SIAM REVIEW
Vol. 36, No. 2, pp. 226–257, June 1994

004

MODEL PROBLEMS IN NUMERICAL STABILITY THEORY FOR INITIAL VALUE PROBLEMS*

A.M. STUART† AND A.R. HUMPHRIES‡

Abstract. In the past numerical stability theory for initial value problems in ordinary differential equations has been dominated by the study of problems with simple dynamics; this has been motivated by the need to study error propagation mechanisms in stiff problems, a question modeled effectively by contractive linear or nonlinear problems. While this has resulted in a coherent and self-contained body of knowledge, it has never been entirely clear to what extent this theory is relevant for problems exhibiting more complicated dynamics. Recently there have been a number of studies of numerical stability for wider classes of problems admitting more complicated dynamics. This on-going work is unified and, in particular, striking similarities between this new developing stability theory and the classical linear and nonlinear stability theories are emphasized.

The classical theories of A, B and algebraic stability for Runge–Kutta methods are briefly reviewed; the dynamics of solutions within the classes of equations to which these theories apply—linear decay and contractive problems—are studied. Four other categories of equations—gradient, dissipative, conservative and Hamiltonian systems—are considered. Relationships and differences between the possible dynamics in each category, which range from multiple competing equilibria to chaotic solutions, are highlighted. Runge–Kutta schemes that preserve the dynamical structure of the underlying problem are sought, and indications of a strong relationship between the developing stability theory for these new categories and the classical existing stability theory for the older problems are given. Algebraic stability, in particular, is seen to play a central role.

It should be emphasized that in all cases the class of methods for which a coherent and complete numerical stability theory exists, given a structural assumption on the initial value problem, is often considerably smaller than the class of methods found to be effective in practice. Nonetheless it is arguable that it is valuable to develop such stability theories to provide a firm theoretical framework in which to interpret existing methods and to formulate goals in the construction of new methods. Furthermore, there are indications that the theory of algebraic stability may sometimes be useful in the analysis of error control codes which are not stable in a fixed step implementation; this work is described.

Key words. numerical stability, Runge–Kutta methods, linear decay, contractivity, gradient systems, dissipativity, conservative systems, Hamiltonian systems

AMS subject classifications. 34C35, 34D05, 65L07, 65L20

1. Introduction.

Many problems of interest in the physical sciences and engineering require the understanding of dynamical features that evolve over long-time periods. Examples include the process of coarsening in solid phase separation, where metastability causes extremely long time-scales, turbulence in fluid mechanics, where statistical measures (such as Lyapunov exponents) require averages over long time intervals, and the simulation of planetary interactions in the solar system. Thus the numerical approximation of evolution equations over long time intervals is of some importance.

For simplicity we concentrate here on the system of ordinary differential equations

$$\frac{du}{dt} = f(u), \quad u(0) = u_0, \tag{1.1}$$

where $u(t) \in C^1(\mathbb{R}^+, \mathbb{C}^p)$ and $f(\bullet) : \mathbb{C}^p \to \mathbb{C}^p$. We will assume that f is, at least, continuously differentiable with respect to its arguments. Throughout the following we will denote

*Received by the editors November 9, 1992; accepted for publication (in revised form) January 18, 1994.

†Scientific Computing and Computational Mathematics Program, Division of Applied Mechanics, Durand-252, Stanford University, Stanford, California 94305-4040. The work of this author was supported by Office of Naval Research under N00014-92-J-1876 and National Science Foundation grant DMS-9201727.

‡School of Mathematical Sciences, University of Bath, Bath, Avon BA2 7AY, United Kingdom and Scientific Computing and Computational Mathematics Program, Department of Computer Science, Stanford University, Stanford, California 94305-2140. Present address, School of Mathematics, University of Bristol, University Walk, Bristol BS8 1TW, United Kingdom. This author's work was supported by the United Kingdom Science and Engineering Research Council, Stanford University, and Office of Naval Research grant N00014-92-J-1876.

the inner product on $\mathbb{C}^p$ by $\langle \bullet, \bullet \rangle$ with corresponding norm $\| \bullet \|$ denoted by $\|u\|^2 = \langle u, u \rangle$. The precise inner-product used will be that which appears in the structural assumptions made on f.

The large time dynamics of (1.1) can exhibit a variety of behavior ranging from very simple, such as reaching steady state, through moderately complex periodic or quasi-periodic behavior, to the extremely complex chaotic behavior observed in, for example, the Lorenz equations. A fundamental question in the numerical analysis of initial value problems is to determine how closely, and in what sense, the numerical approximation relates to the underlying continuous problem. If we let U_n denote an approximation to the true solution $u(t_n)$, where $t_n = n\Delta t$ and the time-step Δt is typically chosen to be small relative to an appropriate time-scale in the problem, then standard analysis on sufficiently smooth problems of the form (1.1) shows that the error satisfies

$$\|u(t_n) - U_n\| \leqslant c_1 e^{c_2 T} \Delta t^r, \tag{1.2}$$

for $0 \leqslant n\Delta t \leqslant T$. Here $r > 0$ is the order of the method and, typically, c_1 and c_2 are positive constants. Notice that, for fixed T, letting $\Delta t \to 0$ results in a proof of convergence of the numerical scheme on finite time intervals. However, fixing Δt and letting $T \to \infty$ gives no error bound; thus standard error analysis tells us nothing about the relationship between the long-time dynamics of the discrete and continuous problems. Understanding the behavior of algorithms for fixed Δt as $T \to \infty$ is what we shall term *numerical stability* for the purposes of this paper. In contrast to the question of convergence on fixed time intervals, it is necessary to impose structural assumptions on $f(\bullet)$ to make substantial progress with the question of numerical stability. These structural assumptions confer certain dynamical properties on the underlying equations and *numerical stability is the question of whether, and in what sense, these dynamical properties are inherited by the numerical approximation.* This form of stability is sometimes termed practical stability.

This article is concerned entirely with aspects of *stability* in the integration of differential equations over long time intervals. The question of the *convergence* properties of dynamical systems under discretization is reviewed in [51]; the existence and convergence of a variety of invariant sets (such as equilibria, unstable manifolds, periodic solutions, and strange attractors) is studied. Note that it is primarily through stability analyses that it is possible to distinguish between the usefulness of different integration techniques over long time intervals. Convergence of invariant sets, if it occurs, typically occurs for all consistent numerical methods and (other than the rate of convergence) such convergence analyses do not distinguish between the usefulness of the different methods [51].

The purposes of the paper are: (1) to unify the classical and the currently evolving numerical stability theories as far as possible; (2) emphasize the somewhat restrictive dynamical properties of the problems covered by classical stability theories and to draw attention to other, more dynamically complex, categories motivated by applications in science, engineering and the theory of differential equations; (3) to show that there are strong relationships between the classical and developing theories and, in particular, to emphasize the unifying role of a form of numerical stability for Runge–Kutta methods termed algebraic stability.

For the purposes of this paper it is possible to think of the numerical methods which approximate (1.1) as mappings of the form

$$U_{n+1} = \phi(U_n; \Delta t). \tag{1.3}$$

We shall only study Runge–Kutta methods in detail here and, for the purposes of this review, it is sufficient to be aware only of the following facts concerning these approximation methods:

(i) while the numerical solution sequence $\{U_0, U_1, U_2, \ldots\}$ remains in a compact set $\mathcal{B}$ there is $\Delta t(\mathcal{B})$ such that the Runge–Kutta method may be thought of as a mapping of the form (1.3) for $0 < \Delta t < \Delta t(\mathcal{B})$;

(ii) Runge–Kutta methods satisfy a local approximation property which may be expressed as

$$\|\phi(u(t_n); \Delta t) - u(t_{n+1})\| \leqslant C\Delta t^{r+1},$$

where $u(t_n)$ satisfies (1.1); this approximation property implies an estimate of the form (1.2);

(iii) Runge–Kutta methods depend on certain parameters (see below) which form a matrix A and vector $\mathbf{b}$. In particular, the matrices M and B formed from A and $\mathbf{b}$ (see below) are important in framing our stability results. The parameters in A and $\mathbf{b}$ are generally adjusted to achieve many different, sometimes conflicting, goals. An example is the choice of the integer r in (1.2). In this paper we shall concentrate on the choices of A and $\mathbf{b}$ which ensure important stability properties, in the sense alluded to earlier. We shall not discuss in detail the important question of how these choices interact with other choices (such as the determination of r).

The notation used for Runge–Kutta methods is now described: given a sequence of points $t_n = n\Delta t$ and approximations $U_n \approx u(t_n)$ to the solution of (1.1) we define a general k-stage Runge–Kutta Method (RKM) by

$$\eta_i = U_n + \Delta t \sum_{j=1}^{k} a_{ij} f(\eta_j), \quad i = 1, \ldots, k,$$

$$U_{n+1} = U_n + \Delta t \sum_{i=1}^{k} b_i f(\eta_i), \quad U_0 = u_0.$$

Let A, I denote the $k \times k$ matrices with entries

$$\{A\}_{ij} = a_{ij}, \qquad \{I\}_{ij} = \delta_{ij},$$

let

$$\mathbf{b} = [b_1, \ldots, b_k]^T, \qquad \mathbf{1} = [1, \ldots, 1]^T,$$

let B denote the $k \times k$ matrix

$$B := \operatorname{diag}(b_1, b_2, \ldots, b_k), \tag{1.4}$$

and let M denote the $k \times k$ matrix

$$M := BA + A^T B - \mathbf{b}\mathbf{b}^T. \tag{1.5}$$

We use the notation

$$m_{ij} = \{M\}_{ij} = b_i a_{ij} + b_j a_{ji} - b_i b_j.$$

Note that, assuming the solvability of the equations for the η_i, the RKM defines a map from $\mathbb{C}^p$ into $\mathbb{C}^p$. For any given U_n, the solvability of the Runge–Kutta equations is ensured for sufficiently small Δt [5]. However, the question of solvability for a complete sequence $\{U_n\}_{n=0}^{\infty}$ and given arbitrary Δt and u_0 is nontrivial and we will return to it throughout the paper when particular structural assumptions on $f(u)$ allow us to make more detailed comments. However, all general statements about the large n behavior of the RKMs are made on the implicit assumption that a solution sequence exists. □

Ensuring stability usually boils down to certain constraints on the coefficients in the matrix A and vector $\mathbf{b}$ which define the Runge–Kutta method. The classical theories of A- and AN-stability (for *linear decay* problems) and B- and algebraic stability (for *contractive nonlinear problems*) are reviewed with emphasis placed on the implications of the structural assumptions for the relatively simple dynamics of the underlying equations. Various other classes of problems, which admit complicated dynamics, are then discussed. Specifically *gradient*, *dissipative*, *conservative*, and *Hamiltonian* equations are considered in turn. (Note that contractive problems are often referred to as dissipative in the numerical analysis literature; this conflicts with the definition of dissipativity in the differential equations literature which we use here; see §5 and [30].) For most of these problems numerical stability theory is far from complete and is currently developing. Nonetheless, we make it clear that there are striking relationships with the classical theory.

Sections 2–7 go through a sequence of model problems relevant to numerical stability, starting with linear decay and ending with Hamiltonian systems. In section 8 we discuss briefly analogous problems for linear multistep and one-leg methods. Section 9 contains a description of the effect of error control on numerical stability; it is shown that algebraic stability is useful in the analysis of variable step-size codes which are not algebraically stable in a fixed step implementation. Section 10 contains the conclusions and several open problems.

In summary we find the following important role played by the matrices M and B in numerical stability theory; the precise meaning of stability in each case can be found by reference to the appropriate section. The symbol $\geqslant$ in the context of matrices means *positive semidefinite*.

- Contractive problems (§3);

$$M \geqslant 0,\ B \geqslant 0 \Rightarrow \quad \textit{stability}.$$

- Dissipative gradient problems (§§4 and 5);

$$M \geqslant 0,\ B \geqslant 0 \Rightarrow \quad \textit{stability}.$$

- Dissipative problems (§5);

$$M \geqslant 0,\ B \geqslant 0 \Rightarrow \quad \textit{stability}.$$

- Conservative problems (§6);

$$M \equiv 0 \Rightarrow \quad \textit{stability}.$$

- Orthonormality preserving matrix equations (§6);

$$M \equiv 0 \Rightarrow \quad \textit{stability}.$$

- Hamiltonian problems (§7);

$$M \equiv 0 \Rightarrow \quad \textit{stability}.$$

It should be emphasized that some numerical methods not covered by these stability theories may behave well in practice. Thus the results should be viewed with some caution. Nonetheless we believe that it is valuable to provide some firm theoretical basis for the analysis of qualatative properties of integration techniques. In addition, such a basis helps to identify questions of interest in the future development and analysis of numerical methods.

Note that the simplest method with $M \geqslant 0$, $B \geqslant 0$ is the *backward Euler scheme*

$$U_{n+1} = U_n + \Delta t f(U_{n+1}),$$

while the simplest method with $M \equiv 0$ is the *implicit midpoint rule*

$$U_{n+1} = U_n + \Delta t f(\frac{1}{2}[U_{n+1} + U_n]). \tag{1.6}$$

Much of the analysis described here can also be developed for the study of time integration methods for partial differential equations; indeed some of it was initially developed in that context. Throughout we illustrate the various categories of equations by considering the following partial differential equation.

Example. The Ginzburg–Landau equation for a complex function $u(x, t)$ satisfies

$$u_t = (\hat{a} + i\hat{b})u_{xx} - (\hat{c} + i\hat{d})|u|^2 u + \hat{e}u, \qquad x \in (0, 1), \tag{1.7}$$

$$u(0, t) = u(1, t), \qquad u_x(0, t) = u_x(1, t). \tag{1.8}$$

Here $\hat{a}, \hat{b}, \hat{c}, \hat{d}, \hat{e} \in \mathbb{R}$. In this context we introduce the inner product

$$\langle u, v \rangle = \int_0^1 Re(u\bar{v})dx$$

and corresponding L_2 norm

$$\|u\|^2 = \int_0^1 |u|^2 dx.$$

Provided that $\hat{a}$ and $\hat{c}$ are positive this equation has a unique bounded solution for all time $t \geqslant 0$ given arbitrary initial data in $L_2((0, 1))$ [53].

Under spatial discretization this equation yields a system of ordinary differential equations in the form (1.1). Thus all statements about the complex partial differential equation have natural analogues for related systems of ordinary differential equations provided that the spatial discretization confers those properties from the infinite dimensional problem to the finite dimensional one. This can be achieved in many cases but the precise form of spatial discretization will vary depending upon the structural assumption under consideration. For simplicity of exposition we shall discuss (1.7), (1.8) directly as an illustrative example and ignore the (important) issue of appropriate spatial discretization. □

2. Linear decay. The analysis of the large-time behavior of numerical methods for initial value problems begins with the study of the linear, constant coefficient test problem (1.1) together with the assumption of *linear decay*

$$f(u) = \lambda u, \quad \mathrm{Re}(\lambda) \leqslant 0, \quad p = 1, \tag{2.1}$$

where $u \in \mathbb{C}$ and p is the dimension of the problem. See [14] and [19] and the references cited therein. In this section we use the standard norm $\|u\|^2 = u\bar{u}$ on $\mathbb{C}$. The following solution behavior may be easily established.

RESULT 2.1. *Any two solutions $u(t)$, $v(t)$ of (1.1), (2.1) satisfy*

$$\|u(t) - v(t)\| \leqslant \|u(0) - v(0)\|$$

for all $t \geqslant 0$. *Furthermore, if the inequality in* (2.1) *is strict, then*

$$\lim_{t\to\infty} u(t) = 0$$

for any $u(0) \in \mathbb{C}$.

Numerical stability analysis focuses on determining conditions under which the numerical method replicates these properties. This is the motivation behind the following definition [7].

DEFINITION 2.2. *A* RKM *is said to be* A-stable *provided that the function*

$$R(z) = 1 + z\mathbf{b}^T(I - zA)^{-1}\mathbf{1}$$

satisfies $|R(z)| < 1$ *for all* $z : Re(z) < 0$.

It is worth noting that there are also algebraic characterisations of A-stability; see [45]. Straightforward analysis shows that, for a RKM applied to (1.1), (2.1), $U_{n+1} = R(\Delta t\lambda)U_n$ and hence (see, for example, [7] and [40]) we obtain the following.

RESULT 2.3. *Any two solution sequences* $\{U_n\}_{n=0}^{\infty}$ *and* $\{V_n\}_{n=0}^{\infty}$ *of an* A-*stable* RKM *applied to the problem* (1.1), (2.1) *satisfy*

$$\|U_{n+1} - V_{n+1}\| \leqslant \|U_n - V_n\| \tag{2.2}$$

for all $n \geqslant 0$. *Furthermore, if the inequality in* (2.1) *is strict, then*

$$\lim_{n\to\infty} \|U_n\| = 0 \tag{2.3}$$

for all $\Delta t > 0$ *and any* $U_0 \in \mathbb{C}$.

Remark. For A-stable RKMs applied to (1.1), (2.1) the unique solvability of the defining equations is guaranteed for all $\Delta t > 0$ if $I - zA$ is invertible for any $z = \lambda\Delta t$ in the left-half plane. Typically $I - zA$ will be invertible in the left-half plane since, where it is not, poles occur in the stability function and A-stability cannot hold. However, cancellation of factors in the stability function can lead to methods that are A-stable but not invertible for certain isolated values of $z = \lambda\Delta t$ in the left-half plane; the scheme

$$\begin{aligned}
\eta_1 &= U_n + \Delta t f(\eta_1),\\
\eta_2 &= U_n + 2\Delta t f(\eta_1) - \Delta t f(\eta_2),\\
U_{n+1} &= U_n + 2\Delta t f(\eta_1) - \Delta t f(\eta_2)
\end{aligned}$$

has a linear stability function that is equivalent to backward Euler (which is A-stable) but $I - zA$ is noninvertible for $z = \lambda\Delta t = -1$.

It is possible to generalize this theory into a conditional theory where the properties of Result 2.1 are inherited for sufficiently small Δt. This leads to the following result (see [7] and [40]).

RESULT 2.4. *The region of absolute stability* $\mathcal{S}$ *for a* RKM *is the open set in the complex plane for which* $z \in \mathcal{S} \leftrightarrow |R(z)| < 1$. *If* $z = \lambda\Delta t \in \bar{\mathcal{S}}$, *then any two solution sequences* $\{U_n\}_{n=0}^{\infty}$ *and* $\{V_n\}_{n=0}^{\infty}$ *of a* RKM *applied to the problem* (1.1), (2.1) *satisfy* (2.2) *and if* $z \in \mathcal{S}$, *then* (2.3) *holds.*

Remark. Remarks analogous to those following Result 2.3 also apply in this case. □

There is an important point to raise about Results 2.3 and 2.4 in the context we are considering: since the problem is linear, conditions for ensuring this correct large-time behavior are *independent of the amplitude of initial conditions.* As we shall show, in general, dependence on initial data is a barrier to complete conditional theories for nonlinear problems.

Nonautonomous analogues of (2.1), with λ depending on t, are considered in [4]. This resulted in the definition of *AN-stability*; the motivation for this definition is to ensure that the numerical solution decays on a step-by-step basis, mimicing the behavior of the differential equation. The AN-stable methods are a subset of the A-stable methods. This nonautonomous linear stability theory arises natrually in the context of a general class of nonlinear problems—see Result 3.4.

3. Contractive nonlinear problems. Clearly the linear problems of §2 are very restrictive and naturally attempts were made to study nonlinear problems. The first class of nonlinear problems studied in any systematic way in the context of numerical stability were *contractive* problems introduced by Dahlquist [15], [16]. For simplicity of exposition we will consider the case where (1.1) is real and $f(u) \in C^1(\mathbb{R}^p, \mathbb{R}^p)$; the condition for contractivity in an inner-product norm is

$$\langle f(u) - f(v), u - v\rangle \leqslant 0 \quad \forall u, v \in \mathbb{R}^p, u \neq v. \tag{3.1}$$

A simple example of an equation satisfying (3.1) is the following.

Example. For $p = 1$ and $f(u) = -u^3$ we have

$$\begin{aligned}\langle f(u) - f(v), u - v\rangle &= -(u^2 + uv + v^2)(u - v)^2 \\ &= -\frac{1}{2}[(u + v)^2 + u^2 + v^2](u - v)^2 \leqslant 0. \qquad \square\end{aligned}$$

Example. Consider (1.7) and (1.8) with $\hat{b} = \hat{d} = \hat{e} = 0$ and $\hat{a} = \hat{c} = 1$ and $u(x, t) \in \mathbb{R}$. This gives the scalar reaction-diffusion equation

$$u_t = u_{xx} - u^3,$$

together with periodic boundary conditions on the unit interval. Then, taking the right-hand side of this equation as $f(u)$ we can show that (3.1) holds, using integration by parts:

$$\begin{aligned}\langle u_{xx} - u^3 - v_{xx} + v^3, u - v\rangle &= \int_0^1 \{(u - v)(u - v)_{xx} - (u^3 - v^3)(u - v)\}dx \\ &= -\int_0^1 \{(u_x - v_x)^2 + \frac{1}{2}[(u + v)^2 + u^2 + v^2](u - v)^2\}dx \leqslant 0.\end{aligned}$$

Thus the problem is contractive and satisfies an infinite-dimensional analog of (3.1). $\square$

The following notation will be useful: $\mathcal{E}$ denotes the set of equilibrium points of (1.1) and $\mathcal{E}_{\Delta t}$ denotes the set of fixed points of the RKM. Throughout we will use the following definition for the distance between a point $x \in \mathbb{R}^p$ and a set $\mathcal{B} \subset \mathbb{R}^p$:

$$\operatorname{dist}(x, \mathcal{B}) = \inf_{y \in \mathcal{B}} \|x - y\|.$$

For problems satisfying (3.1) the following result holds, which, in the case of strict contractivity, shows the existence of a unique globally attracting equilibrium point.

RESULT 3.1. *Any two solutions $u(t)$, $v(t)$ of* (1.1), (3.1) *satisfy*

$$\|u(t) - v(t)\| \leqslant \|u(0) - v(0)\|$$

for all $t \geqslant 0$. Furthermore, $\mathcal{E}$ is a closed convex set and, if the inequality (3.1) *is strict for all $u, v : v \in \mathcal{E}, u \notin \mathcal{E}$, then*

$$\lim_{t \to \infty} \operatorname{dist}(u(t), \mathcal{E}) \to 0.$$

Finally, if the inequality in (3.1) *is strict and* $\exists \bar{u} : f(\bar{u}) = 0$ *then* $\bar{u}$ *is a unique equilibrium point and*

$$\lim_{t\to\infty} u(t) = \bar{u}.$$

Proof. A calculation shows that

$$\frac{1}{2}\frac{d}{dt}\|u - v\|^2 = \langle u - v, f(u) - f(v)\rangle \leqslant 0. \tag{3.2}$$

Thus the first result follows.

To prove that the steady states of the system define a convex set it is sufficient to show that any convex combination of zeros of f is also a zero of f. Let $z = \lambda x + (1-\lambda)y$ where $f(x) = f(y) = 0$, $\lambda \in (0, 1)$ and define $z' = z + \delta f(z)$, $\delta > 0$. Then $z' - x = \delta f(z) + (\lambda - 1)(x - y)$. Now (3.1) implies that

$$\langle f(z'), z' - x\rangle \leqslant 0,$$

and hence

$$\langle f(z'), \delta f(z)\rangle \leqslant (1 - \lambda)\langle f(z'), x - y\rangle.$$

Similarly

$$\langle f(z'), z' - y\rangle \leqslant 0$$

implies that

$$\langle f(z'), \delta f(z)\rangle \leqslant -\lambda\langle f(z'), x - y\rangle.$$

Notice that, since $(1 - \lambda)$ and $-\lambda$ have opposite signs, we have $\langle f(z'), \delta f(z)\rangle \leqslant 0$, which is equivalent to $\langle f(z + \delta f(z)), f(z)\rangle \leqslant 0$ since $\delta > 0$; letting $\delta \to 0$ and using the continuity of f, we obtain $\|f(z)\|^2 \leqslant 0$, and thus $f(z) = 0$. Convexity follows. Let $u_i \to u^*$ be such that $f(u_i) = 0$ for each i. By the continuity of $f(\bullet)$ it also follows that $f(u^*) = 0$ and hence that $\mathcal{E}$ is closed.

Now assume that (3.1) is strict for $v \in \mathcal{E}$ and $u \notin \mathcal{E}$. Define the set U by

$$U = \{u \in \mathbb{R}^p : r \leqslant \operatorname{dist}(u, \mathcal{E}) \leqslant R\}.$$

Then, if $u(0) \in U$ it follows that there exists $\bar{u} \in \mathcal{E}$ for which

$$\|u(0) - \bar{u}\| \leqslant R.$$

Thus it follows that, for all $t \geqslant 0$,

$$\operatorname{dist}(u(t), \mathcal{E}) \leqslant \|u(t) - \bar{u}\| \leqslant \|u(0) - \bar{u}\| \leqslant R. \tag{3.3}$$

Now assume, for the purposes of contradiction, that $\operatorname{dist}(u(t), \mathcal{E}) > r$ for all $t \geqslant 0$. Let

$$\tilde{U} = U \bigcap \{u \in \mathbb{R}^p : \|u - \bar{u}\| \leqslant R\}$$

and then define

$$\epsilon := \epsilon(r, R) = \inf_{u\in\tilde{U}} \langle f(u), \bar{u} - u\rangle.$$

Note that $\tilde{U}$ is compact since it is formed as the intersection of a compact set with a closed set. Clearly $\epsilon > 0$ since $\tilde{U}$ is compact, and since strict inequality holds in (3.1) as $\bar{u} \in \mathcal{E}$, $u \notin \mathcal{E}$. Thus, by assumption and by (3.3) we have $u(t) \in \tilde{U}$ for all $t \geqslant 0$, and hence

$$\frac{1}{2}\frac{d}{dt}\|u - \bar{u}\|^2 \leqslant -\epsilon \quad \forall t \geqslant 0.$$

Hence, as $t \to \infty$,

$$\|u - \bar{u}\|^2 \to -\infty,$$

a contradiction. Thus there exists a time $t^*(r, R)$ for which $\text{dist}(u(t), \mathcal{E}) \leqslant r$. Replacing R by r we deduce from (3.3) that $\text{dist}(u(t), \mathcal{E}) \leqslant r$ for all $t \geqslant t^*(r, R)$. Since r is arbitrary the result follows.

For the case of strict inequality for all $u, v : u \neq v$, uniqueness of $\bar{u}$ follows automatically since otherwise we have a contradiction. Thus $\mathcal{E} = \{\bar{u}\}$ and the preceding argument establishes that

$$\lim_{t\to\infty} u(t) = \bar{u}$$

as required. □

The original motivation for the study of these problems was to generalize the notion of contractivity from linear to nonlinear problems since this notion is fundamental in understanding certain kinds of error propagation for numerical methods when applied to stiff systems. However, as a result, the large-time behavior of (1.1), (3.1) is very closely related to that of the model linear problem (1.1), (2.1) (compare Results 2.1 and 3.1)—essentially all solutions are attracted to the unique fixed point or set of fixed points. Runge–Kutta methods for (1.1), (3.1) were studied in [6], [4]. These studies resulted in the following definitions.

DEFINITION 3.2. *An* RKM *is said to be algebraically stable if the matrices B and M defined by* (1.4), (1.5) *are positive semidefinite. An* RKM *is said to be B-stable if, when applied to* (1.1), (3.1), *any two solution sequences* $\{U_n\}_{n=0}^{\infty}$, $\{V_n\}_{n=0}^{\infty}$ *satisfy*

$$\|U_{n+1} - V_{n+1}\| \leqslant \|U_n - V_n\|$$

for any $U_0, V_0 \in \mathbb{R}^p$ *and any* $\Delta t \geqslant 0$.

Remark. (i) All algebraically stable RKMs are equivalent to an algebraically stable RKM with B strictly positive definite. If $b_l = 0$ for some l, then the equation for η_l decouples from the other η_i and is redundant. In technical jargon all algebraically stable RKMs are *DJ-redicible* to a method with B strictly positive definite—see [29]. Thus for the purposes of this article, in all proofs concerning algebraically stable methods, we will assume that $b_i > 0$ for all i. The case when $b_l = 0$ for some l can be dealt with simply by using the aforementioned equivalence.

(ii) There exist arbitrarily high-order schemes that are algebraically stable, but all of them are implicit, that is, they involve the solution of nonlinear equations at each step. □

Once again, numerical stability is the requirement that a certain qualitative property of the differential equation is inherited by the numerical method. The next result shows that the purely algebraic criterion of Definition 3.2 is important in this context; it is a discrete analogue of Result 3.1 and, for strictly contractive problems yields the existence of a unqiue globally attracting equilibrium for a suitable class of RKMs. The implication between algebraic stability and B−stability was proved in [4].

RESULT 3.3. *Any two solution sequences* $\{U_n\}_{n=0}^{\infty}$ *and* $\{V_n\}_{n=0}^{\infty}$ *of an algebraically stable* RKM *applied to the problem* (1.1), (3.1) *satisfy*

$$\|U_{n+1} - V_{n+1}\| \leqslant \|U_n - V_n\|$$

for all $n \geqslant 0$. *Hence*

$$\textit{algebraic stability} \Rightarrow \textit{B-stability.}$$

Furthermore, if the inequality (3.1) *is strict for all* $u, v : v \in \mathcal{E}, u \notin \mathcal{E}$ *then* $\mathcal{E}_{\Delta t} \equiv \mathcal{E}$ *and*

$$\lim_{n\to\infty} \operatorname{dist}(U_n, \mathcal{E}) \to 0.$$

Finally, if the inequality in (3.1) *is strict and there exists* $\bar{u} : f(\bar{u}) = 0$ *then* $\bar{u}$ *is a unique equilibrium point of the* RKM *and*

$$\lim_{n\to\infty} U_n = \bar{u}$$

for all $\Delta t > 0$ *and any* $U_0 \in \mathbb{R}^p$.

Proof. Let the sequence V_n satisfy

$$\xi_i = V_n + \Delta t \sum_{j=1}^{k} a_{ij} f(\xi_j), \quad i = 1, \ldots, k,$$

$$V_{n+1} = V_n + \Delta t \sum_{i=1}^{k} b_i f(\xi_i), \quad U_0 = u_0.$$

We also define

$$D_n = U_n - V_n, \quad E_i = \eta_i - \xi_i, \quad F_i = f(\eta_i) - f(\xi_i).$$

Then

$$D_{n+1} = D_n + \Delta t \sum_{j=1}^{k} b_j F_j$$

and

$$E_i = D_n + \Delta t \sum_{j=1}^{k} a_{ij} F_j,$$

and it follows that

$$\|D_{n+1}\|^2 = \|D_n\|^2 + 2\Delta t \sum_{j=1}^{k} b_j \langle D_n, F_j \rangle + \Delta t^2 \sum_{i,j=1}^{s} b_i b_j \langle F_i, F_j \rangle.$$

$$= \|D_n\|^2 + \Delta t \sum_{j=1}^{k} b_j \langle D_n, F_j \rangle + \Delta t \sum_{i=1}^{k} b_i \langle D_n, F_i \rangle + \Delta t^2 \sum_{i,j=1}^{s} b_i b_j \langle F_i, F_j \rangle.$$

Using the fact that

$$\langle D_n, F_i \rangle = \langle E_i, F_i \rangle - \Delta t \sum_{j=1}^{k} a_{ij} \langle F_i, F_j \rangle$$

and that

$$\langle D_n, F_j\rangle = \langle E_j, F_j\rangle - \Delta t\sum_{i=1}^{k} a_{ji}\langle F_i, F_j\rangle$$

we obtain, since the scheme is algebraically stable,

$$\begin{aligned}\|D_{n+1}\|^2 &= \|D_n\|^2 + 2\Delta t\sum_{j=1}^{k} b_j\langle E_j, F_j\rangle - \Delta t^2\sum_{i,j=1}^{k} m_{ij}\langle F_i, F_j\rangle\\ &\leqslant \|D_n\|^2 + 2\Delta t\textstyle\sum_{j=1}^{k} b_j\langle E_j, F_j\rangle.\end{aligned}$$

Thus we have

$$\|U_{n+1} - V_{n+1}\|^2 \leqslant \|U_n - V_n\|^2 + 2\Delta t\sum_{j=1}^{k} b_j\langle \eta_j - \xi_j, f(\eta_j) - f(\xi_j)\rangle. \tag{3.4}$$

Using (3.1) it follows that

$$\|U_{n+1} - V_{n+1}\| \leqslant \|U_n - V_n\|$$

and B-stability is established.

Now we assume that (3.1) holds with strict inequality for $v \in \mathcal{E}$ and $u \notin \mathcal{E}$. To obtain a contradiction assume that there exists $\bar{w} \notin \mathcal{E}$, which is a fixed point of the Runge–Kutta method, and let $\bar{u} \in \mathcal{E}$. Since $f(\bar{u}) = 0$ the Runge–Kutta equations have a solution $\eta_i = \bar{u}$, $i = 1, \ldots, k$ and $\bar{u}$ is also a fixed point of the Runge–Kutta method; see [38]. Thus from (3.4), setting $U_n = \bar{u}$ and $V_n = \bar{w}$,

$$\|\bar{u} - \bar{w}\|^2 \leqslant \|\bar{u} - \bar{w}\|^2 + 2\Delta t\sum_{j=1}^{k} b_j\langle \bar{u} - \xi_j, f(\bar{u}) - f(\xi_j)\rangle. \tag{3.5}$$

In addition it is not possible for all the ξ_j to be contained in $\mathcal{E}$ for, if they were, then $f(\bar{w}) = f(\xi_i) = 0$, which implies that $\bar{w} \in \mathcal{E}$ and this is not possible. Hence there exists j such that

$$\langle \bar{u} - \xi_j, f(\bar{u}) - f(\xi_j)\rangle < 0$$

and furthermore it is known that for algebraically stable Runge–Kutta methods we may assume that $b_i > 0$ for all i (see the Remark following Definition 3.2). Thus, from (3.1) and (3.5) we have that

$$\|\bar{u} - \bar{w}\|^2 < \|\bar{u} - \bar{w}\|^2,$$

a contradiction, and hence such a $\bar{w}$ cannot exist.

We now prove that $\mathcal{E}$ is attracting; the same notation is employed as for the proof of Result 3.1. Let $U_0 \in U$. Then, by (3.1) and (3.4) it follows that there exists $\bar{u} \in \mathcal{E}$ such that

$$\operatorname{dist}(U_n, \mathcal{E}) \leqslant \|U_n - \bar{u}\| \leqslant \|U_0 - \bar{u}\| \leqslant R.$$

Assume for the purposes of contradiction that $U_n \in \tilde{U}$ for all $n \geqslant 0$. Now notice that, if $U_n \notin \mathcal{E}$, then there exists $j : \eta_j \notin \mathcal{E}$ since otherwise $U_n = \eta_i \in \mathcal{E}$. Thus

$$\epsilon(r, R) = \inf_{u\in\tilde{U}} \max_{1\leqslant j\leqslant k} \langle f(\eta_j), \bar{u} - \eta_j\rangle > 0$$

and

$$b_{\min} = \min_{1 \leqslant j \leqslant k} b_j > 0.$$

From (3.4) we have that

$$\|U_{n+1} - V_{n+1}\|^2 \leqslant \|U_n - V_n\|^2 - 2\Delta t b_{\min} \max_{1 \leqslant j \leqslant k} \langle \bar{u} - \eta_j, f(\eta_j) \rangle.$$

so that, since $\bar{u} \in \tilde{U}$,

$$\|U_{n+1} - \bar{u}\|^2 \leqslant \|U_n - \bar{u}\|^2 - 2\Delta t b_{\min} \epsilon(r, R) \quad \forall n \geqslant 0.$$

Letting $n \to \infty$ gives a contradiction and hence we deduce that there exists $n^*(r, R)$ for which $\operatorname{dist}(U_n, \mathcal{E}) \leqslant r$. Since r is arbitrary the result follows as for Result 3.1.

Finally, assume that (3.1) holds with strict inequality for all $u \neq v$ and that $f(\bar{u}) = 0$. In Result 3.1 we established that $\bar{u}$ is the unique fixed point of (1.1), (3.1) and hence that $\mathcal{E} = \bar{u}$. Applying the previous part of this result to the case where the inequality (3.1) is strict for all $u, v: v \in \mathcal{E}, u \notin \mathcal{E}$ proves that $\bar{u}$ is the unique fixed point of the RKM. Since $\mathcal{E} = \bar{u}$ the convergence result from the previous case can also be applied to show that $U_n \to \bar{u}$ as $n \to \infty$. □

Remark. (i) The role of algebraic stability in the proof is to enable a certain quadratic form, which is defined by the matrix M, to be bounded above when manipulating inequalities and yielding (3.4). This basic idea, and variants on it, will recur throughout the paper.

(ii) In Result 3.3 we have not considered the solvability of the implicit Runge–Kutta equations. The existence of unique solutions under (3.1), for any U_n and any $\Delta t \geqslant 0$, has been established for many classes of algebraically stable methods including those based on Gauss–Legendre quadrature, for which $M \equiv 0$, the Radau IA, IIA and Lobatto IIIC methods; see [19] and [29].

(iii) In [38] it was observed that $\mathcal{E} \subseteq \mathcal{E}_{\Delta t}$ for RKMs. The class of methods for which $\mathcal{E} \equiv \mathcal{E}_{\Delta t}$ for all $\Delta t > 0$ and *all* autonomous problems (1.1) was termed *regular*; various order barriers for stable regular methods are proved in [28] ruling out high order, regular, stable methods. However, Result 3.3 shows that if a contractive structure is imposed on f, there exist stable methods of arbitrarily high order satisfying $\mathcal{E} \equiv \mathcal{E}_{\Delta t}$. □

Butcher [6] took B-stability as a basic definition and it was only later that the significance of algebraic stability was discovered in [4]. This was achieved through the study of AN-stability as defined in §2. It is clear from Results 2.1 and 3.1 that the problems (1.1), (2.1) and (1.1), (3.1) are very closely related and this is reflected in the close relationship between the stability theories. The following remarkable result proved in [37] is an extension of results proved in [4] and [13]; recall the concept of AN-stability described in §2.

RESULT 3.4. *For S-irreducible* RKMs

algebraic stability $\Leftrightarrow$ *AN-stability* $\Leftrightarrow$ *B-stability* $\Rightarrow$ *A-stability.*

Remark. S-irreducibility is a technical property defined in [29]. We will not reproduce the definition here, but restrict ourselves to noting that these methods are widely occurring in practice. □

This is only a brief overview of the theories for linear decay and contractive nonlinear problems; for further details see [19] and [29]. The theory of nonlinear contractive problems has been extended to contraction in norms other than those induced by an inner product [47], [48] and a very clear account can be found in [39].

In contrast to the linear decay problem, any conditional theory of numerical contractivity (a generalization of the concept of region of absolute stability) will involve dependence of the

allowable time-step on the initial data and is hence much harder to develop. However, explicit methods operating with error control are frequently oberserved to overcome such difficulties automatically. Indeed in some cases it is possible to prove that error control confers desirable stability properties for a range of the tolerance τ that is independent of initial data; see §9 and [52].

To make progress with a conditional theory for nonlinear problems it is necessary to impose still further restrictions on the class of problems. Much of the generalized theory of contractivity reviewed in [39] employs the *circle condition* [17]

$$\exists \rho > 0 : \|f(u) - f(v) + \rho(u - v)\| \leqslant \rho\|u - v\| \quad \forall u, v \in \mathbb{R}^p. \tag{3.6}$$

This condition can only be satisfied by globally Lipschitz functions which limit somewhat the range of direct applications. The motivation behind (3.6) is to combine it with some a priori bounds on the underlying numerical approximations which enables the vector field defining the differential equation to be replaced by a globally Lipschitz one satisfying (3.6). However this important step is rarely addressed in the literature. In §§4 and 5 we describe conditions under which such global *a priori bounds* on numerical solutions may be found *independently of initial data.*

RESULT 3.5. *A function $f(\bullet)$ satisfying* (3.6) *is necessarily globally Lipschitz.*

Proof. Assume to the contrary. Then, for any $K > 0$, there exist $u, v \in \mathbb{R}^p$ with $u \neq v$ such that

$$\|f(u) - f(v) + \rho(u - v)\| \geqslant \|f(u) - f(v)\| - \rho\|u - v\| \geqslant (K - \rho)\|u - v\|.$$

Now choosing $K > 2\rho$ contradicts (3.6). This completes the proof. □

It is worth noting that the circle condition (3.6) is implied by the assumption

$$\exists \alpha > 0 : \langle f(u) - f(v), u - v\rangle \leqslant -\alpha\|f(u) - f(v)\|^2 \quad \forall u, v \in \mathbb{R}^p, u \neq v.$$

The circle condition (3.6) then holds with $\rho = 1/\alpha$. In this sense it can be seen that (3.6) is a very special case of the contractivity condition (3.1).

We now go on to show that the numerical stability theory developed for contractive problems forms a natural bridge for the study of a wide variety of other nonlinear problems.

4. Gradient systems. As in §3, for simplicity of exposition we will consider the case where (1.1) is real and $f(u) \in C^1(\mathbb{R}^p, \mathbb{R}^p)$. It is clear from §§2 and 3 that linear decay and contractivity are such strong conditions that they rule out complicated dynamics, and hence it is natural to relax the notion of contractivity to allow some expansion of trajectories. The function f is said to satisfy a *one-sided Lipschitz condition* if there exists a constant $c > 0$ such that

$$\langle f(u) - f(v), u - v\rangle \leqslant c\|u - v\|^2 \quad \forall u, v \in \mathbb{R}^p. \tag{4.1}$$

This allows exponential separation of trajectories and specifically it is straightforward to prove the following.

RESULT 4.1. *Any two solutions $u(t)$, $v(t)$ of* (1.1), (4.1) *satisfy*

$$\|u(t) - v(t)\| \leqslant e^{ct}\|u(0) - v(0)\|$$

for all $t \geqslant 0$.

Numerical counterparts of Result 4.1 have been studied and these are useful in establishing continuity of the numerical solution with respect to initial data—see Butcher [7] and [29].

Solvability of the Runge–Kutta equations in this context is discussed in [19]. The importance of continuity with respect to initial data will become apparent in Result 4.4.

Since exponential separation of trajectories allows the possibility of exponential growth of the solutions themselves, (4.1) alone is far too broad a class of problems to work with and make substantial progress; for this reason it is sensible to add further structure to the problem. Both linear decay and strictly contractive nonlinear problems are characterised by the property that $u(t)$ approaches a unique equilibrium as time increases. This can be relaxed to the notion that $u(t)$ approaches an equilibrium as time increases but that it is *not necessarily unique.* This leads naturally to the class of *gradient systems* for which there exists $F \in C^2(\mathbb{R}^p, \mathbb{R})$ such that

$$\text{(4.2)} \qquad \begin{cases} f(u) = -\nabla F(u) & \forall u \in \mathbb{R}^p \\ F(u) \geqslant 0 & \forall u \in \mathbb{R}^p, \\ F(u) \to \infty & \text{as } \|u\| \to \infty. \end{cases}$$

For gradient systems it follows that

$$\text{(4.3)} \qquad \frac{d}{dt}(F(u)) = \langle \nabla F(u), u_t \rangle = -\langle f(u), u_t \rangle = -\|u_t\|^2.$$

Hence, arguing loosely, we see that u will be driven to the critical points of F, which are the equilibria of (1.1). If F is convex so that

$$\langle \nabla F(u) - \nabla F(v), u - v \rangle \geqslant 0 \quad \forall u, v \in \mathbb{R}^p.$$

then (4.2) is a contractive problem and the analysis of §2 applies; in particular, the set of equilibria define a convex set. However, for nonconvex F equation (1.1), (4.2) may have multiple isolated equilibria. A simple example is the following.

Example. Consider equation (1.1) in dimension $p = 1$ with $f(u) = u - u^3$. This is in gradient form with

$$F(u) = \frac{1}{4}(u^2 - 1)^2.$$

Notice the three equilibria $0, 1, -1$. □

Example. Consider equation (1.7), (1.8) with $\hat{b} = \hat{d} = 0$ and $\hat{a} = \gamma, \hat{c} = \hat{e} = 1$ and $u(x, t) \in \mathbb{R}$. Then, defining

$$F(u) = \int_0^1 \frac{\gamma}{2} u_x^2 + \frac{1}{4}(u^2 - 1)^2 dx,$$

the equation may be written as

$$u_t = -\nabla F(u),$$

where ∇ is now interpreted as the variational derivative of $F(u)$ with respect to changes in u, confined to an appropriate function space satisfying the boundary conditions. □

Gradient systems arise in a variety of applications; in particular, many phenomenological models of phase transitions such as the solid/solid Cahn–Hilliard equation [22] and the super/normal conducting Ginzburg–Landau equations [9] are in gradient form. Furthermore, gradient systems have been fundamental in the development of many important concepts in the theory of dynamical systems and are important for this reason alone; see [30] and the

references therein. As suggested by (4.3) gradient systems are characterised by the following behavior proved, for example, in [33].

RESULT 4.2. *For any solution $u(t)$ of* (1.1), (4.2) *and any sequence $t_i \to \infty$ for which the ω-limit point*

$$x := \lim_{i\to\infty} u(t_i) \tag{4.4}$$

exists, it follows that $x \in \mathcal{E}$, the set of zeros of f. Furthermore, if all members of $\mathcal{E}$ are isolated then, for each $u(0)$ there exists $\bar{u} := \bar{u}(u(0)) \in \mathcal{E}$ such that

$$\lim_{t\to\infty} u(t) = \bar{u}.$$

Proof. Given $u(0) \in \mathbb{R}^p$ let $w(u(0))$ be the union of points such that (4.4) is defined for some sequence t_i. Then $\omega(u(0))$ is known as the ω-limit set and is a closed, invariant (under forward evolution of the differential equation) set which is connected if compact [2].

Let $x, y \in \omega(u(0))$. Then $F(y) = F(x)$ for otherwise we obtain a contradiction to (4.3). Now consider the solution $u(t)$ of (1.1), (4.2) with $u(0) = x \in \omega(u(0))$; since the ω-limit set is invariant it follows that $u(t) \in \omega(u(0))$ and hence that $F(u(t)) = F(u(0))$ for all $t \geqslant 0$. By (4.3) this implies that $u_t \equiv 0$ for all $t \geqslant 0$ and hence that $u(0) = x \in \mathcal{E}$.

Finally, note that since $0 \leqslant F(u(t)) \leqslant F(u(0))$ it follows from (4.2) that all trajectories are uniformly bounded as $t \to \infty$. Thus $\omega(u(0))$ is compact since it is closed and we deduce that it is also connected. Since the equilibria are isolated it follows that the ω-limit set must be a single point $\bar{u} \in \mathcal{E}$. Since the closure of the trajectory is compact it follows that

$$\lim_{t\to\infty} u(t) = \bar{u}$$

as required. □

For gradient systems it is natural to ask that a numerical approximation replicates the property (4.3) that there is a Lyapunov function which drives the solution to equilibrium. Even if the additional constraint (4.1) is imposed it is unlikely to be possible to find a stability theory which holds for arbitrary Δt, since the problems under consideration admit both contractive and divergent behavior. However, it is both feasible and desirable to find restrictions which are *independent of initial data.* This motivates the following definition.

DEFINITION 4.3. *A* RKM *is said to be gradient stable if, when applied to* (1.1), (4.1), (4.2) *there exists $\Delta t_c > 0$ and a function $F_{\Delta t}(\bullet) : \mathbb{R}^p \to \mathbb{R}$ such that, for all $\Delta t \in (0, \Delta t_c)$:*

(i) *$F_{\Delta t}(U) \geqslant 0$ for all $U \in \mathbb{R}^p$;*

(ii) *$F_{\Delta t}(U) \to \infty$ as $\|U\| \to \infty$.*

(iii) *$F_{\Delta t}(U_{n+1}) \leqslant F_{\Delta t}(U_n)$ for all $U_n \in \mathbb{R}^p$;*

(iv) *if $F_{\Delta t}(U_n) \equiv F_{\Delta t}(U_0)$ for all $n \geqslant 0$ then $U_0 \in \mathcal{E}$, the set of equilibrium points for* (1.1), (4.2).

Such a definition was implicit in the work of Elliott [22] where discrete gradient systems were used in the analysis of numerical approximations of the Cahn-Hilliard equation. A theorem closely related to the following result is proved in [25].

RESULT 4.4. *Assume that, given initial data in $\mathbb{R}^p$, the* RKM *generates a unique C^1 map from $\mathbb{R}^p$ into itself. Then, for any solution of a gradient stable* RKM *applied to* (1.1), (4.2) *with $\Delta t \in (0, \Delta t_c)$ and any sequence $n_i \to \infty$ for which the ω-limit point*

$$x := \lim_{i\to\infty} U_{n_i}$$

exists, it follows that $x \in \mathcal{E}$, the set of zeros of f. Furthermore, if all members of $\mathcal{E}$ are isolated then, for each $u(0)$ there exists $\bar{u} := \bar{u}(u(0)) \in \mathcal{E}$ such that

$$\lim_{n\to\infty} U_n = \bar{u}.$$

Proof. Since the map defining the RKM is C^1 it is Lipschitz continuous with respect to initial data. As for the differential equation, the ω-limit set $\omega(u(0))$ is defined as the union of all possible limit points corresponding to given initial data. A similar argument to that in the Result 4.2 shows that U_n is uniformly bounded in n. From Lemma 2.1.2 in [30] it follows that, since the RKM defines a unique sequence, continuously dependent upon initial data, $\omega(U_0)$ is nonempty, compact and invariant and an argument identical to that in the proof of Result 4.2 shows that $x \in \mathcal{E}$.

However, for dynamical systems defined by mappings it does not follow that $\omega(U_0)$ is connected if compact, and a different argument is needed for the last part of the result. Now assume that the members of $\mathcal{E}$ are isolated. Since the solution sequence is bounded it is contained in a compact set B and this implies that there are a finite number of possible equilibria contained in $\omega(U_0)$, say x_j, $j = 1, \ldots, J$, in $\mathcal{E}$. Let $B_j = B(x_j, \delta) := \{u \in \mathbb{R}^p : \|u - x_j\| < \delta\}$, $B^+ = \bigcup_{j=1,\ldots,J} B_j$ and $B^- = B \backslash B^+$; note that B^- is closed by construction. Assume that δ is sufficiently small that $\operatorname{dist}(x, B_k) \geqslant \Delta > 0$ for all $x \in B_j$, $j \neq k$. Note that $\omega(U_0)$ is nonempty. Assume for the purposes of contradiction that $x_1 \in \omega(U_0)$ and that it is not the unique member of $\omega(U_0)$. Then for all $\delta > 0$ there exists a sequence $n_i \to \infty$ such that $U_{n_i} \in B_1$ and $U_{n_i} \to x_1$ as $n_i \to \infty$. Since x_1 is not the unique limit point there is an infinite sequence of integers m_j such that $U_{m_j} \in B_1$ and $U_{m_j+1} \notin B_1$. Since the mapping defined by the RKM is C^1, it is Lipschitz with constant L on B_1 and since x_1 is a fixed point, we deduce that

$$\|U_{m_j+1} - x_1\| \leqslant L\|U_{m_j} - x_1\| \leqslant L\delta.$$

Hence, if $L\delta < \Delta$ we deduce that $U_{m_j+1} \in B^-$ for each j. But B^- is compact and hence the infinite sequence U_{m_j+1} must have a limit point; such a limit point cannot be contained in $\mathcal{E}$ by definition of B^- and this contradicts the first part of the result. This completes the proof, since the sequence is bounded. □

Remark. The assumption that the RKM generates a unique continuously dependent solution sequence is often made; in some cases this can be a rather strong assumption. However, it is not an unreasonable assumption to make for a system that satisfies (4.1): the one-sided Lipschitz condition implies unique solvability of the Runge–Kutta equations for many classes of implicit methods, if Δt is sufficiently small (but independent of initial data), including those based on Gauss–Legendre quadrature, the Radau IA, IIA and Lobatto IIIA, IIB and IIIC methods; see [19] and [29]. Continuous dependence on initial data can be similarly established. □

Further studies of gradient stability may found in [20] where one-step methods for the Cahn–Hilliard equation are examined. Here we present a proof that the theta method

$$U_{n+1} = U_n + \Delta t[(1-\theta)f(U_n) + \theta f(U_{n+1})] \tag{4.5}$$

is gradient stable for $\theta \in [\frac{1}{2}, 1]$. This illustrates some of the issues involved in establishing gradient stability. Note that the condition on θ is equivalent to the condition that the method be A-stable.

RESULT 4.5. *The theta method* (4.5) *is gradient stable for* $\theta \in [\frac{1}{2}, 1]$ *with* $\Delta t_c = 1/c$, *where* c *is the constant in* (4.1), *and*

$$F_{\Delta t}(U) = F(U) + \frac{\Delta t}{2}(1-\theta)\|f(U)\|^2.$$

Proof. In [34] it is shown that, for a gradient system, (4.1) implies that

$$F(u) - F(v) \leqslant \langle f(u), v - u\rangle + c\|u - v\|^2$$

for any $u, v \in \mathbb{R}^p$. Applying this with $u = U_{n+1}$ and $v = U_n$ we obtain

$$\begin{aligned} F(U_{n+1}) - F(U_n) &\leqslant \langle f(U_{n+1}), U_n - U_{n+1}\rangle + c\|U_{n+1} - U_n\|^2 \\ &= \left\langle \frac{1}{\Delta t}\Big[U_{n+1} - U_n - \Delta t(1-\theta) f(U_n) - \Delta t\theta f(U_{n+1})\Big], U_n - U_{n+1}\right\rangle \\ &\quad + \langle f(U_{n+1}), U_n - U_{n+1}\rangle + c\|U_{n+1} - U_n\|^2 \\ &= \left(c - \frac{1}{\Delta t}\right)\|U_{n+1} - U_n\|^2 + (1-\theta)\langle f(U_n) - f(U_{n+1}), U_{n+1} - U_n\rangle \\ &= \left(c - \frac{1}{\Delta t}\right)\|U_{n+1} - U_n\|^2 + \frac{\Delta t}{2}(1-\theta)\Big[\|f(U_n)\|^2 - \|f(U_{n+1})\|^2\Big] \\ &\quad - \frac{\Delta t}{2}(1-\theta)(2\theta - 1)\|f(U_n) - f(U_{n+1})\|^2. \end{aligned}$$

Hence, for $\theta \in [\frac{1}{2}, 1]$,

$$F_{\Delta t}(U_{n+1}) - F_{\Delta t}(U_n) \leqslant \left(c - \frac{1}{\Delta t}\right)\|U_{n+1} - U_n\|^2.$$

Clearly $F_{\Delta t}$ is bounded below for $\theta \leqslant 1$ and, since $F_{\Delta t}(U) \geqslant F(U)$ for all $U \in \mathbb{R}^p$, (ii) of Definition 4.3 follows. It is also clear that $F_{\Delta t}(U)$ is nonincreasing for $\Delta t \in (0, 1/c)$. Furthermore, if $F_{\Delta t}(U_{n+1}) = F_{\Delta t}(U_n)$ then $U_{n+1} = U_n$. The fixed points $\mathcal{E}_{\Delta t}$ for (4.5) coincide with $\mathcal{E}$ and so gradient stability has been established. □

A similar method of proof establishes that the one-leg counterpart of the theta method (4.5) is also gradient stable; see [34].

Remark. As can be seen a complete theory of gradient stability is not yet developed. However, it is worth observing that, if the additional assumption (5.3) (a form of dissipativity) is appended to (4.2) and the equilibria are isolated, then the conclusion of Result 4.4 follows for any algebraically stable RKM—see [36].

5. Dissipative systems. As in §§3 and 4, for simplicity of exposition we will consider the case where (1.1) is real so that $f(u) \in C^1(\mathbb{R}^p, \mathbb{R}^p)$. Even the one-sided Lipschitz condition which we introduced in the previous section is far too restrictive for many interesting applications and so we relax this condition in our study of dissipative problems. Furthermore, gradient systems only allow solutions to approach equilibria for large time so that periodic, quasi-periodic or chaotic behavior is not admitted; the dissipative problems we study will admit such behavior.

The notion of dissipativity is an important one in many physical applications and naturally there is a mathematical abstraction of this idea in the theory of differential equations; see, for example, [30] and [53]. Roughly speaking an initial value problem is said to be *dissipative* if there is a bounded set, in an appropriate function space for the problem, which all solutions enter after a finite time and thereafter remain inside: thus some measure of energy is dissipated outside the bounded set.

To motivate the study of dissipative problems consider first the equation (1.1) under (3.1), together with the assumption that $f(0) = 0$. Taking $v = 0$ in (3.1) we then deduce that

$$\langle f(u), u\rangle \leqslant 0 \quad \forall u \in \mathbb{R}^p. \tag{5.1}$$

It is straightforward to prove from this that

$$\|u(t)\|^2 \leqslant \|u(0)\|^2 \quad \forall t \geqslant 0. \tag{5.2}$$

This property is often termed *monotonicity* or *weak contractivity*. The numerical analogue of property (5.2), under the assumption (3.1) together with $f(u) = 0$, is studied by a number of authors including [10] and [49]. A straightforward application of the theory in §3 shows that algebraic stability is sufficient for a numerical analogue of (5.2) to hold for all step-sizes $\Delta t > 0$ and all initial data. In fact, a wider class of methods suffices in this context as described in [10].

The monotonicity induced by (5.1) can be weakened to enforce monotonicity only outside a certain bounded region of phase space. This corresponds to a notion of dissipation at sufficiently large amplitude. In this section we will concentrate on a particular class of problems where dissipativity is induced by the structural assumption

$$\exists \gamma, \omega > 0 : \langle f(u), u \rangle \leqslant \gamma - \omega \|u\|^2 \quad \forall u \in \mathbb{R}^p. \tag{5.3}$$

Under (5.3) monotonicity is induced outside the set $\mathcal{B} = \{u \in \mathbb{R}^p : \|u\|^2 \leqslant \gamma/\omega\}$. An example of a system satisfying (5.3) is the Lorenz equations, after translation of the origin. Many other examples exist; in particular, infinite dimensional systems such as the complex Ginzburg–Landau equations (see below) and the Navier–Stokes equation (in two dimensions) satisfy generalizations of (5.3) (see [53]) and, under appropriate spatial discretization, the resulting system of ordinary differential equations satisfy (5.3). (Note that the contractive problems of §3 are sometimes referred to as dissipative in the numerical analysis literature; this conflicts with the terminology in the theory of differential equations which we employ here.)

Example. Consider equation (1.7), (1.8) with $\hat{a} = \hat{b} = \hat{c} = \hat{d} = \hat{e} = 1$. Then we obtain

$$u_t = (1+i)u_{xx} - (1+i)|u|^2 u + u, \qquad x \in (0, 1),$$

together with periodic boundary conditions (1.8). Taking $f(u)$ as the right-hand side of this equation and employing the standard L_2-norm and inner-product we obtain

$$\begin{aligned} \langle (1+i)u_{xx} - (1+i)|u|^2 u + u, u \rangle &= -\int_0^1 |u_x|^2 dx + \int_0^1 |u|^2 - |u|^4 dx \\ &\leqslant \int_0^1 1 - |u|^2 dx = 1 - \|u\|^2. \end{aligned}$$

Thus an infinite-dimensional analog of (5.3) is satisfied with $\gamma = 1, \omega = 1$.

RESULT 5.1. *For* (1.1), (5.3), *any* $u(0) \in \mathbb{R}^p$ *and any* $\rho > 0$ *there exists* $t^* := t^*(\rho, u(0))$ *such that*

$$\|u(t)\|^2 \leqslant \frac{\gamma}{\omega} + \rho$$

for all $t \geqslant t^*$.

Proof. Taking the inner product of (1.1) with u gives

$$\frac{1}{2}\frac{d}{dt}\|u\|^2 = \langle u, u_t \rangle \leqslant \gamma - \omega \|u\|^2.$$

Thus

$$\frac{d}{dt}(e^{2\omega t}\|u\|^2) \leqslant 2\gamma e^{2\omega t}$$

$$\Rightarrow \|u(t)\|^2 \leqslant \frac{\gamma}{\omega} + e^{-2\omega t}\left[\|u(0)\|^2 - \frac{\gamma}{\omega}\right].$$

The result follows. □

Thus all the information about the asymptotic behavior for (1.1), (5.3) is captured in a bounded set; within this set the dynamics may be very complicated, for example, chaotic. It is important to note that problems in the class (5.3) *do not necessarily satisfy a one-sided Lipschitz condition*, as the following example shows.

Example. Consider the two-dimensional problem

$$\dot{x} = -x + xy,$$
$$\dot{y} = -y - x^2.$$

We will show that this problem is dissipative in the sense of (5.3) but that the system does not satisfy a one-sided Lipschitz condition. Let $u = (x, y)^T$ and $f(u) = (-x + xy, -y - x^2)^T$; then

$$\begin{aligned}\langle f(u), u\rangle &= -x^2 - y^2 \\ &= -\|u\|^2.\end{aligned}$$

Thus (5.3) is satisfied with $\gamma = 0$ and $\omega = 1$, and Result 5.1 implies that $\|u\| \to 0$ as $t \to \infty$. Now, to show that a one-sided Lipschitz condition is not satisfied, let $v = (x', y')^T$ so that

$$\langle f(u) - f(v), u - v\rangle = -(x - x')^2 + (x - x')(xy - x'y') - (y - y')^2 + (y - y')(x'^2 - x^2)$$

Suppose that (4.1) holds and let $u = (\beta, \alpha)^T$ and $v = (\alpha, \beta)^T$, where the constants α and β are to be specified below. Notice that $\|u - v\|^2 = 2(\beta - \alpha)^2$ and observe that

$$\begin{aligned}\langle f(u) - f(v), u - v\rangle &= -2(\beta - \alpha)^2 + (\beta - \alpha)(\beta^2 - \alpha^2) \\ &= \left[\frac{1}{2}(\alpha + \beta) - 1\right]\|u - v\|^2.\end{aligned}$$

Choose $\alpha + \beta > 2(c + 1)$ to obtain a contradiction. Thus this system does not satisfy a one-sided Lipschitz condition for any $c > 0$, even though this system is dissipative in the sense of (5.3) and, in fact, the origin is globally attracting. □

This is not an isolated example. In [36] it is shown that the Lorenz equations do not satisfy a one-sided Lipschitz condition and there are many other examples within the class of dissipative systems. Because of this, we will *not* assume that (4.1) holds for systems in the class (1.1), (5.3).

It is natural to ask for a property analogous to Result 5.1 for the numerical method. However, in the light of the example above it perhaps seems too much to ask for a stability theory that is independent of initial data for problems satisfying (5.3) since there is not even a one-sided Lipschitz constant for these problems. However, this view is overly pessimistic as we now show. First we make a definition.

DEFINITION 5.2. *A* RKM *is said to be dissipative stable if, when applied to* (1.1), (5.3), *there exists* Δt_c, $R > 0$ *both independent of* U_0 *such that for all* $\Delta t \in (0, \Delta t_c)$ *and any* $U_0 \in \mathbb{R}^p$ *there exists* $n^* := n^*(U_0, \Delta t)$ *for which any sequence* $\{U_n\}_{n=0}^{\infty}$ *generated by the* RKM *satisfies*

$$\|U_n\|^2 \leqslant R$$

for all $n \geqslant n^*$.

Such a definition is implicit in the work of Foias et al. [24] and similar questions have subsequently been addressed for a variety of partial differential equations and their discretizations; see [35] for a review of the subject.

The following result shows a remarkable correspondence between the contractive nonlinear stability theories and the appropriate theory for problems satisfying (5.3): algebraically stable RKMs are again seen to have desirable stability properties.

RESULT 5.3. *Consider an algebraically stable* RKM *applied to* (1.1), (5.3) *with any* $\Delta t > 0$. *Then the* RKM *is dissipative stable for any* $\Delta t_c > 0$ *and hence*

$$\textit{algebraic stability} \Rightarrow \textit{dissipative stability.}$$

Proof. From the definition of the Runge–Kutta method it follows that

$$\|U_{n+1}\|^2 \leqslant \|U_n\|^2 + 2\Delta t \sum_{j=1}^{k} b_j \langle U_n, f_j\rangle + \Delta t^2 \sum_{i,j=1}^{k} b_i b_j \langle f_i, f_j\rangle,$$

where $f_i := f(\eta_i)$. Using the equation for the η_i we have

$$\langle U_n, f_i\rangle = \langle \eta_i, f_i\rangle - \Delta t \sum_{j=1}^{k} a_{ij} \langle f_i, f_j\rangle$$

and this gives

$$\|U_{n+1}\|^2 \leqslant \|U_n\|^2 + 2\Delta t \sum_{j=1}^{k} b_j \langle \eta_j, f_j\rangle - \Delta t^2 \sum_{i,j=1}^{k} m_{ij} \langle f_i, f_j\rangle.$$

Using algebraic stability and (5.3) we deduce that

$$\|U_{n+1}\|^2 \leqslant \|U_n\|^2 + 2\Delta t \sum_{j=1}^{k} b_j [\gamma - \omega \|\eta_j\|^2].$$

Thus we have, for any given $\epsilon > 0$, that either

$$\|U_{n+1}\|^2 \leqslant \|U_n\|^2 - 2\Delta t \epsilon \tag{5.4}$$

or

$$\sum_{j=1}^{k} b_j [\gamma - \omega \|\eta_j\|^2] \geqslant -\epsilon.$$

In the second case

$$\sum_{j=1}^{k} b_j \|\eta_j\|^2 \leqslant \frac{\gamma + \epsilon}{\omega} \tag{5.5}$$

because

$$\sum_{j=1}^{k} b_j = 1 \tag{5.6}$$

for any convergent RKM. Since the method is algebraically stable it follows that we may assume $b_i > 0$ (see the Remark after Definition 3.2) and thus (5.5) implies that

$$\|\eta_j\|^2 \leqslant \frac{\gamma + \epsilon}{\omega b_j}. \tag{5.7}$$

However, using the bound (5.7) it is possible to deduce a bound on U_{n+1} simply by noting that

$$U_{n+1} = \eta_i + \Delta t \sum_{j=1}^{k} [b_j - a_{ij}] f(\eta_j).$$

Squaring both sides of this expression we obtain

$$\|U_{n+1}\|^2 \leqslant \|\eta_i\|^2 + K\Delta t, \tag{5.8}$$

where K is independent of U_0 and depends only on the bounds (5.7). Performing a sum weighted by the b_i, and recalling (5.6), we obtain from (5.5), (5.8)

$$\|U_{n+1}\|^2 \leqslant \sum_{j=1}^{k} b_i [\|\eta_i\|^2 + K\Delta t] \leqslant \frac{\gamma + \epsilon}{\omega} + K\Delta t. \tag{5.9}$$

Thus either (5.4) or (5.9) holds. An induction based on these quantities yields the desired result with

$$R = \frac{\gamma + \epsilon}{\omega} + K\Delta t. \qquad \square$$

Remark. (i) In [36] it is shown by use of the Brouwer fixed point theorem that under (5.3), for a DJ-irreducible algebraically stable method with invertible A, the Runge–Kutta equations have a solution for all $\Delta t \geqslant 0$ and any $U_n \in \mathbb{R}^p$. However, uniqueness cannot be established under (5.3) alone.

(ii) Notice that the bound R on U_n obtained for sufficiently large n is very close to the bound $(\gamma/\omega) + \rho$ for the differential equation; thus the set into which the large-time dynamics are confined is also closely related to the equivalent set for the differential equation.

(iii) Notice again the role of algebraic stability: it enables us to determine the sign of the quadratic form defined by M, just as in the proof of Result 3.3. $\square$

6. Conservative systems. We shall start this section, as in §§3, 4, and 5 by considering the case where (1.1) is real so that $f(u) \in C^1(\mathbb{R}^p, \mathbb{R}^p)$. We will then go further and look at certain complex *matrix* systems of differential equations.

In many physical models, no energy-loss mechanism is present and conservative systems result. As a simple example of a conservative system, which arises naturally from the limit $\gamma, \omega \to 0$ of the dissipative systems considered in §5, we take the structural assumption

$$\langle f(u), u \rangle = 0 \quad \forall u \in \mathbb{R}^p. \tag{6.1}$$

Example. The equations

$$x_t = -xy^2, \qquad y_t = x^2 y$$

satisfy (6.1). $\square$

Example. The nonlinear Schrodinger equation, which is a nondissipative limit of the complex Ginzburg–Landau equation, satisfies an infinite-dimensional analogue of (6.1) and arises throughout mathematical physics. Specifically we take $\hat{a} = \hat{c} = \hat{e} = 0$ and $\hat{b} = \hat{d} = 1$ in (1.7), (1.8) and we obtain

$$u_t = iu_{xx} - i|u|^2 u. \tag{6.2}$$

Note that, using integration by parts,

$$\langle iu_{xx} - i|u|^2u, u\rangle = -\int_0^1 Re\{i|u_x|^2 + i|u|^4\}dx = 0$$

and so we have an infinite-dimensional analog of (6.1). □

By following the proof of Result 5.1 it is straightforward to see the following.

RESULT 6.1. *The solution* $u(t)$ *of* (1.1), (6.1) *satisfies*

$$\|u(t)\| = \|u(0)\|$$

for all $t \geqslant 0$.

Again it is natural to ask for numerical schemes which mimic this property. This approach was taken by Cooper [12] and a modification of the classical theory of [4] and the use of ideas from the proof of Result 5.3 enables proof of the following result, which shows a remarkable correspondence with both the classical theories of §§2 and 3 and the new theory described in §5.

RESULT 6.2. *Consider the numerical solution of* (1.1), (6.1) *by a* RKM. *If the* RKM *satisfies* $M \equiv 0$ *where* M *is defined by* (1.5), *then*

$$\|U_n\| = \|U_0\|$$

for all $n \geqslant 0$.

Proof. By definition of the RKM we have

$$\|U_{n+1}\|^2 = \|U_n\|^2 + 2\Delta t \sum_{i=1}^{k} b_i \langle U_n, f(\eta_i)\rangle + \Delta t^2 \sum_{i,j=1}^{k} b_i b_j \langle f(\eta_i), f(\eta_j)\rangle.$$

Using the defining equation for the η_i gives

$$\|U_{n+1}\|^2 = \|U_n\|^2 + 2\Delta t \sum_{i=1}^{k} b_i \langle \eta_i, f(\eta_i)\rangle - \Delta t^2 \sum_{i,j=1}^{k} m_{ij} \langle f(\eta_i), f(\eta_j)\rangle.$$

Using the fact that $M \equiv 0$, and the structural assumption (6.1), the result follows. □

Remark. (i) Algebraically stable methods of arbitrarily high order which satisfy $M \equiv 0$ do exist: they are those schemes based on Gauss–Legendre quadrature and discussed in [6] and [3]. In particular, the implicit midpoint rule (1.6) is algebraically stable and satisfies $M \equiv 0$.

(ii) Again the role of the matrix M is crucial; in this case not only is a bound on the quadratic form important but it is necessary to remove its contribution. Setting $M \equiv 0$ does this.

(iii) The solvability of the RK equations has not been investigated for RKMs under (6.1). □

We now consider a stronger kind of conservation: we consider the matrix system of differential equations (with * denoting Hermitian transpose)

$$Q_t = S(Q)Q, \quad Q^*(0)Q(0) = I, \tag{6.3}$$

where $Q(t)$ is a time-dependent $p \times p$ complex-valued matrix, $S(Q)$ is a skew-Hermitian matrix-valued function of Q that satisfies

$$S^*(Q) = -S(Q) \quad \forall Q \in \mathbb{C}^{p\times p} \tag{6.4}$$

and I is the $p \times p$ identity. Equation (6.3) arises in applications such as the continuous SVD and closely related problems arise in the computation of Lyapunov exponents for systems of ordinary differential equations. The system is conservative in a very strong sense: the orthonormality of the columns of the matrix Q are preserved with time evolution.

RESULT 6.3. *The solution* $Q(t)$ *of* (6.3), (6.4) *satisfies*

$$Q^*(t)Q(t) = I$$

for all $t \geqslant 0$.

Proof. Clearly

$$\frac{d}{dt}(Q^*Q) = Q^*Q_t + Q_t^*Q.$$

But

$$Q^*Q_t = Q^*S(Q)Q$$

and hence, by (6.4),

$$\frac{d}{dt}(Q^*Q) = Q^*[S(Q) + S^*(Q)]Q = 0.$$

Thus $Q^*(t)Q(t) = Q^*(0)Q(0) = I$ as required. □

Applying the standard Runge–Kutta method to the matrix system (6.3) gives, for $Q_n \approx Q(n\Delta t)$,

$$\begin{aligned} Q_{n+1} &= Q_n + \Delta t \sum_{j=1}^{k} b_j S(\Gamma_j)\Gamma_j, \\ \Gamma_i &= Q_n + \Delta t \sum_{j=1}^{k} a_{ij} S(\Gamma_j)\Gamma_j, \qquad i = 1, \ldots, k, \end{aligned}$$

where Γ_i is a complex-valued $p \times p$ matrix. We will employ the notation $S_i := S(\Gamma_i)$.

It is important in some contexts to find numerical methods which will automatically enforce the orthonormality of the columns of $Q(t)$ during numerical simulation. This was realized in [18], where the following result is proved.

RESULT 6.4. *The solution of* (6.3), (6.4) *by a* RKM *with* $M \equiv 0$ *where* M *is defined by* (1.5) *satisfies*

$$Q_n^*Q_n = I$$

for all $n \geqslant 0$.

Proof. From the definition of the RKM applied to (6.3) we obtain

$$\begin{aligned} Q_{n+1}^*Q_{n+1} &= \left[Q_n^* + \Delta t \sum_{i=1}^{k} b_i \Gamma_i{}^* S_i^* \right] \left[Q_n + \Delta t \sum_{j=1}^{k} b_j S_j \Gamma_j \right] \\ &= Q_n^*Q_n + \Delta t \sum_{i=1}^{k} b_i \Gamma_i{}^* S_i^* Q_n + \Delta t \sum_{j=1}^{k} b_j Q_n^* S_j \Gamma_j \\ &\qquad + \Delta t^2 \sum_{i,j=1}^{k} b_i b_j \Gamma_i{}^* S_i^* S_j \Gamma_j. \end{aligned}$$

Now, from the defining equation for the Γ_i,

$$\Gamma_i^* S_i^* Q_n = \Gamma_i^* S_i^* \Gamma_i - \Delta t \sum_{j=1}^{k} a_{ij} \Gamma_i^* S_i^* S_j \Gamma_j$$

and

$$Q_n^* S_j \Gamma_j = \Gamma_j^* S_j \Gamma_j - \Delta t \sum_{i=1}^{k} a_{ji} \Gamma_i^* S_i^* S_j \Gamma_j.$$

Combining these three expressions we find that

$$Q_{n+1}^* Q_{n+1} = Q_n^* Q_n + \Delta t \sum_{i=1}^{k} b_i \Gamma_i^* [S_i^* + S_i] \Gamma_i - \Delta t^2 \sum_{i,j=1}^{k} m_{ij} \Gamma_i^* S_i^* S_j \Gamma_j.$$

Setting $M \equiv 0$ and employing (6.4) we obtain

$$Q_{n+1}^* Q_{n+1} = Q_n^* Q_n$$

and the desired result follows. □

Remark. (i) The result presented in [18] employs the theory of symplectic integrators as outlined in the next section and yields an "if and only if" result.

(ii) The proof given here once again makes clear the role of the positive-definite quadratic form defined by M and its annihilation by the choice $M \equiv 0$. Recall again that algebraically stable schemes satisfying $M \equiv 0$ exist and so, once again, the importance of algebraic stability is apparent.

(iii) The solvability of the Runge–Kutta equations has not been addressed here. However, in [18] an explicit iteration scheme is constructed which, if iterated to convergence, satisfies the Runge–Kutta equations but which also retains the orthonormality of the system regardless of the number of iterations used. This then corresponds to a linearly implicit numerical method which is "stable" in an appropriate sense.

(iv) A result unifying Results 6.2 and 6.4 may be found in [42]. □

7. Hamiltonian systems. The class of conservative systems induced by the inner product structure (6.1) is clearly a somewhat restrictive one and it is natural to broaden the scope somewhat to include more general schemes with conservation properties. To this end we consider the case where (1.1) is a real Hamiltonian system of even dimension with $f(u) \in C^1(\mathbb{R}^p, \mathbb{R}^p)$ and $p = 2N$. To establish a connection with §6 we consider first the linear problem

$$u_t = JAu, \tag{7.1}$$

where A is positive definite symmetric and where J is a skew-symmetric matrix satisfying

$$J^T = J^{-1} = -J. \tag{7.2}$$

Then we may define a norm based on A by

$$\|u\|^2 = \frac{1}{2} u^T A u.$$

It follows that

$$\frac{d}{dt}\|u\|^2 = \frac{1}{2}[u_t^T Au + u^T Au_t] = \frac{1}{2}[u^T AJ^T Au + u^T AJAu] = 0.$$

This is equivalent to Result 6.1 and shows conservation of the Hamiltonian

$$H(u) := \frac{1}{2}u^T Au.$$

However, for nonlinear Hamiltonian systems this equivalence does not hold.

Given $H \in C^2(\mathbb{R}^{2N}, \mathbb{R})$, general Hamiltonian systems are of the form (1.1), where

$$f(u) = J\nabla H(u) \tag{7.3}$$

and J is a skew-symmetric matrix satisfying (7.2). Thus we shall consider the case where

$$J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$$

and I is the $N \times N$ identity. Equation (1.1), (7.3) then takes the familiar form

$$p_t = \nabla_q H(u), \qquad q_t = -\nabla_p H(u),$$

where $u^T = (p^T, q^T)$ for $p, q \in \mathbb{R}^N$ and ∇_p (respectively, ∇_q) denotes the gradient with respect to the p (respectively, q) variables.

Example. A simple example is the system

$$p_t = p^2 q, \quad q_t = -pq^2,$$

which corresponds to the Hamiltonian $\frac{p^2q^2}{2}$. □

Example. The nonlinear Schrodinger equation, (6.2), (1.8) is Hamiltonian with conjugation replacing the transpose and i playing the role of the skew symmetric operator J :

$$u_t = -i\nabla F(u),$$

where

$$F(u) = \int_0^1 \frac{1}{2}|u_x|^2 + \frac{1}{4}|u|^4 dx,$$

and ∇ represents the variational derivative with respect to changes in u, confined to an appropriate function space. □

Two important properties of Hamiltonian systems are described in Result 7.2. In order to explain the result we need to define the following.

DEFINITION 7.1. *A mapping $G(U) \in C(\mathbb{R}^{2N}, \mathbb{R}^{2N})$ is said to be symplectic if*

$$DG(U)^T JDG(U) = J \quad \forall U \in \mathbb{R}^{2N}.$$

Here DG denotes the Jacobian of the mapping G with respect to the variable U. We will use an analogous notation for mappings other than G throughout this section.

RESULT 7.2. *Solutions of* (1.1), (7.3) *satisfy*

(i) $H(u(t)) = H(u(0))$ *for all* $t \geqslant 0$;

(ii) *if the solution operator $G(U; t)$ is defined by $u(t) = G(u(0); t)$ for given initial data $u(0)$ then $G(\bullet, t)$ is a symplectic mapping for each $t \in \mathbb{R}^+$.*

Proof. The first fact follows in a straightforward way since

$$\frac{d}{dt}H(u(t)) = \frac{1}{2}[\nabla H(u)^T u_t + u_t^T \nabla H(u)]$$

$$= \frac{1}{2}[\nabla H(u)^T J\nabla H(u) + \nabla H(u)^T J^T \nabla H(u)] = 0,$$

since $J^T = -J$. The result follows.

For the second part, let $R(t)$ denote $DG(U; t)$, where D denotes the Jacobian with respect to U. Then $R(t)$ satisfies the matrix differential equation

$$R_t = JA(t)R, \quad R(0) = I$$

where $A(t)$ is the Hessian of $H(u)$ evaluated at $u = u(t)$ and is hence symmetric. Now let $V(t) = R^T JR$ and note that $V(0) = J$. Clearly

$$V_t = R_t^T JR + R^T JR_t = R^T AJ^T JR + R^T JJAR.$$

Now, using (7.2) we obtain

$$V_t = R^T AR - R^T AR = 0$$

and hence $V(t) = J$ for all t. By definition of V the result follows. □

Clearly (i) is a conservation property; since H is in general not a positive-definite quadratic form this property is not equivalent to Result 6.1 except for the linear problem (7.1) with positive definite A. Although it is heavily disguised, (ii) is also a conservation property: it states that the area of the projection of any set in $\mathbb{R}^{2N}$ onto certain distinguished planes in $\mathbb{R}^2$ is preserved under the solution operator G [1]. Again it is natural to ask that the conservation properties (i) and (ii) are inherited by any numerical approximations. In this context the following result of Sanz-Serna [42] and of Lasagni [41] is of interest since it again shows a close relationship with the classical theory of §3 and in particular the role of the matrix M from algebraic stability theory in preserving (ii).

RESULT 7.3. *Solutions of* (1.1), (7.3) *by the* RKM *with* $M \equiv 0$, *where* M *is defined by* (1.5), *define a symplectic mapping for each* $\Delta t \geqslant 0$.

Proof. The Runge–Kutta method defines a mapping $U \to W$ determined implicitly by the equations

$$W = U + \Delta t \sum_{j=1}^{k} b_j f(\eta_j),$$
$$\eta_j = U + \Delta t \sum_{j=1}^{k} a_{ij} f(\eta_j).$$

We let $R = DW(U)$ and $\Gamma_j = D\eta_j(U)$ and denote the Jacobian of $f(\eta)$ with respect to η evaluated at $\eta = \eta_i$ by $Df_i = Df(\eta_i)$. Then, differentiating the mapping with respect to U gives

$$R = I + \Delta t \sum_{j=1}^{k} b_j Df_j \Gamma_j,$$
$$\Gamma_i = I + \Delta t \sum_{j=1}^{k} a_{ij} Df_j \Gamma_j.$$

Thus we obtain

$$R^T JR = \left[I + \Delta t \sum_{i=1}^{k} b_i {\Gamma_i}^T Df_i^T \right] J \left[I + \Delta t \sum_{j=1}^{k} b_j Df_j \Gamma_j \right]$$

so that

$$\begin{aligned} R^T JR = J &+ \Delta t \sum_{i=1}^{k} b_i {\Gamma_i}^T Df_i^T J \\ &+ \Delta t \sum_{j=1}^{k} b_j JDf_j \Gamma_j + \Delta t^2 \sum_{i,j=1}^{k} b_i b_j {\Gamma_i}^T Df_i^T JDf_j^T \Gamma_j. \end{aligned} \tag{7.4}$$

Now, from the defining equations for the Γ_i,

$$\Gamma_i{}^T Df_i^T J = \Gamma_i{}^T Df_i^T J\Gamma_i - \Delta t \sum_{j=1}^{k} a_{ij} \Gamma_i{}^T Df_i^T JDf_j\Gamma_j$$

and

$$JDf_j\Gamma_j = \Gamma_j{}^T JDf_j\Gamma_j - \Delta t \sum_{i=1}^{k} a_{ji} \Gamma_i{}^T Df_i^T JDf_j\Gamma_j.$$

Combining these expression with (7.4) we obtain

$$R^T JR = J + \Delta t \sum_{j=1}^{k} b_i \Gamma_i{}^T [Df_i^T J + JDf_i]\Gamma_i - \Delta t^2 \sum_{i,j=1}^{k} m_{ij} \Gamma_i{}^T Df_i^T JDf_j\Gamma_j.$$

Since the method satisfies $M \equiv 0$ we obtain

$$R^T JR = J + \Delta t \sum_{j=1}^{k} b_i \Gamma_i{}^T [Df_i^T J + JDf_i]\Gamma_i.$$

Now, $Df_i = JA_i$ where A_i, the Hessian of H evaluated at η_i, is symmetric. Hence, using (7.2),

$$Df_i^T J + JDf_i = A_i^T J^T J + JJA_i = A_i^T - A_i = 0.$$

Thus $R^T JR = J$ so that the RKM defines a symplectic mapping for each $\Delta t \geqslant 0$. □

Remark. (i) Again this result assumes the solvability of the Runge–Kutta equations. This matter has not been investigated in detail for Hamiltonian systems.

(ii) Again the role of the matrix M is clear: a certain quadratic form is annihilated by setting $M \equiv 0$. Indeed Results 6.2, 6.4, and 7.3 all fall under the umbrella of a general result showing that all quadratic first integrals of (1.1) are preserved by Runge–Kutta schemes with $M \equiv 0$; see the discussion in [42]. □

For general nonlinear, nonintegrable Hamiltonian problems it is not possible to enforce both properties (i) and (ii) from Result 7.2 onto a numerical scheme since it would then have to be exact; see [26]. Thus it is an open and interesting question to determine the relative merits of preserving the two properties under discretization; see for example [46] where energy-momentum conserving methods are shown to be superior to symplectic momentum conserving methods for an application in elasto-dynamics.

To discuss Hamiltonian systems in detail is well beyond the scope of this review. Here our purpose is merely to emphasize connections with other classes of problems. For a complete overview of the numerical analysis of Hamiltonian systems see [44].

8. Remarks on multistep methods. Throughout the paper we have concentrated on Runge–Kutta methods; this has allowed a unified exposition and the theme of algebraic stability has run throughout. Nonetheless, much of the theory for RKMs was developed in tandem with that for Linear Multistep and One-Leg Methods (LMMs and OLMs) and indeed in the 1960s and 1970s the theory for RKMs was often predated by that for multistep methods. Thus it is in order to briefly sketch how the theory for LMMs and OLMs fits in to that described here. An important point to appreciate is that LMMs and OLMs naturally define a dynamical system on a space of higher dimension than the original problem—specifically in $\mathbb{R}^{pk}$ or $\mathbb{C}^{ppk}$ for a k-step method—and this is a source of some difficulty.

For linear decay problems the properties of A-stability and absolute stability have natural analogues for multistep methods and indeed this was the starting point for numerical stability theory [14]. The importance of contractive problems in numerical analysis was recognized by Dahlquist in [15] in the context of multistep methods; the form of G-stability was defined for multistep methods applied to (1.1), (3.1) and inheriting a notion of contractivity. Subsequently a remarkable equivalence theorem, analogous to Result 3.4, was proved: *G-stability is equivalent to A-stability* for OLMs [16]. For gradient systems there has been a little work on multistep methods; in particular in [23] it is proved that the first three backward differentiation formulae are gradient stable, employing a natural generalization of Definition 4.3. The concept of dissipative stability is generalized to multistep methods in [32] where the result *dissipative stability is equivalent to A-stability* is proved for LMMs and OLMs, using the equivalence of G-stability with A-stability. This result is analogous to Result 5.3 for RKMs. Ideas relating to conservation properties and symplectic structure for multistep methods are considered in [21]. The preservation of orthonormality properties in matrix differential equations are studied in [18].

9. The effect of error control. An important question which we briefly discuss here is whether the variation of time-step according to local error control will automatically enforce some form of numerical stability, even for explicit schemes. Such results are conjectured in [27] and [43], based on illuminating studies of particular examples. In our opinion it would be valuable to develop further the mathematical theory of the stability of variable step-size codes; in particular it would be of interest to identify error-controlled schemes which yield the correct long-time qualitative behavior for an interval of the error tolerance τ independent of initial data. This is the natural generalization of the contractive, gradient and dissipative stability theories described in this review for fixed step schemes.

However, it is not immediately clear why such results concerning error-controlled schemes should be true since local error control is an *accuracy requirement* whilst we are seeking *stability* results. In a notable paper, Hall [31] established a remarkable connection between accuracy and stability for error control schemes. We illustrate this with a simple example modified from [31] and [27]: consider (1.1) with $p = 1$ and

$$f(u) = -u.$$

If we apply the explicit Euler scheme with variable time-step, then we obtain

$$U_{n+1} = U_n - \Delta t_n U_n$$

where the time-step Δt_n now varies with n. This is a first-order accurate approximation to the true solution; that is the error over one step of length Δt is proportional to Δt^2. A second-order accurate approximation is formed, with error of $\mathcal{O}(\Delta t^3)$ over one step, by calculating the first step of a Trapezoidal rule correction:

$$V_{n+1} = U_n - \frac{\Delta t_n}{2}[U_{n+1} + U_n].$$

A simple error estimate for U_{n+1} is then formed as the difference between U_{n+1} and V_{n+1} on the assumption that Δt is small. The *error per unit step* strategy requires that Δt_n is chosen so that

$$\|U_{n+1} - V_{n+1}\| \leqslant \frac{1}{2}\Delta t_n \tau, \tag{9.1}$$

where $\tau \ll 1$ is an error tolerance. (The factor $\frac{1}{2}$ is chosen to simplify (9.5) below; it is simply a matter of definition.) Under this local error control we deduce that, since the standard

Euclidean norm $\|\bullet\|$ is equivalent to $|\bullet|$ in dimension $p = 1$,

$$|U_n - U_{n+1}| \leqslant \tau$$

is required for the step to be acceptable and hence that

$$\Delta t_n \leqslant \frac{\tau}{|U_n|}$$

is required. If we choose the largest time-step compatible with this error control, then we obtain

$$U_{n+1} = U_n \left(1 - \frac{\tau}{|U_n|}\right), \quad \Delta t_n = \frac{\tau}{|U_n|}. \tag{9.2}$$

Straightforward analysis shows that

$$|U_n| > \frac{\tau}{2} \Rightarrow |U_{n+1}| < |U_n|$$

while

$$|U_n| \leqslant \frac{\tau}{2} \Rightarrow |U_{n+1}| \leqslant \tau.$$

Using this it is possible to show that the local error control forces iterates to enter and remain in interval $[-\tau, \tau]$ about the origin; during this process the time-step approaches the linear stability limit. In other words the error control acts to force the correct long-time behavior, up to an error proportional to the tolerance.

This kind of desirable behavior can be generalized to the dissipative and gradient systems studied in §§4 and 5—see [52] for details. Here we outline the key to that analysis which revolves around the fact that certain error control mechanisms force the RKM to behave like an algebraically stable RKM even if the underlying method is not algebraically stable in a fixed time-step implementation.

One of the simplest error control strategies for the solution of (1.1) is to take the explicit Euler scheme

$$U_{n+1} = U_n + \Delta t_n f(U_n) \tag{9.3}$$

and then form the more accurate approximation

$$V_{n+1} = U_n + \frac{\Delta t_n}{2}[f(U_n) + f(U_{n+1})]. \tag{9.4}$$

This generalizes what we did for the linear problem above. Thus the difference of U_{n+1} and V_{n+1} is an estimate of the error incurred in (9.3) and the error per unit step strategy then requires that Δt_n is chosen so that (9.1) is satisfied. This implies that

$$\|f(U_n) - f(U_{n+1})\| \leqslant \tau \tag{9.5}$$

and hence that, under error control, the explicit scheme (9.3) is never far from the backward Euler scheme (9.3). Specifically we have that

$$U_{n+1} = U_n + \Delta t_n f(U_{n+1}) + \Delta t_n E,$$

where $\|E\| \leqslant \tau$ by (9.5). The backward Euler scheme is algebraically stable and for this reason we might expect that the error control confers desirable stability properties on the explicit scheme. This intuition is placed on a firm mathematical foundation in [52] for the contractive, gradient, and dissipative problems studied here in sections 3,4 and 5 and a class of error control schemes (including (9.3), (9.4), (9.1), and the Fehlberg (2,3) pair). These schemes are shown to be stable in the sense that desirable long-time behavior is guaranteed for an interval of τ independent of initial data.

10. Conclusions. It will be clear from reading this article that the numerical stability theory for problems in §§4–9 is far from complete. Nonetheless, it should also be clear that the classes of problems in §§4–9 all form a natural progression from the simple problems in §§2 and 3. Furthermore, the problems described in §§2–9 arise in a variety of different application areas and admit many interesting and complicated dynamical features such as exponential attraction to a unique equilibrium point, multiple competing equilibria, dissipative chaos, conservation properties and finally Hamiltonian systems which can exhibit both integrability and Hamiltonian chaos. An important point is that there are clear indications of connections in the numerical stability theory for *all these problems*. In particular, algebraic stability plays a fundamental role. We make a subjective list of open problems:

- To further explore the classes of numerical methods which are gradient or dissipative stable in the sense of Definition 4.2 and Definition 5.2. Relatedly to determine whether the definitions themselves are appropriate or whether they should be modified.
- To assess the relative merits of Hamiltonian conserving algorithms which preserve the property of Result 7.2(i), and symplectic algorithms which inherit the property of Result 7.2(ii). In particular, it is of interest to determine what can be said about the behavior of the Hamiltonian for symplectic schemes. It will probably be beneficial to impose a variety of structural assumptions on the Hamiltonian H in an attempt to further assess the relative merits of symplectic and conserving algorithms in different contexts.
- To close the gap between methods deemed to be "good" according to rigorous mathematical stability theories and the often different and larger class of methods which "work well in practice." In this context it is perhaps important to make clear mathematical statements about what it means for a code to work well in the context of long-time integration. To this end it may be valuable to develop a rigorous mathematical framework for the evaluation of the stability of variable time-step codes.
- To identify other classes of problems motivated either by real applications or by a need for theoretical understanding of the differential equations, for which it would be valuable to develop numerical stability theories.

Finally we conclude with a disclaimer: it is not our purpose to completely review the subject of numerical stability theory for initial value problems. We have concentrated on the mathematical properties of the underlying problems and this has been our unifying theme. For this reason there are numerous references to related work in the numerical analysis literature that have not been made here.

Acknowledgments. We are grateful to Luca Deici, Kjell Gustafsson, Arieh Iserles, Bob Russell, Juan Simo, and Marc Spijker for helpful conversations and to the referees for many useful suggestions.

REFERENCES

[1] V. ARNOLD, *Mathematical Methods of Classical Mechanics*, Springer, New York, 1978.
[2] N. BHATIA AND G. SZEGO, *Stability Theory of Dynamical Systems*, Springer, New York, 1970.
[3] K. BURRAGE, *High order algebraically stable Runge–Kutta methods*, BIT, 18 (1978), pp. 373–383.
[4] K. BURRAGE AND J. BUTCHER, *Stability criteria for implicit Runge–Kutta processes*, SIAM J. Numer. Anal., 16 (1979), pp. 46–57.
[5] J. BUTCHER, *Implicit Runge–Kutta processes*, Math. Comp., 18 (1964), pp. 50–64.
[6] ———, *A stability property of implicit Runge–Kutta methods*, BIT, 15 (1975), pp. 358–361.
[7] ———, *The Numerical Analysis of Ordinary Differential Equations*, Wiley, Chichester, 1987.
[8] M. CALVO AND J. SANZ-SERNA, *The development of variable-step symplectic integrators, with applications to the two-body problem*, in Proceedings of the 14th Dundee Conference on Numerical Analysis, London, 1991, Pitman.

[9] J. CHAPMAN, S. HOWISON, AND J. OCKENDON, *Macroscopic models for superconductivity*, SIAM Rev., 34 (1992), pp. 529–560.
[10] G. COOPER, *A modification of algebraically stability for implicit Runge–Kutta methods*, Technical Report, University of Sussex, 1986.
[11] ———, *On the existence of algebraically stable Runge–Kutta methods*, IMA J. Numer. Anal., 6 (1986), pp. 325–330.
[12] ———, *Stability of Runge–Kutta methods for trajectory problems*, IMA J. Numer. Anal., 7 (1987), pp. 1–13.
[13] M. CROUZIEX, *Sur la B-stabilite des methodes de Runge–Kutta*, Numer. Math., 32 (1979), pp. 75–82.
[14] G. DAHLQUIST, *A special stability problem for linear multistep methods*, BIT, 3 (1963), pp. 27–43.
[15] ———, *Error analysis for a class of methods for stiff non-linear initial value problems*, in Numerical Analysis, Dundee 1975, Springer, 1975, pp. 60–74.
[16] ———, *G-stability is equivalent to A-stability*, BIT, 18 (1978), pp. 384–401.
[17] G. DAHLQUIST AND R. JELTSCH, *Generalized disks of contractivity for explicit and implicit Runge–Kutta methods*, TRITA-NA-7906, Dept. Numer. Anal. and Comp. Sci., Stokholm, 1979.
[18] L. DEICI, R. RUSSELL, AND E. VAN VLEICK, *Unitary integrators and applications to continuous orthonormalization techniques*, SIAM J. Numer. Anal., 31 (1994), to appear.
[19] K. DEKKER AND J. VERWER, *Stability of Runge–Kutta Methods for Stiff Nonlinear Equations*, North Holland, Amsterdam, 1984.
[20] Q. DU AND R. NICOLAIDES, *Numerical analysis of a continuum model of phase transition*, SIAM J. Numer. Anal., 28 (1991), pp. 1310–1322.
[21] T. EIROLA AND J. SANZ-SERNA, *Conservation of integrals and symplectic structure of differential equations by multistep methods*, Numer. Math., 61 (1992), pp. 281–290.
[22] C. ELLIOTT, *The Cahn–Hilliard model for the kinetics of phase seperation*, in Mathematical models for phase change problems, J. Rodrigues, ed., Birkhauser, Basel, Switzerland, 1988.
[23] C. ELLIOTT AND A.M. STUART, *The global dynamics of semilinear parabolic equations*, SIAM J. Numer. Anal., 30 (1993).
[24] C. FOIAS, M. S. JOLLY, I. G. KEVREKIDIS, AND E. S. TITI, *Dissipativity of numerical schemes*, Nonlinearity, 4 (1991), pp. 591–613.
[25] D. FRENCH AND S. JENSEN, *Long-time behavior of arbitrary order continuous time Galerkin schemes for some one-dimensional phase transition problems*, IMA J. Numer. Anal., (1992), to appear.
[26] Z. GE AND J. MARSDEN, *Lie-Poisson Hamilton–Jacobi theory*, Phys. Lett. A., 133 (1988), pp. 134–139.
[27] D. GRIFFITHS, *The dynamics of some linear multistep methods with step-size control*, in Proc. 12 Biennial Dundee Conference on Numerical Analysis, Pitman, London, 1987.
[28] E. HAIRER, A. ISERLES, AND J. SANZ-SERNA, *Equilibria of Runge–Kutta methods*, Numer. Math., 568 (1989), pp. 243–254.
[29] E. HAIRER AND G. WANNER, *Solving Ordinary Differential Equations* II: *Stiff Problems*, Springer-Verlag, Berlin, 1991.
[30] J. HALE, *Asymptotic behavior of dissipative systems*, American Mathematical Society, Providence, RI, 1988.
[31] G. HALL, *Equilibrium states of Runge–Kutta schemes*, ACM Trans on Math. Software, 11 (1985), pp. 289–301.
[32] A. HILL, *Global dissipativity for A-stable nethods*, to be submitted to SIAM J. Numer. Anal., (1994).
[33] M. HIRSCH AND S. SMALE, *Differential Equations, Dynamical Systems and Linear Algebra*, Academic Press, London, 1974.
[34] A. HUMPHRIES, *Numerical Analysis of Dynamical Systems*, Ph.D. thesis, University of Bath, 1994.
[35] A. R. HUMPHRIES, D. A. JONES, AND A. M. STUART, *Approximation of dissipative partial differential equations over long time intervals*. Appears in Numerical Analysis, Dundee, 1993, D. F. Griffiths and G. A. Watson, eds., Longman, New York, 1994.
[36] A. HUMPHRIES AND A. M. STUART, *Runge–Kutta methods for dissipative and gradient dynamical systems*, SIAM J. Numer. Anal., (1994), to appear.
[37] W. HUNDSDORFER AND M. SPIJKER, *A note on B-stability of Runge–Kutta methods*, Numer. Math., 36 (1981), pp. 319–331.
[38] A. ISERLES, *Stability and dynamics of numerical methods for nonlinear ordinary differential equations*, IMA J. Numer. Anal., 10 (1990), pp. 1–30.
[39] J. KRAAIJEVANGER, *Contractivity of Runge–Kutta methods*, preprint TW-89-12, University of Leiden, 1989.
[40] J. D. LAMBERT, *Numerical Methods for Ordinary Differential Systems*, Wiley, Chichester, 1992.
[41] F. LASAGNI, *Canonical Runge–Kutta methods*, J. Appl. Math. Phys., 39 (1988), pp. 951–953.
[42] J. SANZ-SERNA, *Runge–Kutta schemes for Hamiltonian systems*, BIT, 28 (1988), pp. 877–883.
[43] ———, *Numerical ordinary differential equations vs. dynamical systems*, in Proceedings of the IMA Conference on Dynamics of Numerics and Numerics of Dynamics, Bristol, 1990, Cambridge University Press, Cambridge, 1992.

[44] J. SANZ-SERNA, *Symplectic integrators for Hamiltonian problems: an overview*, Acta Numerica, 1 (1992), pp. 243–286.

[45] R. SCHERER AND H. TURKE, *Algebraic characterization of a-stable Runge–Kutta schemes*, Appl. Numer. Math., 5 (1989), pp. 133–144.

[46] J. SIMO AND N. TARNOW, *The discrete energy-momentum method. Part* I. *conserving algorithms for elastodynamics*, 1992. Stanford University Division of Applied Mechanics Report 92–3.

[47] M. SPIJKER, *Contractivity in the numerical solution of initial value problems*, Numer. Math., 42 (1983), pp. 271–290.

[48] ———, *Stepsize restrictions for stability of one-step methods in the numerical solution of initial value problems*, Math. Comp., 45 (1985), pp. 377–392.

[49] ———, *A note on contractivity in the numerical solution of initial value problems*, BIT, 27 (1987), pp. 424–437.

[50] A. M. STUART, *The global attractor under discretization*, in Continuation and Bifurcations: Numerical Techniques and Applications, D. Roose, B. De Dier, and A. Spence, eds., Kluwer Academic Publishers, Amsterdam, 1990.

[51] ———, *Numerical analysis of dynamical systems*, in Acta Numerica, Cambridge University Press, 1994.

[52] A. M. STUART AND A. HUMPHRIES, *The essential stability of local error control for dynamical systems*, SIAM J. Numer. Anal., to appear.

[53] R. TEMAN, *Infinite Dimensional Dynamical Systems in Mechanics and Physics*, Springer-Verlag, New York, 1988.