P(\pi + \phi) = O(\phi^4)$;
(b) $P(\pi + \phi) = O(\phi^2)$ and $P(\phi) = O(\phi^4)$;
(c) $P(\phi) = O(\phi^6)$.
When $P(\theta + \phi) = O(\phi^4)$, (3.18) implies that $a_m(e^{i(\theta+\phi)}) = O(\phi^2)$, for $m = 1, \ldots, r$. So, $da_m(e^{i\theta})/d\theta = 0$, and thus,

$$\sum_{j=1}^{q+1} a_m(j-1)e^{i(j-1)\theta} = 0 \implies \mathbf{w}(\theta) \equiv [0,\, e^{i\theta},\, 2e^{i2\theta},\, \ldots,\, qe^{iq\theta}]^T \in \mathrm{Ker(S)}.$$

When $P(\theta + \phi) = O(\phi^6)$, (3.18) implies that $a_m(e^{i(\theta+\phi)}) = O(\phi^3)$, for $m = 1, \ldots, r$. Therefore, $da_m(e^{i\theta})/d\theta = d^2a_m(e^{i\theta})/d\theta^2 = 0$, and so

$$\sum_{j=1}^{q+1} a_m(j-1)^2e^{i(j-1)\theta} = 0 \implies \mathbf{y}(\theta) \equiv [0,\, e^{i\theta},\, 4e^{i2\theta},\, \ldots,\, q^2e^{iq\theta}]^T \in \mathrm{Ker(S)}.$$

In case **(a)**, $\mathbf{v}(0)$, $\mathbf{v}(\pi)$ and $\mathbf{w}(\pi)$ are in $\mathrm{Ker}(S)$. In case **(b)**, $\mathbf{v}(0)$, $\mathbf{w}(0)$ and $\mathbf{v}(\pi)$ are in $\mathrm{Ker}(S)$. In case **(c)**, $\mathbf{v}(0)$, $\mathbf{w}(0)$ and $\mathbf{y}(0)$ are in $\mathrm{Ker}(S)$. In each case, the three vectors are linearly independent, and so we obtain a contradiction to $\mathrm{Dim}(\mathrm{Ker}(S)) \leq 2$.

The remaining alternative to the method (ρ, σ) not being strictly A–stable is that (ρ, σ) is not strictly stable. Since $P(\theta) > 0$ for $\theta \in (0, \pi)$, and consistency implies that $\rho'(1) = \sigma(1) \neq 0$, this can only be so if $\rho(-1) = 0$. In this case, $\rho(\xi) = (1 - \xi^2)r(\xi)$, for some $r \in \mathbb{P}^{q-2}$. Since $\mathbf{v}(0)$ and $\mathbf{v}(\pi)$ are null vectors of S, Theorem 3.2 implies that S must have rank $q - 1$ for (ρ, σ). We deduce from (3.18) that 1 and -1 are both roots of $a_m(\xi)$, for $m = 1, \ldots, q-1$, and thus $a_m(\xi) = (1 - \xi^2)b_m(\xi)$ for some $b_m \in \mathbb{P}^{q-2}$. Since the a_m's are linearly independent, the b_m's must span $\mathbb{P}^{q-2}$, and so $r(\xi) \in \mathrm{Span}\{b_1(\xi), \ldots, b_{q-1}(\xi)\}$. Hence, $\rho(\xi) \in \mathrm{Span}\{a_1(\xi), \ldots, a_{q-1}(\xi)\}$, and therefore

$$\mathrm{Span}\{\sigma(\xi),\, \rho(\xi),\, a_1(\xi),\, \ldots,\, a_{q-1}(\xi)\} \neq \mathbb{P}^q.$$

(3.10) now implies that (ρ, σ) is not strictly dissipative if $\rho(-1) = 0$. Hence, we have shown that strict dissipativity requires strict A–stability. **Q.E.D.**

The following technical proposition is used frequently in the sequel.

Proposition 3.5 *Let $(\rho,\,\sigma)$ be an irreducible A–stable method of order $p \ge 1$. Then,*

$$\rho(e^z) - z\sigma(e^z) = c_2 z^2 + c_3 z^3 + c_4 z^4 + O(z^5), \tag{3.20}$$

and $P(\phi) = O(\phi^4)$ if and only if $p = 2$. If $p = 2$, then

$$\sigma(1)c_4 - \sigma'(1)c_3 \ge 0, \tag{3.21}$$

and $P(\phi) = O(\phi^6)$ if and only if equality holds in (3.21).

Proof: (2.8) and the continuity of $\rho(\xi)$ and $\sigma(\xi)$ imply that $P(\phi) \ge 0$. Hence, for some $d_2 \in \mathbb{R}$,

$$\begin{aligned} & P(\phi) = 2\Re e[\rho(e^{i\phi})\sigma(e^{-i\phi})] = 2\Re e[(\rho(e^{i\phi}) - i\phi\sigma(e^{i\phi}))\sigma(e^{-i\phi})] \\ = \;& 2\Re e[(c_2(i\phi)^2 + c_3(i\phi)^3 + c_4(i\phi)^4 + O(\phi^5))(\sigma(1) - i\phi\sigma'(1) + d_2\phi^2 + O(\phi^3))] \\ = \;& -2c_2\sigma(1)\phi^2 + 2(\sigma(1)c_4 - \sigma'(1)c_3 - c_2 d_2)\phi^4 + O(\phi^6) \ge 0. \end{aligned}$$

Hence, $P(\phi) = O(\phi^4)$ if and only if $c_2 = 0$; i.e. if and only if $p = 2$. When $p = 2$, $c_2 = 0$, and we immediately obtain the latter parts of the proposition. **Q.E.D.**

Remarks: We conclude that an $\overline{A}$–stable second–order method must satisfy inequality (3.21) strictly, and that $\sigma(-1) \ne 0$. $\sigma(-1)$ can be zero for a first–order $\overline{A}$–stable method, but the method cannot satisfy $P(\pi + \phi) = O(\phi^4)$. In terms of the stability domain S, $\overline{A}$–stable methods must contain all of the imaginary axis in the interior of S, apart, possibly, from 0 and ∞. If the method is second–order, however, it must also be stable at ∞. We note that (3.21) is equivalent to $\tilde{c}_4 \ge 0$ in the following alternative expansion, valid when $p = 2$,

$$\rho(e^z)/\sigma(e^z) = z + \tilde{c}_3 z^3 + \tilde{c}_4 z^4 + O(z^5).$$

Below, we give the proof of the sufficiency of $\overline{A}$–stability for strict dissipativity when $q = 2$, which has to be treated as a special case, employing more elementary arguments than are required when $q \ge 3$.

Lemma 3.6 *If an irreducible consistent two–step method $(\rho,\,\sigma)$ is strictly A–stable, it is strictly dissipative.*

Proof: Taking the normalization $\sigma(1) = 1$, and noting that $\mathbb{P}^2 = \mathrm{Span}\{(\xi + 1)^2,\,(\xi - 1)^2,\,(\xi^2 - 1)\}$, a two–step consistent method must be of the form

$$\rho(\xi) = 1/2((\xi^2 - 1) + \gamma(\xi - 1)^2), \quad \sigma(\xi) = 1/4((\xi + 1)^2 + \delta(\xi^2 - 1) + \epsilon(\xi - 1)^2), \tag{3.22}$$

for $\gamma,\,\delta,\,\epsilon \in \mathbb{R}$. The method is irreducible if $|\epsilon| + |\gamma(\delta - \gamma)| \ne 0$. Considering the A–stability of the method $(\rho,\,\sigma)$, it is clear that unless $\gamma,\,\epsilon \ge 0$, at least one of $\rho(\xi)$ and $\sigma(\xi)$ has a root $|\xi| > 1$. The identity

$$\begin{aligned} & 4(\rho(\xi)\sigma(\omega) + \rho(\omega)\sigma(\xi)) \\ = \;& (\xi\omega - 1)\{[(1 + \gamma)\xi + (1 - \gamma)][(1 + \gamma)\omega + (1 - \gamma)] \\ + \;& (\epsilon + \gamma(\delta - \gamma))(\xi - 1)(\omega - 1)\} \\ + \;& \epsilon\gamma(\xi - 1)^2(\omega - 1)^2 + (\delta - \gamma)(\xi^2 - 1)(\omega^2 - 1) \end{aligned} \tag{3.23}$$

follows from manipulations that are made more transparent by Theorem 3.7. Putting $\xi = \overline{\omega} = e^{i\theta}$, we see that $P(\theta) = \epsilon\gamma(1-\cos\theta)^2 + (\delta-\gamma)\sin^2\theta$. Hence, a further requirement for A–stability is that $\delta - \gamma \geq 0$. (Note that $\delta = \gamma$ implies that the method is second–order.) As a result, we conclude that, for the method to be irreducible and A–stable, it is necessary that

$$0 \leq \gamma,\ \delta-\gamma,\ \epsilon, \quad 0 < \max\{\epsilon,\ \gamma(\delta-\gamma)\}. \tag{3.24}$$

We now show that (3.24) is also sufficient for G–stability, and hence, by Theorem 2.2, for A–stability. The term in square brackets on the right–hand side of (3.23) is of the form $4\sum_{i,j=1}^{2} g_{ij}\xi^{i-1}\omega^{j-1}$ for

$$g_{11} = (1+\epsilon+\gamma\delta+2\gamma)/4,\ \ g_{22} = (1+\epsilon+\gamma\delta-2\gamma)/4,\ \ g_{12} = g_{21} = (1-\epsilon-\gamma\delta)/4.$$

Given (3.24), $g_{11} > 0$ and $\mathrm{Det}(G) = (\epsilon+\gamma(\delta-\gamma))/4 > 0$, and hence G is positive definite. The last line of (3.23) is of the form $4\sum_{i,j=1}^{3} s_{ij}\xi^{i-1}\omega^{j-1}$, where

$$S = 1/4[\epsilon\gamma\mathbf{a}_1\mathbf{a}_1^T + (\delta-\gamma)\mathbf{a}_2\mathbf{a}_2^T], \quad \text{for} \quad \mathbf{a}_1 = [1,\ -2,\ 1]^T, \quad \mathbf{a}_2 = [1,\ 0,\ -1],$$

and is thus non–negative definite. Hence, $\|\cdot\|_G$ is a G–norm for $(\rho,\ \sigma)$.

$\overline{A}$–stability requires strict stability; i.e. $\gamma > 0$. Below, we will use Theorem 3.2 to show that $\|\cdot\|_G$ is a strict G–norm for the method $(\rho,\ \sigma)$, when $\gamma > 0$. This will establish, by Lemma 3.4, that (3.24) plus $\gamma > 0$ is a necessary and sufficient condition for $\overline{A}$–stability, and, thus, will prove the lemma (and also that every consistent irreducible two–step strictly stable A–stable method is strictly A–stable). $\gamma > 0$ implies $\rho(-1) \neq 0$, and thus $a_1(\xi),\ a_2(\xi) \notin \mathrm{Span}\{\rho(\xi),\ \sigma(\xi)\}$. On the other hand, at least one of the coefficients $\epsilon\gamma$, $\delta-\gamma$, is positive, and so the method satisfies (3.10), and is thus strictly dissipative. **Q.E.D.**

Remark: As noted above, the two–step case is in some sense special, and doesn't bring out the full force of the definition of $\overline{A}$–stability, which imposes rather more restrictions on A–stable methods, when $q \geq 3$, than is suggested by the above proof. The following example shows that an A–stable method may fail to be $\overline{A}$–stable because $P(\phi) = O(\phi^6)$. Consider

$$\rho(\xi) = \xi^3 - \xi^2, \quad \sigma(\xi) = 2\xi^3 - 3\xi^2 + 5\xi/2 - 1/2. \tag{3.25}$$

This second–order method satisfies

$$8\Re e[\rho(\xi)\overline{\sigma(\xi)}] = (|\xi|^2-1)[4|\xi|^4 + |3\xi^2-4\xi+1|^2 + |\xi-1|^4] + 2|\xi-1|^6.$$

Hence, $P(\phi) = 2(1-\cos\phi)^3 = \phi^6/4 + O(\phi^8)$.

To complete the proof of Theorem 3.3 for $q \geq 3$, we have to construct, or otherwise establish the existence of, a pair $(G,\ S)$ for any $\overline{A}$–stable method $(\rho,\ \sigma)$, such that S satisfies the requirements of Theorem 3.2. Our construction is based around the original G–norm construction of Dahlquist in [7], together with the observation that, if a degenerate A–stable method satisfies $P(\theta) > 0$ for all $\theta \in \mathbb{R}$, one can find a pair $(G,\ S)$ where S is of full rank. We start by quoting the following lemma.

Lemma 3.7 (Dahlquist [7]) *Suppose that* (ρ_q, σ_q) *is an irreducible (possibly inconsistent) A–stable q–step method. Then, if* $\sigma_q(-1) = 0$*, there exist a constant* $\gamma_q = -\rho_q(-1)/(2\sigma_q'(-1)) > 0$ *and an irreducible A–stable method* $(\rho_{q-1}, \sigma_{q-1})$ *of exactly* $(q-1)$ *steps such that*

$$\rho_{q-1}(\xi) = \frac{\rho_q(\xi) - \gamma_q \frac{\xi-1}{\xi+1}\sigma_q(\xi)}{\xi+1}, \qquad \sigma_{q-1}(\xi) = \frac{\sigma_q(\xi)}{\xi+1}. \tag{3.26}$$

If, instead, $\sigma_q(\xi)$ *has no roots on the unit circle* $S(0, 1)$*, but* $\rho_q(1) = 0$*, then there is a constant* $\gamma_q = \sigma_q(1)/(2\rho_q'(1)) > 0$*, and an A–stable method* $(\rho_{q-1}, \sigma_{q-1})$ *of exactly* $(q-1)$ *steps, such that*

$$\rho_{q-1}(\xi) = \frac{\sigma_q(\xi) - \gamma_q \frac{\xi+1}{\xi-1}\rho_q(\xi)}{\xi-1}, \qquad \sigma_{q-1}(\xi) = \frac{\rho_q(\xi)}{\xi-1}. \tag{3.27}$$

We remark that Dahlquist actually proved a more general result than the above, in the context of a pair of polynomials (ρ_q, σ_q) with complex coefficients satisfying (2.8) — see [7]. The first step in our proof of the sufficiency of $\overline{A}$–stability for strict dissipativity establishes that, after at most two applications of Lemma 3.6, one arrives at a method for which $P(\theta) > 0$ for all $\theta \in \mathbb{R}$.

Lemma 3.8 *Suppose that, for* $q \geq 3$*,* (ρ_q, σ_q) *is an irreducible strictly A–stable method of order* $p \geq 1$*. Then, the method falls into one of the following catagories.*
(i) $p = 1$ *and* $\sigma(-1) \neq 0$*.*
(ii) $p = 1$ *and* $\sigma(-1) = 0$*.*
(iii) $p = 2$*.*
In each case, there is an $\overline{A}$*–stable method* $(\rho_{r-1}, \sigma_{r-1})$ *of* $r-1$ *steps, such that* $P_{r-1}(\theta) = 2\Re e[\rho_{r-1}(e^{i\theta})\sigma_{r-1}(e^{-i\theta})] > 0$ *for all* $\theta \in \mathbb{R}$*. In case* **(i)**, $r = q$ *and there is a method* $(\rho_{q-1}, \sigma_{q-1})$ *such that*

$$\begin{aligned} \rho_q(\xi)\sigma_q(\omega) + \rho_q(\omega)\sigma_q(\xi) &= (\xi\omega - 1)\sigma_{q-1}(\xi)\sigma_{q-1}(\omega) \\ &+ (\xi-1)(\omega-1)[\rho_{q-1}(\xi)\sigma_{q-1}(\omega) + \rho_{q-1}(\omega)\sigma_{q-1}(\xi)]. \end{aligned} \tag{3.28}$$

In case **(ii)**, $r = q-1$ *and there exist a method* $(\rho_{q-1}, \sigma_{q-1})$ *and a constant* $\gamma_{q-1} > 0$ *such that*

$$\begin{aligned} & \rho_q(\xi)\sigma_q(\omega) + \rho_q(\omega)\sigma_q(\xi) \\ = \; & (\xi\omega - 1)[\sigma_{q-1}(\xi)\sigma_{q-1}(\omega) + 2\gamma_{q-1}(\xi+1)(\omega+1)\sigma_{q-2}(\xi)\sigma_{q-2}(\omega)] \\ + \; & (\omega^2 - 1)(\xi^2 - 1)[\rho_{q-2}(\xi)\sigma_{q-2}(\omega) + \rho_{q-2}(\omega)\sigma_{q-2}(\xi)]. \end{aligned} \tag{3.29}$$

In case **(iii)**, $r = q-1$ *and there exist a method* $(\rho_{q-1}, \sigma_{q-1})$ *and a constant* $\gamma_{q-1} > 0$ *such that*

$$\begin{aligned} & \rho_q(\xi)\sigma_q(\omega) + \rho_q(\omega)\sigma_q(\xi) \\ = \; & (\xi\omega - 1)[\sigma_{q-1}(\xi)\sigma_{q-1}(\omega) + 2\gamma_{q-1}(\xi-1)(\omega-1)\sigma_{q-2}(\xi)\sigma_{q-2}(\omega)] \\ + \; & (\omega - 1)^2(\xi - 1)^2[\rho_{q-2}(\xi)\sigma_{q-2}(\omega) + \rho_{q-2}(\omega)\sigma_{q-2}(\xi)]. \end{aligned} \tag{3.30}$$

Proof: Definition 3.2 and Theorem 2.1 imply that consistent $\overline{A}$–stable methods fall into the three catagories mentioned. In case **(i)**, (3.28) and the A–stability of $(\rho_{q-1},\,\sigma_{q-1})$ follow directly from Lemma 3.7 and formulae (3.27). Substituting $\xi = e^{i\theta}$, $\omega = e^{-i\theta}$ into (3.28), we obtain

$$P_q(\theta) = 2\Re e[\rho_q(e^{i\theta})\sigma_q(e^{-i\theta})] = 2|1-e^{i\theta}|^2\Re e[\rho_{q-1}(e^{i\theta})\sigma_{q-1}(e^{-i\theta})] = |1-e^{i\theta}|^2 P_{q-1}(\theta).$$

We know that $P_q(\theta) > 0$ for all $\theta \neq 2n\pi$. On the other hand, since $p = 1$, Proposition 3.4 implies that $P_q(\phi) \neq O(\phi^4)$. Hence, $P_{q-1}(\phi) \neq O(\phi^2)$, which implies that $P_{q-1}(0) \neq 0$ and, thus, that $P_{q-1}(\theta) > 0$ for all $\theta \in \mathbb{R}$. The $\overline{A}$–stability of $(\rho_{q-1},\,\sigma_{q-1})$ is now clear.

In case **(ii)**, $\sigma(-1) = 0$, and so we may apply formula (3.26) of Lemma 3.7 to obtain a new irreducible A–stable method $(\rho_{q-1},\,\sigma_{q-1})$. Clearly, $\rho_{q-1}(1) = \rho_q(1) = 0$, and we may therefore apply (3.27) to obtain a third irreducible A–stable method $(\rho_{q-2},\,\sigma_{q-2})$. (3.29) follows directly from these formulae. Substituting $\xi = e^{i\theta}$, $\omega = e^{-i\theta}$ into (3.29), we obtain

$$P_q(\theta) = |e^{i\theta}+1|^2 P_{q-1}(\theta) = |1-e^{i\theta}|^2|1+e^{i\theta}|^2 P_{q-2}(\theta). \tag{3.31}$$

Here, $O(\phi^2) = P_q(\pi+\phi)$, $P_q(\phi) \neq O(\phi^4)$, and so $P_{q-2}(\pi)$, $P_{q-2}(0) \neq 0$. Thus, $P_{q-2}(\theta) > 0$ for all $\theta \in \mathbb{R}$. Hence, $(\rho_{q-2},\,\sigma_{q-2})$ is $\overline{A}$–stable.

For case **(iii)**, Proposition 3.4 implies that $P_q(\phi) = O(\phi^4)$, and therefore, by (3.19), $P_q(\pi) > 0$. After a first application of (3.27), we obtain the irreducible A–stable method $(\rho_{q-1},\,\sigma_{q-1})$. Since $P_{q-1}(\phi) = O(\phi^2)$ and $\sigma_{q-1}(1) = \rho_q'(1) \neq 0$, we deduce that $\rho_{q-1}(1) = 0$. Thus, we may again apply (3.27) to the method $(\rho_{q-1},\,\sigma_{q-1})$ to obtain the irreducible A–stable method $(\rho_{q-2},\,\sigma_{q-2})$. (3.30) follows from a rearrangement of the formulae used. Substituting $\xi = e^{i\theta}$ and $\omega = e^{-i\theta}$ into (3.30) yields $P_q(\theta) = |1-e^{i\theta}|^4 P_{q-2}(\theta)$. Since $P_q(\theta) > 0$ for $\theta \neq 2n\pi$ and $P_q(\phi) \neq O(\phi^6)$, we deduce that $P_{q-2}(\theta) > 0$ for all $\theta \in \mathbb{R}$. **Q.E.D.**

The last result showed how an $\overline{A}$–stable method $(\rho_q,\,\sigma_q)$ can be reduced to a method $(\rho_{r-1},\,\sigma_{r-1})$ such that $P_{r-1}(\theta) > 0$ for all $\theta \in \mathbb{R}$. The next lemma shows that, for such a method, a G–norm exists for which the corresponding matrix S has full rank. This finally leads on to the completion of Theorem 3.3, in which we construct a strict G–norm for $(\rho_q,\,\sigma_q)$, where S satisfies the sufficient conditions of Theorem 3.2.

Lemma 3.9 *Suppose that $r \geq 2$, and that $(\rho,\,\sigma)$ is an $(r-1)$–step irreducible A–stable method, such that $P(\theta) > 0$ for all $\theta \in \mathbb{R}$. Then, there is a real symmetric positive definite $(r-1)\times(r-1)$ matrix H, a real basis $\{b_1(\xi),\,\ldots,\,b_r(\xi)\}$ for $\mathbb{P}^{r-1}$, positive constants $\mu_1,\,\ldots,\,\mu_r$ and a constant $\lambda \in (0,\,1)$ such that, for any $N,\,n \in \mathbb{N}$ and $U_n,\,\ldots,\,U_{n+r-1} \in \mathbb{C}^N$,*

$$\begin{aligned} 2\Re e\langle\rho(E)U_n,\,\sigma(E)U_n\rangle \;&=\; \|\mathbf{U}_{n+1}\|_H^2 - \|\mathbf{U}_n\|_H^2 + \sum_{m=1}^{r}|\mu_m b_m(E)U_n|^2 \\ &\geq\; \|\mathbf{U}_n\|_H^2 - \lambda^2\|\mathbf{U}_n\|_H^2. \end{aligned} \tag{3.32}$$

Proof: $P(\theta) > 0$ for $\theta \in \mathbb{R}$, implies that $\Re e[\rho(\xi)\overline{\sigma(\xi)}] > 0$ for all $|\xi| \geq 1$. Continuity implies that there is a $\lambda \in (0, 1)$ such that

$$\Re e[\rho(\xi)\overline{\sigma(\xi)}] > 0, \qquad \text{for } |\xi| \geq \lambda. \tag{3.33}$$

Thus, $\Re e[\rho(\xi)/\sigma(\xi)]$ is well–defined and continuous on $\partial B(0, \lambda)$, and achieves a minimum $\delta > 0$ on this compact set. Let us consider the method

$$\rho^*(\xi) \equiv \rho(\lambda\xi) - \delta\sigma(\lambda\xi), \qquad \sigma^*(\xi) \equiv \sigma(\lambda\xi).$$

By (3.33), and the harmonic minimum principle, $\Re e[\rho^*(\xi)\sigma^*(\overline{\xi})] > 0$ for $|\xi| > 1$. Hence, Theorem 2.1 implies that the method (ρ^*, σ^*) is A–stable, and Theorem 2.2 shows the existence of an $(r-1)\times(r-1)$ symmetric positive definite matrix D and an $r \times r$ symmetric non–negative definite matrix T, such that

$$\rho^*(\zeta)\sigma^*(\eta) + \rho^*(\eta)\sigma^*(\zeta) = (\zeta\eta - 1)\sum_{i,\,j=1}^{r-1} d_{ij}\zeta^{i-1}\eta^{j-1} + \sum_{i,\,j=1}^{r} t_{ij}\zeta^{i-1}\eta^{j-1}.$$

Setting $\xi = \zeta/\lambda$ and $\omega = \eta/\lambda$, we obtain

$$\begin{aligned}\rho(\xi)\sigma(\omega) + \rho(\omega)\sigma(\xi) &= (\xi\omega - \lambda^2)\sum_{i,\,j=1}^{r-1} d_{ij}\lambda^{-i-j}\xi^{i-1}\omega^{j-1} \\ &+ 2\delta\sigma(\xi)\sigma(\omega) + \sum_{i,\,j=1}^{r} t_{ij}\lambda^{-i-j+2}\xi^{i-1}\omega^{j-1}.\end{aligned}$$

Now, let us define $H = \Lambda_1^T D\Lambda_1$, positive definite, and $W = \Lambda_2^T T\Lambda_2$, non–negative definite, where $\Lambda_1 = \mathrm{diag}[\lambda^{-1}, \ldots, \lambda^{-(r-1)}]$ and $\Lambda_2 = \mathrm{diag}[1, \lambda^{-1}, \ldots, \lambda^{-(r-1)}]$. Then, from the bijection $\xi^i\omega^j \longleftrightarrow \langle U_{n+i}, U_{n+j}\rangle$, we obtain

$$\begin{aligned}2\Re e\langle\rho(E)U_n, \sigma(E)U_n\rangle &= \|\mathbf{U}_{n+1}\|_H^2 - \lambda^2\|\mathbf{U}_n\|_H^2 + 2\delta|\sigma(E)U_n|^2 \\ &+ \sum_{i,\,j=1}^{r} w_{ij}\langle U_{n+i}, U_{n+j}\rangle.\end{aligned}$$

For $\boldsymbol{\beta} = [\beta_0, \ldots, \beta_{r-1}]^T$, we consider the matrix

$$\Sigma = (1-\lambda^2)\left[\begin{array}{ccc|c} & & & 0 \\ & H & & \vdots \\ & & & 0 \\ \hline 0 & \ldots & 0 & 0 \end{array}\right] + 2\delta\boldsymbol{\beta}\boldsymbol{\beta}^T + W.$$

If $\mathbf{x} \in \mathbb{R}^r \setminus \{\mathbf{0}\}$, then $\mathbf{x} = [\mathbf{y}^T, x_r]^T$ for $\mathbf{y} \in \mathbb{R}^{r-1}$. Because $\beta_q \neq 0$, $\mathbf{x}^T \Sigma \mathbf{x} \geq (1-\lambda^2)\mathbf{y}^T H \mathbf{y} + \beta_{r-1}^2 x_r^2 > 0$, by A–stability and Theorem 2.1, and therefore Σ is real symmetric positive definite.

Hence, the pair (H, Σ) satisfy (2.12) for the method (ρ, σ). Furthermore, the positive definiteness of Σ implies there is a basis of orthogonal eigenvectors $\mathbf{b}_1, \ldots, \mathbf{b}_r$ of S, with corresponding eigenvalues $\mu_1^2, \ldots, \mu_r^2 > 0$, such that the polynomials $b_m(\xi) = \sum_{i=1}^r b_m \xi^{i-1}$ satisfy (3.32). **Q.E.D.**

Note: the above method of proof recombines several ideas of Dahlquist [7], but the consideration of $\lambda < 1$ is novel.

Because H in the statement of Lemma 3.9 is positive definite, there is an orthogonal basis of eigenvectors $\{\mathbf{c}_1, \ldots, \mathbf{c}_{r-1}\}$ for $\mathbb{R}^{r-1}$, whose corresponding eigenvalues $\nu_1^2, \ldots, \nu_{r-1}^2$ are all positive. If we define $c_m(\xi) = \sum_{i=1}^{r-1} c_{mi}\xi^{i-1}$, then (3.32) implies that

$$\rho(\xi)\sigma(\omega) + \rho(\omega)\sigma(\xi) = (\xi\omega - 1)\sum_{m=1}^{r-1} \nu_m^2 c_m(\xi)c_m(\omega) + \sum_{m=1}^{r} \mu_m^2 b_m(\xi) b_m(\omega). \quad (3.34)$$

Lemma 3.10 *A consistent irreducible strictly A–stable method (ρ, σ) of more than two steps is strictly dissipative.*

Proof: We again consider separately the categories **(i)**–**(iii)** from Lemma 3.8, and show that, in each case, there is a pair (G, S), such that the sufficient conditions of Theorem 3.2 are satisfied by a particular G–norm.

In case **(i)**, for $r = q \geq 3$, combining (3.28) with (3.34) yields

$$\begin{aligned} & \rho_q(\xi)\sigma_q(\omega) + \rho_q(\xi)\sigma_q(\omega) \\ = \quad & (\xi\omega - 1)\left\{\sigma_{q-1}(\xi)\sigma_{q-1}(\omega) + \sum_{m=1}^{q-1}\nu_m^2(\xi-1)c_m(\xi)(\omega-1)c_m(\omega)\right\} \\ + \quad & \sum_{m=1}^{q}\mu_m^2(\xi-1)b_m(\xi)(\omega-1)b_m(\omega) \\ \equiv \quad & (\xi\omega-1)\sum_{i,j=1}^{q} g_{ij}\xi^{i-1}\omega^j + \sum_{i,j=1}^{q+1} s_{ij}\xi^{i-1}\omega^{j-1}. \end{aligned} \quad (3.35)$$

Since, (from the proof of Lemma 3.8), $\sigma_{q-1}(1) = \rho_q'(1) \neq 0$, $\sigma_{q-1}(\xi) \notin \oplus_{m=1}^{q-1}(1-\xi)c_m(\xi)$, and so G is symmetric positive definite. $\sigma_q(1) \neq 0$ implies $\sigma_q(\xi) \notin \oplus_{m=1}^{q}(1-\xi)b_m(\xi)$. Thus, S satisfies (3.10) and (ρ_q, σ_q) is strictly dissipative.

In case **(ii)**, we combine (3.29) with (3.34), for $r = q-1 \geq 2$, to give

$$\begin{aligned} & \rho_q(\xi)\sigma_q(\omega) + \rho_q(\omega)\sigma_q(\xi) \\ = \quad & (\xi\omega - 1)\Big\{\sigma_{q-1}(\xi)\sigma_{q-1}(\omega) + 2\gamma_{q-1}(\xi+1)\sigma_{q-2}(\xi)(\omega+1)\sigma_{q-2}(\omega) \end{aligned}$$

$$
\begin{aligned}
&+ \left.\sum_{m=1}^{q-2} \nu_m^2 (\xi^2-1) c_m(\xi)(\omega^2-1)c_m(\omega)\right\} + \sum_{m=1}^{q-1} \mu_m^2 (\xi^2-1) b_m(\xi)(\omega^2-1) b_m(\omega) \\
&\equiv (\xi\omega - 1) \sum_{i,j=1}^{q} g_{ij} \xi^{i-1}\omega^{j-1} + \sum_{i,j=1}^{q+1} s_{ij}\xi^{i-1}\omega^{j-1}. \qquad (3.36)
\end{aligned}
$$

From Lemma 3.8, $\sigma_{q-1}(-1)$, $\sigma_{q-2}(1) \neq 0$. Hence, $\sigma_{q-1}(\xi)$, $(\xi+1)\sigma_{q-2}(\xi) \notin \oplus_{m=1}^{q-2}(\xi^2-1)c_m(\xi)$, and so G is symmetric positive definite. On the other hand, S has rank $r = q-1$, and $\sigma_q(1)$, $\rho_q(-1) \neq 0$ imply that $\sigma_q(\xi)$, $\rho_q(\xi) \notin \oplus_{m=1}^{q-1}(\xi^2-1)b_m(\xi)$. Thus, S satisfies (3.10), and the method (ρ_q, σ_q) is strictly dissipative.

In case **(iii)**, (3.30) and (3.34), with $r = q-1 \geq 2$, combine to give

$$
\begin{aligned}
&\rho_q(\xi)\sigma(\omega) + \rho_q(\omega)\sigma_q(\xi) \\
&= (\xi\omega-1)\left\{\sigma_{q-1}(\xi)\sigma_{q-1}(\omega) + 2\gamma_{q-1}(\xi-1)\sigma_{q-2}(\xi)(\omega-1)\sigma_{q-2}(\omega)\right. \\
&+ \left.\sum_{m=1}^{q-2} \nu_m^2 (\xi-1)^2 c_m(\xi)(\omega-1)^2 c_m(\omega)\right\} + \sum_{m=1}^{q-1} \mu_m^2 (\xi-1)^2 b_m(\xi)(\omega-1)^2 b_m(\omega) \\
&\equiv (\xi\omega - 1) \sum_{i,j=1}^{q} g_{ij} \xi^{i-1}\omega^{j-1} + \sum_{i,j=1}^{q+1} s_{ij}\xi^{i-1}\omega^{j-1}. \qquad (3.37)
\end{aligned}
$$

Lemma 3.8 implies that $\sigma_{q-1}(1)$, $\sigma_{q-2}(1) \neq 0$. Hence, $\sigma_{q-1}(\xi)$, $(\xi-1)\sigma_{q-2}(\xi) \notin \oplus_{m=1}^{q-2}(\xi-1)^2 c_m(\xi)$, and so G is symmetric positive definite. S again has rank $r = q-1$; $\sigma_q(1)$, $\rho_q'(1) \neq 0$ imply that $\sigma_q(\xi)$, $\rho_q(\xi) \notin \oplus_{m=1}^{q-1}(\xi-1)^2 b_m(\xi)$. Thus, S satisfies (3.10), and the method (ρ_q, σ_q) is strictly dissipative. **Q.E.D.**

Proof of Theorem 3.3: The necessity of a strictly dissipative method to be strictly A–stable for $q \geq 2$, or A–stable when $q = 1$ was established by by Lemma 3.4. Sufficiency follows from Theorem 3.2 ($q = 1$), Lemma 3.6 ($q = 2$), and Lemma 3.10 ($q \geq 3$). **Q.E.D.**

Remark: The results proved in this section may be extended to the case of variable step sizes $k \in (0, k_0]$, as follows. Because the norm $\|\cdot\|_G$ is invariant under k, $(\|\mathbf{U}_n\|_G)$ is still a strictly decreasing sequence until $\|\mathbf{U}_{n+1}\|_G \leq R_1(k_0)$. (As may be seen from the proof of Theorem 3.2, $R_1(k)$ may be chosen to be an increasing linear function of k.) Once inside the ball $\overline{B}(0, R_1(k_0))$, (3.2) implies that the sequence $(\mathbf{U}_n)$ cannot leave.

4 Infinite–dimensional dissipativity

There is more than one possible framework in which one may generalize to an infinite number of dimensions the problem considered in Section 2. We choose to work in a setting that includes (1.1, 1.2) as a special case, and in which one may analyse the incompressible Navier–Stokes equations in two space dimensions.

Let H be a complex Hilbert space with inner product $\langle\cdot,\,\cdot\rangle$ and norm $|\cdot|$, and let W be a dense continuously imbedded Hilbert subspace of H. Suppose that $f : W \longrightarrow H$ is a locally Lipschitz continuous function such that

$$\Re e\langle f(w),\, w\rangle \le a - b|w|^2, \qquad \text{for some } a \ge 0,\, b > 0, \qquad \text{for all } w \in W. \tag{4.1}$$

(Note that the inner product and norm here are those of H, not W.) We assume that, for all $u_0 \in H$, there is a unique solution of the equation

$$u_t = f(u), \quad t > 0, \qquad u(0) = u_0, \tag{4.2}$$

such that $u \in C[[0,\,\infty);\,H] \cap C[(0,\,\infty);\,W]$.

We again approximate, (in time only), using linear multistep and one–leg discretizations. As for the differential equation itself, our framework is too general for there to be sufficient information to prove existence of solutions, and we will consider the following global existence assumption in the one–leg case for some fixed f and k:

(A) There exist solution sequences $(\mathbf{U}_n) \subset H^q$ for (1.6, 4.1), for all $\mathbf{U}_0 \in H^q$, such that $\sigma(E)U_n \in W$ for all $n \in \mathbb{N}$.

Note that, in general, **(A)** makes no sense unless $\beta_q \ne 0$.

We employ the following definition of dissipativity in the one–leg case.

Definition 4.1 *A one–leg method (1.6) is* **dissipative**, *under the structural condition (4.1), if, for all pairs H and W, all locally Lipschitz $f : W \to H$ obeying (4.1), and all $k > 0$ for which* **(A)** *is satisfied, there is a bounded set $\mathcal{B} \subset H^q$ such that, for all bounded subsets D of H^q, there is an $n_0(D)$, such that $\mathbf{U}_n \in \mathcal{B}$ for all $n \ge n_0$, whenever $\mathbf{U}_0 \in D$.*

We remark that initial data for linear multistep methods cannot be taken from $(H \setminus W)^q$, unless $\sigma(\xi) = \beta_q \xi^q$. (In the case of consistent methods, the latter condition forces the multistep method to coincide with the one–leg method.) If assumption **(A)** is satisfied by (1.6, 4.1), then we may take arbitrary initial data $\mathbf{V}_0 \in W^q \cap D$ for (1.5, 4.1), where D is a bounded subset of H. An initial solution $U_0,\,\dots,\,U_{2q-1} \in H$ may be obtained for (1.6, 4.1) by solving

$$E^i\sigma(E)U_0 = V_i, \qquad E^i\rho(E)U_0 = kf(V_i), \quad i = 0,\,\dots,\,q-1, \tag{4.3}$$

as in Lemma 2.6. The existence of a solution sequence (U_n) for (1.6, 4.1) follows from **(A)**, and then, as in Theorem 2.5, the sequence $V_n = \sigma(E)U_n$, $n \ge 0$, satisfies (1.5, 4.1). This establishes the existence of solutions for (1.5, 4.1).

$\mathbf{U}_0$ is contained in some ball $\overline{B}(0,\,r) \subset H^q$, so if (1.6, 4.1) is dissipative, with absorbing set $\overline{B}(0,\,R) \subset H^q$, then $\mathbf{V}_n \in \overline{B}(0,\,|\sigma(1)|R)$ for $n \ge n_0(r)$. In general, however, the set D does not determine r and $n_0(r)$, which depend on the arbitrarily large norm $\|\mathbf{V}_0\|_{W^q}$. Thus, in general, linear multistep methods only satisfy a weakened concept of dissipativity.

Henceforth, therefore, we shall only consider one–leg discretizations, in the knowledge that only a weak equivalence holds with linear multistep methods in this context. We combine two main objectives in this section:
(i) To classify dissipative methods for condition (4.1).
(ii) To find conditions under which a quantitative analysis analogous to (1.3, 1.4) is possible, which yields bounds on the rate of decay $|U_n|$ for large initial data that are in good agreement with those given for $|u(t)|$ by (1.4).

We find it convenient to complete **(i)** before considering **(ii)**.

Lemma 4.1 *If $(\rho,\,\sigma)$ is irreducible, then the corresponding one–leg method (1.6) is dissipative under condition (4.1), only if $(\rho,\,\sigma)$ is strongly A–stable.*

Proof: For all $k>0$, (**A**) is satisfied when $f:\mathbb{C}\to\mathbb{C}$ is defined by $f(u)=\lambda u$, for $\Re e[\lambda]<0$. So, Lemma 3.4 shows that the method $(\rho,\,\sigma)$ must be A–stable.

Consider $H=L_2(0,\,\pi)$, $W=H^2(0,\,\pi)\cap H_0^1(0,\,\pi)$, and $f(u)\equiv L[u]\equiv u_{xx}$. The inequality $\langle u,\,u_{xx}\rangle=-|u_x|^2\leq -|u|^2$, for $u\in W$, shows that f satisfies (4.1). $(-N^2,\,(2/\pi)^{1/2}\sin Nx)_{N=1}^{\infty}$ are the normalized eigenpairs of L in H. Hence, if $(\rho,\,\sigma)$ is A–stable, and $\xi(\lambda)$ denotes a root of

$$\rho(\xi)-\lambda\sigma(\xi)=0, \tag{4.4}$$

$U_n=\xi(-kN^2)^n\sin Nx$, $n\geq 0$, is a solution sequence for (1.6, 4.1).

Suppose, now, that $\sigma(\xi_0)=0$ for some $|\xi_0|=1$, but that (1.6) is dissipative under condition (4.1). Fixing the step–size k, we may assume, without loss of generality, that the absorbing set is a ball $\overline{B}(0,\,r)\subset H^q$, for some $r>0$, where we use the infinity product norm. By definition of dissipativity, there is an $n_0\in\mathbb{N}$, such that $\mathbf{U}_n\in\overline{B}(0,\,r)$ for all $n\geq n_0$, whenever $\mathbf{U}_0\in\overline{B}(0,\,2r)$.

Let us choose $\epsilon>0$ such that $(1-\epsilon)^{n_0}>1/2$. Rouché's theorem shows that there is a $\lambda_0>0$ such that, when $|\lambda|\geq\lambda_0$, there is a root $\xi(\lambda)$ of (4.4) such that $|\xi(\lambda)-\xi_0|\leq\epsilon$, and thus $|\xi(\lambda)|\in[1-\epsilon,\,1)$. Taking N such that $kN^2\geq\lambda_0$, and considering the solution $U_n=2r(2/\pi)^{1/2}\Re e[e^{-i\phi}\xi(-kN^2)^n]\sin Nx$ of (1.6, 4.1), for $\phi=\arg(\xi(-kN^2)^{n_0})$, we observe that $\mathbf{U}_0\in\overline{B}(0,\,2r)$, but that

$$\|\mathbf{U}_{n_0}\|_{H^q}=\max_{1\leq i\leq q}|U_{n_0+i}|\geq|U_{n_0}|=2|\xi(-kN^2)|^{n_0}r>r.$$

This contradiction proves the lemma. **Q.E.D.**

We remark that one may show, using Fourier decomposition, that when $u_t=u_{xx}$ is semidiscretized in time by any A–stable method, all trajectories tend to 0 in H. What the above proof shows is that, when $\sigma(\xi)$ has a root on the unit circle, there is no uniformity in the rate of decay of the H norm. Thus, dissipativity fails. The linear counterexample of the above proof equally serves to show that strong A–stability is necessary for a linear multistep method (1.5) to be dissipative under condition (4.1).

The following lemma helps to show that strong A–stability is also a sufficient condition for the dissipativity of (1.6) under condition (4.1).

Lemma 4.2 *Suppose that the method* $(\rho,\,\sigma)$ *is irreducible and strongly A–stable. Then, there is a constant* $\nu_0 > 0$*, and a continuous strictly decreasing function* $\lambda : [0,\,\nu_0] \to (0,\,1]$*, such that*

$$\Re e\left[\frac{\rho(\xi)}{\sigma(\xi)} + \nu\right] \geq 0 \qquad \textit{for } |\xi| \geq \lambda(\nu), \quad \textit{for all } \nu \in [0,\,\nu_0]. \tag{4.5}$$

Proof: When $(\rho,\,\sigma)$ is strongly A–stable, uniform continuity in the neighbourhood of the compact set $\partial B(0,\,1)$ implies that there is a $\nu_0 > 0$, such that σ has no roots in $D \equiv \overline{B}(0,\,1) \setminus B(0,\,1-\nu_0)$. Hence, ρ/σ is continuously differentiable on compact D, and we may take positive $L \geq \sup_{\xi\in D} |(\rho/\sigma)'(\xi)|$. Thus, (4.5) is satisfied for $\lambda(\nu) \equiv 1 - \nu/L$. **Q.E.D.**

A further variation on the G–stability theme is needed for our main result.

Definition 4.2 *Suppose that there is a constant* $\nu_0 > 0$ *and a matrix function* $G \in C[[0,\,\nu_0];\,\mathcal{L}(\mathbb{R}^q)]$*, such that* $G(\nu)$ *is symmetric positive definite, and that* $\langle\cdot,\,\cdot\rangle$ *is an inner product on some complex Hilbert space* H*. Then a continuum of norms may be defined on* H^q*, for* $\nu \in [0,\,\nu_0]$*, by*

$$\|\mathbf{U}\|^2_{G(\nu)} = \sum_{i=1}^{q}\sum_{j=1}^{q} g_{ij}(\nu)\langle U_i,\,U_j\rangle, \qquad \textit{for } \mathbf{U} \equiv [U_1,\,\ldots,\,U_q]^T, \quad U_i \in H. \tag{4.6}$$

If, for a given method $(\rho,\,\sigma)$*, there is such a function* G *and a continuous strictly decreasing function* $\lambda : [0,\,\nu_0] \to (0,\,1]$ *such that, for all inner products on all complex Hilbert spaces* H *and any sequence* (U_n) *defined on* H*,*

$$2\Re e\langle(\rho(E) + \nu\sigma(E))U_n,\,\sigma(E)U_n\rangle \geq \|\mathbf{U}_{n+1}\|^2_{G(\nu)} - \lambda(\nu)^2\|\mathbf{U}_n\|^2_{G(\nu)}, \tag{4.7}$$

for all $n \geq 0$ *and all* $\nu \in [0,\,\nu_0]$*, then we say that the method* $(\rho,\,\sigma)$ *is* **strongly G–stable***, and that the function* $\|\cdot\|_{G(\nu)}$ *is a* **strong G–norm** *for the method.*

We remark that, by continuity, the matrices $G(\nu)$ are uniformly positive definite over $[0,\,\nu_0]$, and thus the norms $\|\cdot\|_{G(\nu)}$ are uniformly equivalent.

Theorem 4.3 *An irreducible method* $(\rho,\,\sigma)$ *is strongly G–stable if and only if it is strongly A–stable if and only if (1.6) is dissipative under condition (4.1).*

Proof: We assume that the method $(\rho,\,\sigma)$ is strongly G–stable, and we consider the one–leg approximation of equation (4.1, 4.2) for some f and k, for which assumption **(A)** is satisfied. Let $\nu = \min\{\nu_0,\,bk\}$. Taking the real part of the inner product of (1.6) with $\sigma(E)U_n$, we obtain, from (4.1) and (4.11),

$$\begin{aligned} 2ka \;&\geq\; 2\Re e\langle\rho(E)U_n + bk\sigma(E),\,\sigma(E)U_n\rangle \geq 2\Re e\langle\rho(E)U_n + \nu\sigma(E)U_n,\,\sigma(E)U_n\rangle \\ &\geq\; \|\mathbf{U}_{n+1}\|_{G(\nu)} - \lambda(\nu)^2\|\mathbf{U}_n\|^2_{G(\nu)}, \end{aligned}$$

where $\lambda(\nu) \in (0,\, 1)$. Summing the last inequality, we obtain

$$\|\mathbf{U}_n\|_{G(\nu)}^2 \leq \lambda(\nu)^{2n} \|\mathbf{U}_0\|_{G(\nu)}^2 + \frac{2ka(1-\lambda(\nu)^{2n})}{1-\lambda(\nu)^2}. \tag{4.8}$$

Hence, for any $\epsilon > 0$, $\mathcal{B} = B(0,\, r) \subset (H^q,\, \|\cdot\|_{G(\nu)})$ is an absorbing set for $r = 2ka(1-\lambda(\nu)^2)^{-1} + \epsilon$, and so (1.6) is dissipative under (4.1).

From Lemma 4.1, we know that dissipativity of the one–leg method (1.6), under condition (4.1), implies that $(\rho,\, \sigma)$ is strongly A–stable.

Now, we assume that the method $(\rho,\, \sigma)$ is strongly A–stable. Lemma 4.2 implies that there is a constant $\nu_0 > 0$, and a strictly decreasing continuous function $\lambda : [0,\, \nu_0] \to (0,\, 1]$, such that (4.5) holds. Defining $\rho^*(\xi) = \rho(\lambda(\nu)\xi) + \nu\sigma(\lambda(\nu)\xi)$ and $\sigma^*(\xi) = \sigma(\lambda(\nu)\xi)$, we obtain $\Re e\langle \rho^*(\xi),\, \sigma^*(\xi)\rangle > 0$ for $|\xi| > 1$, from (4.5) and the harmonic minimum principle. Since $\lambda \in C[[0,\, \nu_0];\, (0,\, 1]]$, the coefficients of $\rho^*(\xi)$ and $\sigma^*(\xi)$ are continuously dependent on ν.

Baiocchi and Crouzeix [1] have given a proof of Theorem 2.2, (see also [10]), showing that S may always be chosen to be of rank one at most. Thus, there exist a G–norm $\|\cdot\|_{H(\nu)}$ and a polynomial $a(\nu)(\xi) = \sum_{i=1}^{q+1} a_i(\nu)\xi^{i-1}$ for $(\rho^*,\, \sigma^*)$, such that

$$2\Re e\langle \rho^*(\xi),\, \sigma^*(\xi)\rangle = (\xi\omega - 1)\sum_{i,\, j=1}^{q} h_{ij}(\nu)\xi^{i-1}\omega^{j-1} + a(\nu)(\xi)a(\nu)(\omega).$$

On substituting $\omega = 1/\xi$, we see that the coefficients of $a(\nu)(\xi)$, and hence those of H, may be chosen to be continuously dependent on those of $\rho^*(\nu)$ and $\sigma^*(\nu)$. Following the proof of Lemma 3.9, we see that if $G(\nu) = \Lambda_1^T H \Lambda_1$, for $\Lambda_1 = \mathrm{diag}[\lambda(\nu)^{-1},\, \ldots,\, \lambda(\nu)^{-q}]$, then $G \in C[[0,\, \nu_0];\, \mathcal{L}(\mathbb{R}^q)]$ and $\|\cdot\|_{G(\nu)}$ is a strong G–norm for the method $(\rho,\, \sigma)$. **Q.E.D.**

We note that inequality (4.7) may also be used to show that strongly A–stable one–leg approximations of the monotone system (1.1, 1.8) have solutions which decay at the rate of $\lambda(\nu)^n$ for large initial data.

Objective **(ii)**, in this section, is to determine the asymptotic (in terms of size of initial data) rate of decay of one–leg approximations of (4.1, 4.2). Inequality (1.4) demonstrates that solutions of (4.1, 4.2) satisfy

$$\limsup_{|u_0|\to\infty} |u(t)|/|u_0| \leq e^{-bt}. \tag{4.9}$$

For a given irreducible strongly A–stable method $(\rho,\, \sigma)$, we aim to find uniform constants C, $\nu_0 > 0$ and a function $\lambda : [0,\, \nu_0] \to (0,\, 1]$, such that the bound

$$\limsup_{\|\mathbf{U}_0\|_{H^q}\to\infty} |U_n|/\|\mathbf{U}_0\|_{H^q} \leq C\lambda(bk)^n \tag{4.10}$$

is valid for the corresponding one–leg discretization of all equations (4.1, 4.2) such that $bk \in (0,\, \nu_0]$. For any strongly A–stable method, the proof of Lemma

4.2, combined with (4.8), provides one such λ. However, we are interested in identifying cases where a sharper value of λ can be found, so that $\lambda(bk)^n \approx e^{-bnk}$; i.e. the bound for numerical rate of decay emulates (4.9).

Let us define $g(r,\,\theta) \equiv \Re[\rho(re^{i\theta})/\sigma(re^{i\theta})]$ for $\theta \in [0,\,2\pi)$. If $g(1,\,\theta_0) = 0$, for some $\theta_0 \in [0,\,2\pi)$, the boundary point lemma for harmonic functions (see [9]) implies that $g_r(1,\,\theta_0) > 0$ — note that $g(1,\,0) = 0$ and $g_r(1,\,0) = \rho'(1)/\sigma(1) = 1$ for a consistent method. Just as for $P(\theta)$ in Section 3, $g(1,\,\theta)$ must have even multiplicity at θ_0, so that $g_\theta(1,\,\theta_0) = 0$. We deduce that

$$g_r(1,\,\theta_0) = \Re e[e^{i\theta_0}(\rho/\sigma)'(e^{i\theta_0})] > 0, \quad g_\theta(1,\,\theta_0) = -\mathcal{I}m[e^{i\theta_0}(\rho/\sigma)'(e^{i\theta_0})] = 0,$$

and hence, $g_r(1,\,\theta_0) = |(\rho/\sigma)'(e^{i\theta_0})|$. Clearly, $\lambda_\nu(0) \geq -1/g_r(1,\,\theta_0)$ for any such θ_0. For $\lambda(\nu)$ to approximate $e^{-\nu}$ well, we need λ to be determined solely by the behaviour of $\Re e[\rho(\xi)/\sigma(\xi)]$ near $\xi = 1$, where the order conditions of the method $(\rho,\,\sigma)$ come into play. This will be the case if we assume that

$$\Re e[\rho(\xi_0)/\sigma(\xi_0)] = 0 \quad \text{for } \xi_0 \in \partial B(0,\,1) \setminus \{1\}, \quad \text{implies } |(\rho/\sigma)'(\xi_0)| < 1. \tag{4.11}$$

Before our main results, we need the following technical lemma.

Lemma 4.4 *Let $(\rho,\,\sigma)$ be an irreducible consistent strongly A–stable method such that*

$$\rho(e^z)/\sigma(e^z) = c_1 z + c_2 z^2 + \ldots + c_{2q} z^{2q} + O(z^{2q+1}). \tag{4.12}$$

Then, for some $m \in [1,\,q]$, $c_{2s} = 0$ for all $s < m$, and $c_{2m} > 0$.

Proof: Assume that the result is false. Then, for some $m \in [1,\,q]$, $c_{2s} = 0$ for all $s < m$, and $c_{2m} \leq 0$. If the last inequality is strict, then $\Re e[\rho(\xi)/\sigma(\xi)] < 0$ for some $|\xi| = 1$, in contradiction to (2.18). Hence, $c_{2s} = 0$ for all $s \in [0,\,q]$. The method of proof of Lemma 3.4 shows that, for any pair $(G,\,S)$ satisfying (2.12), S has $q+1$ independent kernel vectors, and thus $S = 0$. From (2.12), if

$$h(r,\,\theta) \equiv \Re e[\rho(e^{r+i\theta})/\sigma(e^{r+i\theta})] \quad \text{for } (r,\,\theta) \in \mathbb{R} \times [0,\,2\pi), \tag{4.13}$$

then $h(0,\,\theta) = 0$ for all $\theta \in \mathbb{R}$. Thus, the boundary point lemma implies that $h_r(0,\,\theta) > 0$ for all $\theta \in \mathbb{R}$. By uniform continuity, there is a $\mu \in (0,\,1)$ such that $h(\log\mu,\,\theta) < 0$ for all $\theta \in \mathbb{R}$ — i.e. $|\rho(\xi)+\sigma(\xi)| < |\rho(\xi)|+|\sigma(\xi)|$ for all $|\xi| = \mu$ — and neither $\rho(\xi)$ nor $\sigma(\xi)$ has a root on $\partial B(0,\,\mu)$. From Rouché's theorem, we infer that $\rho(\xi)$ has no roots outside $\overline{B}(0,\,\mu)$, which is a contradiction. **Q.E.D.**

Theorem 4.5 *Suppose that $(\rho,\,\sigma)$ is an irreducible strongly A–stable method of order $p \geq 1$, satisfying (4.11). If the pair $(\lambda(\nu),\,\nu_0)$ satisfies (4.5), then $\lambda(\nu) > e^{-\nu}$ for all $\nu \in (0,\,\nu_0]$. Furthermore, there is a λ such that $\lambda(\nu) = e^{-\nu} + O(\nu^2)$ if and only if, either $p = 1$, or $p = 2$ and $c_4 > 0$ in (4.12). In addition, if $p = 2$, $c_{2s} = 0$ in (4.12) for all $s < m$, and $c_{2m} > 0$ for some $m \in [3,\,q]$, the best λ satisfies $\lambda(\nu) = e^{-\nu} + O(\nu^{m/(m-1)})$.*

Sketch Proof: Assumption (4.11) implies that the behaviour of $\lambda(\nu)$, for sufficiently small ν, is determined only by the behaviour of $\Re e[\rho(\xi)/\sigma(\xi)]$ in a neighbourhood of $\xi = 1$, though ν_0 may be reduced by behaviour elsewhere.

When $p = 1$, $c_1 = 0$ and $c_2 < 0$ in (4.12). Thus, from (4.13),

$$h(r, \theta) = r + c_2(r^2 - \theta^2) + \Re e[O(r + i\theta)^3]$$
$$h_r(0, 0) = 1, \quad h_{\theta\theta}(0, 0) = -2c_2 > 0, \quad \text{and } h_\theta(r, 0) = 0 \quad \text{for all } r \in \mathbb{R}.$$

So, for fixed small negative r_0, $h(r_0, 0)$ locally minimizes $h(r_0, \theta)$. Since $h(-\nu - c_2\nu^2, 0) + \nu = O(\nu^3)$, we conclude that $\lambda(\nu) \geq e^{-\nu} - c_2\nu^2 + O(\nu^3)$, but that there is a $\nu_0(\epsilon) > 0$ such that $\lambda(\nu) \equiv \exp(-\nu + (\epsilon - c_2)\nu^2)$ satisfies (4.5) for any $\epsilon > 0$.

When $p = 2$, $c_1 = c_2 = 0$ and $c_3 \leq -1/12$, by Theorem 2.1, and $c_{2s} = 0$ for all $s < m$, and $c_{2m} > 0$ for some $m \in [2, q]$, by Lemma 4.4. When $m = 2$, (which occurs whenever (ρ, σ) is also $\overline{A}$–stable, by Proposition 3.5),

$$h(r, \theta) = r - 3c_3 r\theta^2 + c_4\theta^4 + O(r^3 + r^2\theta^2 + r\theta^4 + \theta^6).$$

When $p = 2$ and $m \in [3, q]$,

$$h(r, \theta) = r - 3c_3 r\theta^2 + c_{2m}\theta^{2m} + O(r^3 + r^2\theta^{2(m-2)} + r\theta^{2(m-1)} + \theta^{2(m+1)}).$$

For fixed small negative r_0, $f(r_0, \theta)$ has local minimum

$$r_0 - \frac{m-1}{m}\left(\frac{(3c_3r_0)^m}{mc_{2m}}\right)^{\frac{1}{m-1}} + O(r_0^{\frac{m+1}{m-1}}), \quad \text{near } \theta = \pm\left(\frac{3c_3r_0}{mc_{2m}}\right)^{\frac{1}{2m-2}}. \tag{4.14}$$

Hence, the best λ is given by $\lambda(\nu) = \exp(-\nu + (C_m + \epsilon)\nu^{m/(m-1)})$, for $\epsilon > 0$, where $-C_m$ is the coefficient of $r_0^{m/(m-1)}$ in (4.14). $\square$

Remarks: It is to some degree counter–intuitive that the results for second–order methods should be no better, and in some cases worse, than those for first–order methods. However, it should be remembered that we have asked for very demanding conditions to be satisfied by a restricted class of methods.

We believe that it is only when one has behaviour similar to $f(u) = (-b + c(u)i)u$, for b as in (4.1) and $c(u) \in \mathbb{R}$ arbitrarily large, that the results of Theorem 4.5 are optimal. When f satisfies (4.1), but such behaviour is not present, we speculate that it should be possible to accommodate f in some extension of $A(\alpha)$–stable theory, and to use methods of high order which could achieve an asymptotic decay rate for the numerical method closer to that of the underlying system.

Acknowledgements:

The author wishes to thank Andrew Stuart for pointing out this open problem, Tony Humphries, for suggesting the work in Section 3, and Endre Süli for materially helpful discussions.

References

[1] C. Baiocchi and M. Crouzeix, *On the equivalence of A–stability and G–stability,* Appl. Numer. Math., 5 (1989), pp. 19–22.

[2] K. Burrage and J. C. Butcher, *Stability criteria for implicit Runge–Kutta processes,* SIAM J. Numer. Anal., 16 (1979), pp. 1095–1113.

[3] J. C. Butcher, *A stability property of implicit Runge–Kutta methods,* BIT, 15 (1975), pp. 358–361.

[4] J. B. Conway, *Functions of One Complex Variable*, 2nd Ed., Springer, Berlin, 1978.

[5] G. Dahlquist, *A special stability problem for linear multistep methods,* BIT, 3 (1963), pp. 27–43.

[6] G. Dahlquist, *Error analysis for a class of methods for stiff nonlinear initial value problems,* Numerical Analysis, Dundee 1975, Springer Lecture Notes in Mathematics 506, pp. 60–74.

[7] G. Dahlquist, *G–stability is equivalent to A–stability,* BIT, 18 (1978), pp. 384–401.

[8] T. Eirola and O. Nevanlinna, *What do multistep methods approximate?* Numer. Math., 53 (1988), pp. 559–569.

[9] D. Gilbarg and N. S. Trudinger, *Elliptic Partial Differential Equations of Second Order*, 2nd Ed., Springer, Berlin, 1983.

[10] E. Hairer and G. Wanner, *Solving Ordinary Differential Equations II*, Springer, Berlin, 1991.

[11] A. T. Hill and E. Süli, *Upper semicontinuity of attractors for linear multistep methods approximating sectorial evolution equations,* Math. Comp. (to appear).

[12] A. R. Humphries and A. M. Stuart, *Runge–Kutta methods for dissipative and gradient dynamical systems,* SIAM J. Numer. Anal. (to appear).

[13] J. K. Hale, *Asymptotic Behaviour of Dissipative Systems, A.M.S.*, New York, 1988.

[14] J. K. Hale, X.–B. Lin and G. Raugel, *Upper semicontinuity of attractors for approximations of semigroups and partial differential equations*, Math. Comp., 50 (1988), pp. 89–123.

[15] L. V. Kantorovich and G. P. Akilov, *Functional Analysis in Normed Spaces*, (English translation) Pergamon Press, Oxford, 1964.

[16] R. Temam, *Infinite–Dimensional Dynamical Systems in Mechanics and Physics*, Appl. Math. Sci. vol. 68, Springer, Berlin, 1988.